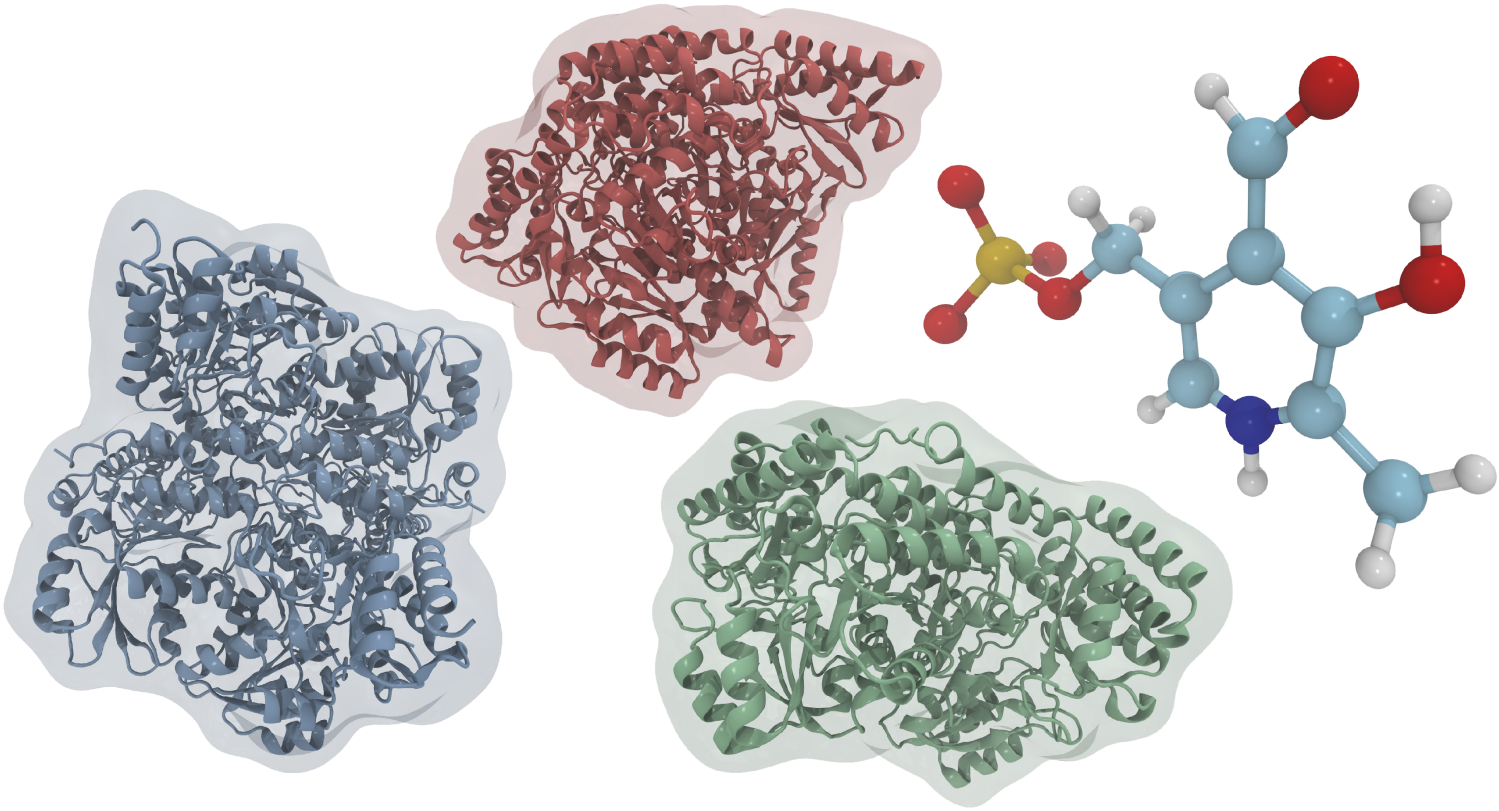


Computational Studies Addressed to PLP-Dependent Enzymes

Henrique Silva Fernandes

Doutoramente em Química

BioSIM, UCIBIO@REQUIMTE, Faculdade de Medicina da Universidade do Porto

2021

Supervisor

Nuno Manuel Ferreira de Sousa de Azevedo Cerqueira, Researcher, BioSIM, UCIBIO@REQUIMTE, Faculdade de Medicina da Universidade do Porto

Co-Supervisor

Sérgio Filipe Maia de Sousa, Researcher, BioSIM, UCIBIO@REQUIMTE, Faculdade de Medicina da Universidade do Porto

This Ph.D. dissertation was conducted at the UCIBIO@REQUIMTE, in the BioSIM Research group of the Faculty of Medicine of University of Porto under the supervision of Nuno Manuel Ferreira de Sousa de Azevedo Cerqueira and Sérgio Filipe Maia de Sousa.

This work was financed by the “Fundação para a Ciência e Tecnologia” (FCT) through a PhD scholarship (SFRH/BD/115396/ 2016).

The projects developed in this thesis were supported by the Applied Molecular Biosciences Unit – UCIBIO, which is financed by national funds from FCT (UIDP/04378/2020 and UIDB/04378/2020). This work was also partially funded by FCT through the project PTDC/QUI-QFI/31689/2017.

Some of the calculations were produced with the support of INCD funded by FCT and FEDER under the project 01/SAICT/2016 number 022153, and CPCA/A00/7140/2020 and CPCA/A00/7145/2020 projects.

Some of the calculations were produced using the QTREX high-throughput computing (HPC) cluster at the Faculty of Sciences of the University of Porto.

Most calculations were produced using the Flash HPC cluster at the Faculty of Medicine of the University of Porto.

This page was intentionally left blank.

Agradecimentos

A escrita e a apresentação desta tese estabelecem um marco importante em mais uma etapa da minha vida académica. Por isso, não poderia deixar de agradecer a todas as pessoas, que de alguma forma, deixaram a sua marca ao longo destes últimos anos.

Em primeiro lugar, quero prestar o meu agradecimento ao Doutor Nuno Cerqueira, com quem tive oportunidade de trabalhar desde o mestrado, por todos os ensinamentos, desafios e motivação. Este agradecimento estende-se também ao Doutor Sérgio Sousa, que na qualidade de coorientador, muito contribuiu para o meu desenvolvimento como investigador, espicaçando o interesse por diferentes problemas químicos e bioquímicos.

Não poderia deixar de agradecer à Professora Doutora Maria João Ramos pelo apoio e ensinamentos durante o período em tive a oportunidade de fazer parte do seu grupo de investigação.

A todos os amigos do grupo BioSIM, pelo companheirismo, partilha de ideias e conhecimentos, e momentos de descontração, nomeadamente, a Carla, o André, a Juliana, a Rita e a Cátia. Estendo este agradecimento também ao Rui, João, Eduardo, Pedro Ferreira, Pedro Paiva, Fabiola, Ritinha e Óscar.

Aos Maiores, Nuno e João (Chaves), um muito obrigado por todas as aventuras e um grande abraço. Aproveito também para agradecer à Bárbara, Daniela, Ana e Joana pela amizade de longa data.

À minha irmã, pelo apoio e amizade.

Um agradecimento muito especial aos meus pais pelo esforço que fizeram pela minha educação e o encorajamento para seguir os meus sonhos e objetivos.

Não sei palavras que possam descrever o quão importante és para mim e o quão compreensiva, paciente e encorajadora foste ao longo de todos os anos que já partilhámos juntos. A ti, Jéssica, dedico esta tese.

This page was intentionally left blank.

Abstract

Pyridoxal-5'-phosphate (PLP) is the catalytic form of vitamin B_6 and is employed, as a cofactor, by several enzymes with more than 140 distinct catalytic activities. PLP-dependent enzymes can catalyze decarboxylation, racemization, transamination, and α, β, and γ-elimination/replacement reactions. In this thesis, the catalytic mechanisms of histidine decarboxylase, serine hydroxymethyltransferase, and cystathionine-γ-lyase were studied using molecular dynamics simulations and hybrid quantum mechanics/molecular mechanics calculations. These studies enlighten the unknown mechanisms of these enzymes with an atomistic level of detail and a Gibbs free energy profile. Moreover, they unveiled some key features about how PLP-dependent enzymes can take advantage of PLP to catalyze so many different reactions.

Besides, the catalytic mechanisms for the xanthine oxidation and nitrate/nitrite reduction by xanthine oxidase were investigated employing similar methodologies. These two studies provided, for the first time, a complete description of the mechanism behind these two reactions.

As a consequence of the current COVID-19 pandemic situation, a scientific contribution was also produced during the course of this thesis. Specifically, the catalytic mechanism of the main protease of SARS-CoV-2 was described following a similar methodology.

Parallel to studying the mechanisms of the previously mentioned enzymes, some software programs were also developed and published, namely VMD Store, ToolBar, molUP, Energy Split, and chemTool. The main purpose of these software applications is to assist the setup of calculations and help the analysis of data that come from different types of calculations. These programs were essential to overcome the well-known limitations of the available software at that time.

The approximation of scientific knowledge to the general society was also taken into consideration during this thesis. Thus, some tools, such as BioSIM^{AR}, Protein Wars, and Criminal Investigation, were developed and used during science divulgation events. The main purpose of these tools is to captivate the attention and interest of the general public to more complex scientific aspects using playful technologies.

Keywords: pyridoxal-5'-phosphate (PLP), histidine decarboxylase, serine hydroxymethyltransferase, cystathionine-γ-lyase, molybdenum, xanthine oxidase, SARS-CoV-2, enzyme, catalytic mechanism, QM/MM, molecular dynamics, software development

Resumo

O piridoxal-5'-fosfato (PLP) é a forma cataliticamente ativa da vitamina B_6 e é usado, como cofator, por diversas enzimas na catálise de mais de 140 reações diferentes. As enzimas dependentes de PLP catalisam reações de descarboxilação, racemização, transaminação e α-, β-, e γ-eliminação/substituição. Nesta tese, os mecanismos catalíticos da histidina descarboxilase, serina hidroximetiltransferase e cistationina-γ-liase foram estudados usando métodos computacionais, nomeadamente dinâmica molecular e cálculos híbridos de mecânica quântica/mecânica molecular. Estes trabalhos permitiram desvendar os mecanismos, até então desconhecidos, dessas enzimas com detalhe atomístico e com perfis de energia livre de Gibbs. Além disso, foi possível desvendar algumas particularidades de como as enzimas dependentes de PLP conseguem utilizar este cofator em reações tão distintas.

Os mecanismos catalíticos de oxidação da xantina e redução de nitrate/nitrito catalisados pela xantina oxidase também foram estudados usando técnicas computacionais similares. Estes dois estudos permitiram, pela primeira vez, uma descrição completa do mecanismo associado a essas duas reações.

Devido à atual situação de pandemia por COVID-19, uma contribuição científica no sentido de travar este vírus foi produzida durante a tese. Mais especificamente, o mecanismo da protease principal do vírus SARS-CoV-2 foi estudado através dos mesmos métodos.

Paralelamente ao estudo dos mecanismos das enzimas supracitadas, alguns programas foram desenvolvidos e publicados, nomeadamente a VMD Store, ToolBar, molUP, Energy Split, e o chemTool. O principal objetivo destes programas é ajudar na preparação dos cálculos computacionais e também na análise dos resultados obtidos através dos diferentes métodos. Estes programas foram essenticais para ultrapassar algumas das limitações dos programas disponíveis na altura.

Por fim, também a aproximação do conhecimento científico à sociedade em geral foi tida em consideração. Consequentemente, algumas tecnologias, como o BioSIM^{AR}, Protein Wars, e o Criminal Investigation, foram desenvolvidas e usadas em eventos de divulgação científica. O principal propósito destas tecnologias é cativar a atenção e o interesse da sociedade para aspetos científicos de uma forma lúdica.

Palavras-chave: piridoxal-5'-fosfato (PLP), histidina descarboxilase, serina hidroximetiltransferase, cistationina-γ-liase, molibdénio, xantina oxidase, SARS-CoV-2, enzima, mecanismo catalítico, QM/MM, dinâmica molecular, desenvolvimento de software

Preamble

This dissertation is organized in six main chapters:

A. Introduction

Chapter A provides a first quick look about enzymes and their biological role. This description is followed by a reflection about the importance of studying the catalytic mechanisms of enzymes, in particular, using computational methods. In this chapter, enzymes are grouped in two different groups regarding the presence or not of a cofactor. Within the cofactor-dependent enzymes group, they are organized into enzymes that employ organic or inorganic cofactors. For organic cofactors, particular relevance will be given to pyridoxal-5'-phosphate (PLP) cofactor, which is used by the enzymes whose catalytic mechanism was detailed studied in this thesis. The catalytic mechanism of xanthine oxidase (a Mo-containing), having different substrates, was also studied in this Ph.D. project. Therefore, molybdenum is discussed in more detail to provide a quick overview about enzymes that use this metal as an important catalytic player.

B. Computational Methodologies

Chapter B aims to review the main theoretical principles applied to the computational methods used in the projects conducted during this thesis. This chapter is divided in four main sections devoted to:

- 1. Molecular Mechanics
- 2. Quantum Mechanics
- 3. Hybrid Methods
- 4. Enzymatic Kinetics

Moreover, this chapter provides an explanation for the chosen methods employed in each type of calculation.

C. Results and Discussion

This chapter consists of the major part of the thesis, since it includes the specific background contextualization (introductions), results, discussion, and conclusions for each developed study. In addition to the study of PLP-dependent enzymes, some

complementary projects were also developed. Consequently, the chapter is organized to guide the reader throughout the different studies:

- 1. PLP-dependent Enzymes
- 2. Mo-dependent Enzymes
- 3. Cofactor-independent Enzymes
- 4. Development of Scientific Software
- 5. Pedagogic Applications
- 6. Databases

In this thesis, the catalytic mechanism of different enzymes was studied using computational methods. Enzymes present different particularities, one of them is the presence or not a cofactor and whether that cofactor is organic or not. Accordingly, the PLP-dependent enzymes studied in this work were grouped in the first section (1. PLP-dependent Enzymes). The second section includes the two catalytic mechanisms studied for an enzyme that used a Mo ion as cofactor (2. Mo-dependent Enzymes). Finally, the third section was devoted to the description of the catalytic mechanism of the main protease of SARS-CoV-2 (3. Cofactor-independent Enzymes).

Sections 4, 5, and 6 address different projects of software development regarding scientific and pedagogic applications, and web-based databases.

D. Final Remarks

Chapter D summarizes the main contributions of this work for the scientific community and society.

E. References

This chapter includes all the bibliographic references appropriately mentioned through this document.

F. Appendix

Finally, chapter F includes all the appendixes that could be important as a complement of what is presented in the main text of this thesis. All the appendixes have an appropriate cross-mention in the main text.

Reading Guide

The amino acid residues are mentioned using the 3-letters nomenclature with the first capital letter (e.g., Glu). When a specific amino acid residue from a protein is mentioned, the respective residue number is added in superscript (e.g., Glu^{802}).

In order to assist the reader along the reading of the different chapters, some specific acronyms are presented in its extended form for the first time they appear in each main chapter. Nevertheless, a “List of Abbreviations” is available to clarify the meaning of all acronyms mentioned in this thesis.

All scientific publications that resulted from the work developed during the period of this Ph.D. are highlighted in infographics. Each infographic details the titles, list of authors, Digital Object Identifier (DOI), year of publication, journal/book, and impact factor (if applicable). The impact factor corresponds to the one available at the time of the publication. Moreover, the infographics include a Quick Response (QR) code to easily redirect the reader to the publication at the journal/book website.

A set of QR codes were also included to facilitate the access to certain key hyperlinks mentioned in the thesis.

This page was intentionally left blank.

All scientific publications, oral communications, and posters produced in the course of this Ph.D. project are presented in the following lists.

List of Scientific Publications (17)

Book Chapters (2)

1. Teixeira, C. S. S., **Fernandes, H. S.**, Fernandes, P. A., Ramos, M. J., & Cerqueira, N. M. F. S. A. (2017). Cancer therapies based on enzymatic amino acid depletion. In *Nanostructures for Cancer Therapy* (pp. 623-651).
2. Teixeira, C. S. S., **Fernandes, H. S.**, Sousa, S. F., & Cerqueira, N. M. F. S. A. (2019). Enzymatic Amino Acid Deprivation Therapies Targeting Cancer. In *Pharmaceutical Biocatalysis* (pp. 311-347).

Reviews (3)

1. **Fernandes, H. S.**, Silva Teixeira, C. S., Fernandes, P. A., Ramos, M. J., & Cerqueira, N. M. (2017). Amino acid deprivation using enzymes as a targeted therapy for cancer and viral infections. *Expert Opin Ther Pat, 27*(3), 283-297. doi:10.1080/13543776.2017.1254194
2. **Fernandes, H. S.**, Teixeira, C. S. S., Sousa, S. F., & Cerqueira, N. (2019). Formation of Unstable and very Reactive Chemical Species Catalyzed by Metalloenzymes: A Mechanistic Overview. *Molecules, 24*(13). doi:10.3390/molecules24132462 [Cover: page **391**]
3. Magalhães, R. P., **Fernandes, H. S.**, & Sousa, S. F. (2020). Modelling Enzymatic Mechanisms with QM/MM Approaches: Current Status and Future Challenges. *Israel Journal of Chemistry, 60*(7), 655-666. doi:10.1002/ijch.202000014

Articles (8)

1. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. (2017). The Catalytic Mechanism of the Pyridoxal-5'-phosphate-Dependent Enzyme, Histidine Decarboxylase: A Computational Study. *Chemistry (Weinheim an der Bergstrasse, Germany), 23*(38), 9162-9173. doi:10.1002/chem.201701375 [**Hot paper**]

2. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. M. F. S. A. (2018). molUP: A VMD plugin to handle QM and ONIOM calculations using the gaussian software. *J Comput Chem, 39*(19), 1344-1353. doi:10.1002/jcc.25189 [Cover: page 392]
3. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. M. F. S. A. (2018). Catalytic Mechanism of the Serine Hydroxymethyltransferase: A Computational ONIOM QM/MM Study. *ACS Catalysis, 8*(11), 10096-10110. doi:10.1021/acscatal.8b02321
4. **Fernandes, H. S.**, Sousa, S. F., & Cerqueira, N. (2019). VMD Store - A VMD Plugin to Browse, Discover, and Install VMD Extensions. *J Chem Inf Model, 59*(11), 4519-4523. doi:10.1021/acs.jcim.9b00739
5. Magalhaes, R. P., Vieira, T. F., **Fernandes, H. S.**, Melo, A., Simoes, M., & Sousa, S. F. (2020). The Biofilms Structural Database. *Trends Biotechnol, 38*(9), 937-940. doi:10.1016/j.tibtech.2020.04.002
6. Ribeiro, P. M. G., **Fernandes, H. S.**, Maia, L. B., Sousa, S. F., Moura, J. J. G., & Cerqueira, N. M. F. S. A. (2021). The complete catalytic mechanism of xanthine oxidase: a computational study. *Inorganic Chemistry Frontiers, 8*(2), 405-416. doi:10.1039/d0qi01029d
7. Gonçalves, I.G., **Fernandes, H.S.**, Melo, A., Sousa, S.F., Simões, L.C., and Simões, M. (2021). LegionellaDB – A Database on *Legionella* Outbreaks. *Cell: Trends in Microbiology*. doi:10.1016/j.tim.2021.01.015
8. Lapaillerie, D., Charlier, C., **Fernandes, H.S.**, Sousa, S.F., Lesbats, P., Weigel, P., Favereaux, A., Guyonnet-Dupérat, V., and Parissi, V. (2021). *In silico*, *in vitro* and *in cellulo* models for monitoring SARS-CoV-2 spike/human ACE2 complex, viral entry and cell fusion. *Viruses*, 13(3), 365. doi:10.3390/v13030365

Proceedings (4)

1. **Fernandes, H.**, Ramos, M. J., & Cerqueira, N. (2017). *Development of computational tools to enhance the study of catalytic mechanisms*. Paper presented at the Proceedings of MOL2NET 2017, International Conference on Multidisciplinary Sciences, 3rd edition.
2. **Fernandes, H.**, Sousa, S., & Cerqueira, N. (2018). *Is the 5,10-methylenetetrahydrofolate cofactor synthesized through a non-enzymatic or*

enzymatic mechanism? Paper presented at the Proceedings of MOL2NET 2018, International Conference on Multidisciplinary Sciences, 4th edition.

3. Magalhães, R., Vieira, T., **Fernandes, H.**, Cerqueira, N., Melo, A., Simões, M., & Sousa, S. (2018). *Creation of a Structural Database for Inhibition of Biofilm Formation*. Paper presented at the Proceedings of MOL2NET 2018, International Conference on Multidisciplinary Sciences, 4th edition.

4. **Fernandes, H. S.**, Sousa, S. F., & Cerqueira, N. M. F. S. A. (2019). Enhancing the catalytic power of serinehydroxymethyltransferase to producecommercially valuable compounds. *FEBS Open Bio, 9*(S1), 411. doi:10.1002/2211-5463.12675

List of Communications (22)

Posters (9)

1. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. M. F. S. A., (2017) Serine HydroxyMethylTransferase: Inhibit it or Improve it?. *III Symposium on Medicinal Chemistry of University of Minho*, Universidade do Minho, Braga, Portugal (May 26th) (Appendix: page 381)

2. **Fernandes, H. S.**, Ramos, M. J., Sousa, S. F., & Cerqueira, N. M. F. S. A., (2018) The catalytic mechanism of Serine Hydroxymethyltransferase: a drug target against Malaria, *UCIBIO Annual Meeting*, Lisbon, Portugal (September 28th to 29th) (Appendix: page 382)

3. **Fernandes, H. S.**, Sousa, S. F., & Cerqueira, N. M. F. S. A., (2018) Is the 5,10methylenetetrahydrofolate cofactor synthesized through a non-enzymatic or enzymatic mechanism?. *6º Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Faculdade de Medicina da Universidade do Porto, Porto, Portugal (December 21st) (Appendix: page 383) [**Best Poster Award**]

4. Magalhães, R. P., Vieira, T. F., **Fernandes, H. S.**, Cerqueira, N. M. F. S. A., Melo, A., Simões, M., & Sousa, S. F., (2018) Creation of a Structural Database for Inhibition of Biofilm Formation. *6º Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Faculdade de Medicina da Universidade do Porto, Porto, Portugal (December 21st) (Appendix: page 384)

5. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) Enhancing the catalytic power of Serine Hydroxymethyltransferase to produce commercially valuable compounds. *19th FEBS Young Scientists' Forum*, Krakow's Jagiellonian University Campus, Krakow, Poland (July 3rd-6th) (Appendix: page 385)
6. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) Enhancing the catalytic power of Serine Hydroxymethyltransferase to produce commercially valuable compounds. *The 44th FEBS Congress*, ICE Krakow Congress Centre, Krakow, Poland (July 6th-11th) (Appendix: page 385)
7. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) Using a computational approach to enhance enzymes as industrial biocatalysts. *7º Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Lisbon, Portugal (December 20th) (Appendix: page 386)
8. Vieira, T.F., Magalhães, R.P., **Fernandes, H.S.**, Simões, M., Sousa, S.F., (2019) Development of a structural database to inhibit biofilm formation. *7º Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Lisbon, Portugal (December 20th) (Appendix: page 387)
9. Sousa, C.E.A., Pereira, C., **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) RDB-TS: Development of a Database for Storage of Transition State Structures for Organic Reactions. *7º Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Lisbon, Portugal (December 20th)

Oral Communications (6)

1. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. M. F. S. A., (2016) Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase. *Oncology & Molecular Biology Parallel Oral Session of the Eleventh YES – Young European Scientist – Meeting*, CIM – Centro de Investigação Médica da Faculdade de Medicina da Universidade do Porto, Portugal (September 16th)
2. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. M. F. S. A., (2017) Development of computational tools to enhance the study of catalytic mechanisms. *Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Departamento de Física da Universidade de Coimbra, (December 22nd)

3. **Fernandes, H. S.**, Ramos, M. J., & Cerqueira, N. M. F. S. A., (2018) Serine HydroxyMethylTransferase: Inhibit it or Improve it?. *7th EuCheMS Chemistry Congress*, Liverpool, United Kingdom (August 26th-30th)
4. **Fernandes H.S.**, Teixeira, C. S. S., Rocha, J. F., Pina, A. F., Sousa, S. F., & Cerqueira, N. M. F. S. A., (2018) Desenvolvimento de Software com Aplicação no Ensino Em Química e Biologia. *VII Encontro da Divisão de Ensino e Divulgação da Química*, Escola Secundária Almeida Garrett, Vila Nova de Gaia, Portugal, (November 17th)
5. **Fernandes, H.S.**, Sousa, S.F., Maia, L.B., Moura, J.J.G, & Cerqueira, N.M.F.S.A., (2019) Production of NO by Xanthine Oxidase – a QM/MM study. *24th European Nitrogen Cycle Meeting*, Universidade NOVA de Lisboa, Lisbon, Portugal (September 12th)
6. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) molUP: How to start a computational QM/MM study of an enzyme?. *XXV Encontro Galego-Português de Química*, Santiago de Compostela, Spain (November 22nd)

Invited Oral Communications (7)

1. **Fernandes, H. S.**, Fernandes, P. A., & Ramos, M.J., (2018) Moléculas Magníficas. *V Encontro Internacional da Casa das Ciências*, Auditório Grande do Complexo Cultural de Vila Flor, Guimarães, Portugal (July 9th)
2. **Fernandes, H.S.**, (2019) Bioquímicos Extraordinários. *VII AEICBAS Biomedical Congress*, Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto, Porto, Portugal (March 17th)
3. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) The Chemistry of Vitamin B6 in Enzymes: a computational approach. *VI Jornadas da Química*, Departamento de Química e Bioquímica da Faculdade de Ciências da Universidade do Porto, Porto, Portugal (April 10th)
4. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) Software Development in Computational (Bio)Chemistry. *Bioinformatics class*, Departamento de Química e Bioquímica da Faculdade de Ciências da Universidade do Porto, Porto, Portugal (April 26th)
5. **Fernandes, H.S.**, Sousa, S.F., Maia, L.B., Moura, J.J.G, & Cerqueira, N.M.F.S.A., (2019) Production of NO by Xanthine Oxidase – a QM/MM study. *1st*

BioSIM Cycle of Conferences, Faculdade de Medicina da Universidade do Porto, Porto, Portugal (September 18th)

6. **Fernandes, H.S.**, Sousa, S.F., & Cerqueira, N.M.F.S.A., (2019) molUP: How to start a computational QM/MM study of an enzyme?. *1st BioSIM Cycle of Conferences*, Faculdade de Medicina da Universidade do Porto, Porto, Portugal (November 18th)
7. **Fernandes, H.S.**, & Sousa, S.F., (2020) A Utilização das Simulações Computacionais no Ensino e Investigação em Química. *Evento da Semana Acadêmica de Ciência e Tecnologia – IFRO*, Ji-Paraná, Brazil (November 25th)

Index

AGRADECIMENTOS ..... **III**

ABSTRACT ..... **V**

RESUMO ..... **VII**

PREAMBLE ..... **IX**

LIST OF SCIENTIFIC PUBLICATIONS (17) XIII

Book Chapters (2) ..... *xiii*

Reviews (3) ..... *xiii*

Articles (8) ..... *xiii*

Proceedings (4) ..... *xiv*

LIST OF COMMUNICATIONS (22) XV

Posters (9) ..... *xv*

Oral Communications (6) ..... *xvi*

Invited Oral Communications (7) ..... *xvii*

INDEX ..... **XIX**

INDEX OF FIGURES ..... **XXV**

INDEX OF EQUATIONS ..... **XXXV**

INDEX OF TABLES ..... **XXXIX**

LIST OF ABBREVIATIONS ..... **XLI**

A. INTRODUCTION **1**

1. ENZYMES 3

1.1. Coenzyme-dependent enzymes *7*

1.1.1. Pyridoxal-5'-Phosphate (PLP) 7

1.1.2. Tetrahydrofolate (THF) 14

1.2. Metalloenzymes *16*

1.2.1. Molybdenum 16

2. THE CLINICAL AND INDUSTRIAL RELEVANCE OF ENZYMES 19

2.1. Enzymes as Drug Targets *20*

2.2. Enzymes as Therapeutic Agents *22*

2.3. Enzymes as Industrial Catalysts *25*

B. COMPUTATIONAL METHODOLOGIES 27

1. MOLECULAR MECHANICS 30
1.1. Classic Force Fields 31
1.1.1. Bonded Interactions 31
1.1.2. Non-Bonded Interactions 33
1.2. Molecular Energy Minimization 34
1.3. Molecular Dynamics Simulations 36
1.3.1. Integration Time Step 37
1.3.2. Ensembles 38
1.3.3. Pressure and Temperature Control 38
1.3.4. Periodic Boundary Conditions 40
1.3.5. Non-Bonded Interactions Cutoff 40
2. QUANTUM MECHANICS 43
2.1. Wave Function 44
2.1.1. Born-Oppenheimer Approximation 47
2.1.2. Linear Combination of Atomic Orbitals 48
2.2. Hartree-Fock 53
2.2.1. Many-Electrons System 53
2.2.2. Hartree-Fock Theory 53
2.2.3. Slater Determinant 54
2.2.4. Variational Principle 55
2.2.5. Hartree-Fock Limit: Electron Correlation 56
2.3. Post-Hartree-Fock Methods 58
2.3.1. Configuration State Methods 58
2.3.2. Multiconfiguration SCF Methods 59
2.3.3. Møller-Plesset Perturbation Theory 60
2.3.4. Coupled-Cluster Theory 62
2.3.5. Domain Based Local Pair Natural Orbital Coupled-Cluster Method 63
2.4. Density Functional Theory 64
2.4.1. Hohenberg-Kohn Theorems 64
2.4.2. Born-Oppenheimer 65
2.4.3. Kohn-Sham Equations 65
2.4.4. Exchange-Correlation Density Functionals 66
2.5. Nuclear Hamiltonian 73
3. HYBRID METHODS 76
3.1. Quantum Mechanics/Molecular Mechanics 77
3.2. Mechanical, Electrostatic, and Polarizable Embedding 79
3.2.1. Mechanical Embedding 79
3.2.2. Electrostatic Embedding 79
3.2.3. Polarizable Embedding 80

3.3. Boundary Conditions *81*

4. Enzymatic Kinetics 82

C. RESULTS AND DISCUSSION 85

1. PLP-dependent Enzymes 87

1.1. Decarboxylation *89*

1.1.1. Introduction 91

1.1.2. Methodology 94

1.1.3. Results and Discussion 97

1.1.4. Conclusion 108

1.2. α-Elimination *111*

1.2.1. Introduction 113

1.2.2. Methodology 117

1.2.3. Results and Discussion 121

1.2.4. Conclusion 134

1.3. α-Addition *139*

1.3.1. Introduction 140

1.3.2. Methodology 143

1.3.3. Results and Discussion 146

1.3.4. Conclusion 151

1.4. γ-Elimination *153*

1.4.1. Introduction 154

1.4.2. Methodology 158

1.4.3. Results and Discussion 162

1.4.4. Conclusion 169

1.5. Chemistry of PLP-dependent enzymes *171*

1.5.1. Dunathan's theory, Intermolecular Interactions, and Electron Sink 171

1.5.2. Substrate Specificity 174

1.5.3. Proton storage 176

1.5.4. The Phosphate Group 176

1.5.5. The Pyridine Ring 177

2. Mo-dependent Enzymes 179

2.1. The Catalytic Mechanism of Xanthine Oxidase *181*

2.1.1. Introduction 183

2.1.2. Methodology 185

2.1.3. Results and Discussion 189

2.1.4. Conclusion 197

2.2. Nitrate and Nitrite Reductase Activity of Xanthine Oxidase *199*

2.2.1. Introduction 201

2.1.1. Methodology 203

2.1.2. Results and Discussion 204

2.1.3. Conclusion 211
3. Cofactor-independent Enzymes 213
3.1. The Catalytic Mechanism of the Main Protease of SARS-CoV-2 215
3.1.1. Introduction 217
3.1.2. Methodology 219
3.1.3. Results and Discussion 221
3.1.4. Conclusion 227
4. Development of Scientific Software 229
4.1. VMD Store 231
4.1.1. Introduction 233
4.1.2. Development 234
4.1.3. VMD Store Popularity and Statistics 235
4.2. ToolBar 237
4.2.1. Introduction 237
4.2.2. The ToolBar 239
4.2.3. Technical Implementation 244
4.2.4. Conclusion 244
4.3. molUP 245
4.3.1. Introduction 247
4.3.2. Development 247
4.3.3. molUP Features 247
4.3.4. Conclusion 258
4.4. Energy Split 259
4.4.1. Introduction 261
4.4.2. Development 263
4.4.3. Conclusion and Future Perspectives 273
4.5. chemTool 275
4.5.1. Introduction 277
4.5.2. Development 278
4.5.3. chemTool Features 278
4.5.4. Conclusion and Future Perspectives 288
5. Pedagogic Applications 289
5.1. BioSIM^{AR} 291
5.1.1. Introduction 293
5.1.2. Development 296
5.1.3. Implementation 301
5.1.4. Conclusion 303
5.2. Protein Wars 305
5.2.1. Introduction 307
5.2.2. Development 308
5.2.3. Conclusion 311

5.3. Criminal Investigation 313
5.3.1. Development 315
5.3.2. Conclusion 320
6. DATABASES 321
6.1. The Biofilms Structural Database 323
6.1.1. Introduction 325
6.1.2. Data Selection 326
6.1.3. Organization of the Database 327
6.1.4. Conclusion 329

D. FINAL REMARKS 331

E. REFERENCES 337

F. APPENDIX 375

1. POSTERS 377
1.1. Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase 379
1.2. Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase 380
1.3. Serine HydroxyMethylTransferase: Inhibit it or Improve it? 381
1.4. The catalytic mechanism of Serine Hydroxymethyltransferase: a drug target against Malaria 382
1.5. Is the 5,10-methylenetetrahydrofolate cofactor synthesized through a non-enzymatic or enzymatic mechanism? 383
1.6. Creation of a Structural Database for Inhibition of Biofilm Formation 384
1.7. Enhancing the catalytic power of SHMT to produce commercially valuable compounds .. 385
1.8. Using a Computational Approach to Enhance Enzymes as Industrial Biocatalysts 386
1.9. Development of a structural database to inhibit biofilm formation 387
2. COVERS 389
2.1. Cover of the Molecules 391
2.2. Cover of the Journal of Computational Chemistry 392
3. REVIEWS AND BOOK CHAPTERS 393

This page was intentionally left blank.

Index of Figures

Figure 1. Structures of pyridoxine (vitamin B_6), pyridoxamine, pyridoxal, pyridoxine-5'-phosphate, pyridoxamine-5'-phosphate, and pyridoxal-5'-phosphate.....7

Figure 2. Distribution of all already known PLP-dependent enzymes over the different enzyme classes. [13].....8

Figure 3. PLP salvage pathway present in all living organisms. (Adapted from [11]) 9

Figure 4. Overview of all types of reactions catalyzed by PLP-dependent enzymes. ...10

Figure 5. Representation of the general catalytic mechanism for the formation of the internal (IA) and external (EA) aldimines in PLP-dependent enzymes. 13

Figure 6. Structures of vitamin B_9 and THF.....14

Figure 7. Intracellular folate metabolic pathways. cSHMT: cytosolic SHMT, mSHMT: mitochondrial SHMT. [38, 39].....15

Figure 8. Schematic representation of the pterin cofactor structure. Schematic representation of the general Moco cofactor structure for enzymes belonging to each of the three main families of Mo-dependent enzymes. 17

Figure 9. Normal modes commonly applied in MM biomolecular force fields. 33

Figure 10. Schematic representation of the behavior of steepest descent and conjugate gradients algorithms used sequentially for potential-energy minimization. 35

Figure 11. Flowchart for the preparation, energy minimization, and MD simulation of a biomolecular system. 42

Figure 12. Flowchart of HF SCF method. 56

Figure 13. Saddle point at the surface of the function: $f(x, y) = 12k(x2 - y2)$. (Adapted from [168]).....74

Figure 14. Representation of an ONIOM QM/MM Model. The red atoms are included in the model system/high level (treated with a QM method) and the blue representation corresponds to the surrounding system/low level (treated with a MM method). The black atoms correspond to the link atoms that span across the boundary between both QM/MM layers.....78

Figure 15. (Left) New cartoon representation of the X-ray structure of HDC from *Homo sapiens* (PDB code 4E1O), with the external aldimine. Subunits A and B are colored in red and blue, respectively. (Right) Representation of the atoms that were placed in the HL layer during the geometry optimizations. The carbon atoms from amino acid residues and the water molecule are colored in gray whereas the ones from the external aldimine are colored in yellow. The representation was simplified by hiding the hydrogen atoms bonded to nonpolar atoms of the amino acid residue. The green spheres represent the hydrogen atoms that were used as link atoms.....95

Figure 16. RMSd values for the MD simulation of HDC with (Left) the external aldimine and (Right) the quinonoid intermediate. The blue, orange, and red lines correspond to the RMSd values of all atoms in the system; only the α-carbons; or the active site amino acid residues plus the external aldimine/quinonoid intermediate, respectively. The simulation was considered equilibrated during the last 20 ns. 97

Figure 17. Structures of the external aldimine (Left) available in the PDB file 4E1O where the PLP cofactor is bonded to the HME inhibitor; (Center) where the PLP cofactor is bonded to the natural substrate; (Right) where the imine linkage was re-orientated to fit the common configuration of other PLP-dependent enzymes. 98

Figure 18. Main models that were built to study the first step of the catalytic mechanism of HDC. The structures colored by green correspond to those were modified comparing to the previous model. The graphs show the activation and reaction energies for all the main models presented above. Dashed lines colored by blue represent important non-bond interactions at the active site. * Energy calculated from the difference between the higher energy from the scan and the reactant. ** Energy calculated from the difference between the optimized transition state and the reactant. No ZPE or thermal corrections were included in these calculations.....100

Figure 19. Structures of reactant, TS, and product of the first step of the reaction catalyzed by HDC. (Left) Wedge-Dash representation of the structures of reactant and product. (Right) *New cartoon* and licorice representations of the TS structure of the active site.....102

Figure 20. Two hypotheses were studied to address the catalytic mechanism for the second step. The structures colored in green correspond to those were modified comparing to the previous hypothesis. The energy profiles are shown for each hypothesis. Dashed lines colored by blue represent important non-bonded interactions at the active site. Energies were calculated from the difference between the optimized TS and the reactant.....104

Figure 21. Structures of reactant, TS, and product of the second step of the reaction catalyzed by HDC. (Left) Wedge-Dash representation of the structures of reactant and product. (Right) New cartoon and licorice representations of the TS structure of the active site.....106

Figure 22. (Top) The energy profile for the two specific steps catalyzed by HDC. (Bottom) Surface mapping of the Merz-Singh-Kollman charges of the external aldimines with the reactant and product and quinonoid intermediate.....108

Figure 23. Analysis of the relevant dihedral angles of the reactant, TS, and product for the first step. Dihedral angles: C2-C7-Cα-C; C2-C7-Cα-Cβ; C2-C7-C8-N. 110

Figure 24. Structure of the external aldimine (PLP + L-serine) (Left) and the THF cofactor (Right). 114

Figure 25. Scheme representing the mechanism behind the internal and external aldimine formation for PLP-dependent enzymes and the specific α-elimination reaction. 115

Figure 26. (A) Alignment of the protein sequences of the *E.coli* and *P. falciparum* forms of the SHMT (BLASTP 2.8.0+). Overlapping of the structures of (B) all dimer and (C) the

active site residues and PLP cofactor from both forms of the SHMT enzyme: *E.coli* (green) and *P. falciparum* (blue). 117

Figure 27. (Left) Structure of the SHMT dimer (subunit A: green; subunit B: blue) and licorice representation of the HL layer. (Right) The substrate, cofactors and the active site residues are represented in licorice and colored according to the subunit that they belong to (subunit A: green; subunit B: blue). Each molecule is identified by a circular tag. 119

Figure 28. Schematic representation of the two mechanistic proposals for the α-elimination of serine that is bonded to the PLP cofactor and forms the external aldimine, considering different protonation states of the Glu^{57} (Top: protonated and Bottom: unprotonated). The spheres represent the phosphate group of the PLP cofactor. 121

Figure 29. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the α-elimination reaction catalyzed by SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction. 123

Figure 30. Energy profile of the two evaluated mechanisms for the α-elimination process catalyzed by SHMT with Glu^{57} protonated (ΔE) (gray) or unprotonated (ΔG) (black). 124

Figure 31. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the three proton transfers catalyzed by SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction. 126

Figure 32. Energy profile for the dihedral rotation across the N1-Cβ bond. Newman representation was used to show the relative configuration of the Cβ and N1 atoms. The energies were computed using the ONIOM(DLPNO-CCSD(T)/CBS:ff99SB) scheme. 127

Figure 33. Structures of reactant (R), transition state (TS), and product (P) of the dehydration of the THF intermediate catalyzed by the SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped at the surface of the molecules involved in the reaction. 128

Figure 34. Structures of reactant (R), transition state (TS), and product (P) of the steps involved in cyclization of the THF intermediate catalyzed by the SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows corresponds to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction. 129

Figure 35. Structures of reactant (R), transition state (TS), and product (P) of the steps involved in cyclization of the THF intermediate in the cytoplasm. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. 131

Figure 36. Energy profile of the two proposals for the conversion of the THF intermediate into 5,10-methylene-THF: (black) the process is catalyzed by the SHMT enzyme or (gray) the reactions occur in the cytoplasm. The values in brackets correspond to the energy of each single step. 132

Figure 37. Structures of reactant (R), transition state (TS), and product (P) of the protonation of the quinonoid intermediate (QI) catalyzed by the SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequencies associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction. 133

Figure 38. Overview of the catalytic mechanism of the PLP-dependent enzyme SHMT. 134

Figure 39. Complete energy profile (DLPNO-CCSD(T)/CBS::ff99SB) for the catalysis of L-serine and THF into glycine and 5,10-methylene-THF by the SHMT. 136

Figure 40. Schematic representation of the behavior of the main groups and charges across the entire catalytic mechanism. 137

Figure 41. Structures of myriocin, sphingofungines E and F, and lactacystin.....140

Figure 42. Schematic representation of the mechanism for the IA and EA formation for PLP-dependent enzymes and the specific retroaldol cleavage reaction.....142

Figure 43. (Left) Secondary structure representation of the SHMT dimer (subunit A: red; subunit B: blue) and licorice representation of the HL layer. (Right) Licorice representation of the HL layer where the carbon atoms of the EA are colored green. 144

Figure 44. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the retroaldol cleavage catalyzed by the wildtype form of SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism. 147

Figure 45. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the retroaldol cleavage catalyzed by the Glu57Gln form of SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism. 149

Figure 46. Energy profile of the (Left) retroaldol cleavage and (Right) α-addition catalyzed by the WT (blue) and E57Q (red) mutated form of SHMT. The energy profile of the α-addition is the profile of the reverse reaction presented in the retroaldol cleavage profile. * Energies estimated by relaxed scan calculations. 150

Figure 47. Reactions catalyzed by CGL.....154

Figure 48. Map of amino acid residue of CGL for which there are mutagenesis studies. The residues are labelled in blue circles (dark: residues at the active site; light: residues far from the active site). Each reported mutation is shown with the respective ratio of enzymatic activity relative to the wild type form. [358, 359, 361] 157

Figure 49. Structure of the EA and Lys^{212} at the active site of CGL that minimizes the distances 1, 2, 3, 4, and 5 among all frames of the MD simulation.....159

Figure 50. (Left) Structure of chain A (red) and B (blue) of CGL, representation of the QM layer, and part of the 5 Å coat of water molecules. (Right) Structure of the EA and sidechains of the active site residues included in the QM layer.....160

Figure 51. Structures of reactant (R), transition state (TS), and product (P) of the two initial and sequential steps associated with the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism. 163

Figure 52. Scheme of the three mechanistic hypotheses tested for the catalytic mechanism of step 2. 163

Figure 53. Structures of reactant (R), transition state (TS), and product (P) of the step 3 of the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism. 165

Figure 54. Structures of reactant (R), transition state (TS), and product (P) of the step 4 of the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism. 166

Figure 55. Structures of reactant (R), transition state (TS), and product (P) of steps 5, 6, and 7 of the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism. 168

Figure 56. Key interactions of the EA with the active site of the HDC, SHMT, and CGL enzymes. (Bottom) Top view of the EA with labels for the angles of each atom/group

bonded to the Cα of the substrate. The angle is measured starting from the pyridine's plane. 172

Figure 57. Structural comparison of the active sites of HDC and AADC. The X-ray crystallographic structures of (wheat) HDC (PDB code 4E1O) and (green) AADC (PDB code 1JS6) were superimposed. The key active site residues are labelled and colored according to the respective structure. 175

Figure 58. Reactions catalyzed by XO: (1) oxidation of HPA to XAN, and (2) oxidation of XAN to URC. 183

Figure 59. Proposals of Yamaguchi Y. (2007) [72] and Metz and Thiel (2009) [73, 74] for the catalytic mechanism of XO. 184

Figure 60. Overlapping of the structures of the active site residues of bovine and human forms of xanthine oxidase. 185

Figure 61. Representation of the XO structure with the three regions modeled by homology highlighted in red. The active site residues and pterin cofactor are represented in licorice colored by yellow. 186

Figure 62. Representation of all the atoms included in the high-level (HL) layer treated using quantum mechanics (licorice colored by element). The residues that are up to 5 Å from the HL layer were represented in grey lines, and the charged residues within that shell are highlighted in blue. A 180 ° rotation across the z-axis was applied from the left to the right images. 187

Figure 63. Wedge-dash representation of XAN and Moco molecules. 189

Figure 64. Tridimensional representation of the active site residues, Moco, and XAN molecule at the active site of XO in the TS of step 1. The red arrows represent the vectors assigned to the atoms' vibrations associated with the imaginary frequency of the TS structure. Relevant interatomic distances for reactant, TS, and product are labeled. Wedge-dash representation of the reaction and product of the reaction. The activation and reaction Gibbs free energies for this step, and imaginary frequency assigned to the TS are shown. 190

Figure 65. Electrostatic potential mapped in the surface of the molecules involved in the reaction of steps 1 and 2. 190

Figure 66. Tridimensional representation of the active site residues, Moco, and XAN molecule at the active site of XO in the TS of step 2. The red arrows represent the vectors assigned to the atoms' vibrations associated with the imaginary frequency of the TS structure. Relevant interatomic distances for reactant, TS, and product are labeled. Wedge-dash representation of the reaction and product of the reaction. The activation and reaction Gibbs free energies for this step, and imaginary frequency assigned to the TS are shown. 192

Figure 67. Representation of the two hypotheses for the catalytic mechanism of step 3. (Red) Protonation of URC molecule by (Red) Arg880 residue or (Green) Glu^{1261} residue. 192

Figure 68. Tridimensional representation of the active site residues, Moco, and URC molecule at the active site of XO in the TS of step 3. The red arrows represent the vectors assigned to the atoms' vibrations associated with the imaginary frequency of the TS structure. Relevant interatomic distances for reactant, TS, and product are labeled. Wedge-dash representation of the reaction and product of the reaction. The activation and reaction Gibbs free energies for this step, and imaginary frequency assigned to the TS are shown. 193

Figure 69. Tridimensional representation of the active site residues, the two [2Fe-2S] clusters, and the FAD cofactor. Scheme of the oxidation of the Mo ion, the consequent flow of electrons through the [2Fe-2S] clusters, and reduction of the FAD cofactor into $FADH_2$. 194

Figure 70. Tridimensional representation of the Moco, Glu^{1261}, Arg^{880}, and water molecules filling the protein tunnel from the solvent to the bottom region of the active site obtained from the MD simulations. 195

Figure 71. Wedge-dash representation of the reaction and product of step 4, and consequent deprotonation of the water molecule closing the enzymatic cycle. The reaction Gibbs free energy for this step is shown. 196

Figure 72. Complete energy profile (B3LYP/6-311++G(3df,3pd),LanL2DZ:ff99SB) for the catalysis of the oxidation of XAN into URC by XO. 197

Figure 73. Overview of the catalytic mechanism of the XAN oxidation into URC by XO. 198

Figure 74. Schematic representation of the catalytic cycle of XO under (green) normal and (red) stress conditions. 202

Figure 75. Schematic representation of the mechanism of reduction of nitrate to nitrite by XO. The wedge-dash schemes illustrate the proposed mechanism based on the calculations of this work. The computed Gibbs-free energies and an illustration of the TS of step 2 are also included. 206

Figure 76. Schematic representation of the mechanism of reduction of nitrite to nitric oxide by XO. Each stage corresponds to a single-electron reduction of nitrite to nitric oxide. The wedge-dash schemes illustrate the proposed mechanism based on the calculations of this work. The computed Gibbs-free energies and an illustration of the TS of step 2 are also included. Colored circles enclose the total charge of the HL and different colors correspond to different spin multiplicities (1, 2, and 3). 209

Figure 77. Energy profile for the PES scan for the homolytic rupture of O-NO bond of nitrite (Step 2 of the 1st Cycle), on the singlet and triplet stages. 209

Figure 78. Complete energy profile (B3LYP/6-311++G(3df,3pd)//B3LYP/6-31G(d,p),LanL2DZ:ff99SB) for the reduction of (Top) nitrate into nitrite, and (Bottom) nitrite into nitric oxide by XO. 210

Figure 79. Overview of the (green arrows) natural catalytic mechanism of XO and the (red arrows) catalytic mechanism for the reduction of nitrate and nitrite into nitrite and nitric oxide, respectively. 212

Figure 80. The complete genome sequence of SARS-CoV-2 virus. [397] Each red label (A-K) represents the respective protein cleavage regions catalyzed by M^{pro}. Each cleavage region is also shown with associated flanking amino acid sequences. 217

Figure 81. The general reaction catalyzed by M^{pro}. “R1” and “R2” represent the remaining part of the substrate (peptide). “X” represents the sidechain of the possible amino acid residues (Ser, Ala, or Asn). 218

Figure 82. Representation of all the 4 steps of the catalytic mechanism of SARS-CoV-2 M^{pro}. The activation and reaction Gibbs free energies are expressed in kcal/mol units. The eigenvectors of each TS structures are shown as red arrows. 224

Figure 83. Charge fluctuation of the key atoms over the catalytic mechanism of the M^{pro}. Circles, squares, and lozenges identify atoms belonging to the catalytic dyad, the substrate, and the water molecule, respectively. Markers are colored by element: nitrogen (blue), sulfur (yellow), hydrogen (white), carbon (orange), and oxygen (red). 225

Figure 84. Complete energy profile for the catalytic mechanism of SARS-CoV-2 M^{pro}. The values placed between parenthesis correspond to the activation/reaction energies of each step. 226

Figure 85. The main window of molUP extension on the VMD Store. On the left panel, a list of all available plugins is presented, organized by category. In the main section, the VMD Store provides information about the plugin’s name, latest version, description, screenshots, and citation. Moreover, there are buttons to install or uninstall the plugin, as well as links for the citable paper and its repository on GitHub. 234

Figure 86. Statistical data regarding the number of downloads of the VMD Store during the time period from February 2018 to August 2020. 236

Figure 87. Screenshot showing the ToolBar (left), the VMD OpenGL Display window (middle), and the VMD Main window (right). The structure of Nitrate Reductase (PDB code: 2NAP) [460] was loaded as an example. 240

Figure 88. Description of the functions associated with all elements available in the ToolBar. The “NEW” tag identifies all the new features added by the ToolBar to the VMD’s experience. All other features were already available in VMD, but they were included in the Toolbar, in a pleasant and more accessible manner. 240

Figure 89. Screenshot showing a nitrogen atom picked using the ToolBar. The ToolBar displays information about the atom that was picked, namely the element, atom’s *name*, *chain*, *ResName*, *ResID*, and *index*. 241

Figure 90. Screenshot of the “Presets” menu available on ToolBar. The “Presets” menu provides options to modify the appearance of the molecular representation in VMD. 242

Figure 91. Screenshot showing the bond length editor provided by the ToolBar. Since the tool is activated, the user picks two atoms consecutively to enable the “Bond Editor” window, where the distance between those atoms can be adjusted. 243

Figure 92. Screenshot of the main GUI of molUP. The main areas of the GUI are tagged. 248

Figure 93. Flowchart of the common strategy used to carry on a QM/MM study using Gaussian software, VMD, and molUP. 250

Figure 94. molUP guide to load a Gaussian-related file using the molUP GUI. 251

Figure 95. molUP guide to adjust the bond length between two atoms. 252

Figure 96. Screenshot of the “Layer”, “Freeze”, and “Charges” subtabs of the “Model” section of the molUP GUI. 253

Figure 97. Screenshot of the “Input” section of molUP and the tools available on the dropdown menu of each block. 253

Figure 98. molUP guide to save Gaussian input file using molUP. Alternative file formats and advanced options are shown on the right. 255

Figure 99. Screenshot of different layout of the “Energies” subtab of the “Output” section of molUP. In the case of multiple structures (right), the plot’s adjustments are shown. 256

Figure 100. (Left) Screenshot of the “Frequencies” subtab of the “Output” section of molUP. (Right) Representation of the vibrational frequency animation and related vectors. 257

Figure 101. Example of a hypothetical molecular system split into two different fragments. The circles and lines represent atoms and covalent bonds, respectively. 267

Figure 102. Screenshot of the Energy Split extension for VMD. The labeled numbers and associated text provide quick instruction on how to use the extension. 269

Figure 103. Screenshot of the Energy Split extension for VMD showing a system split into three fragments. 269

Figure 104. Screenshot of the main GUI of chemTool. The main areas of the GUI are tagged. 279

Figure 105. chemTool guide to load ORCA-related files. 281

Figure 106. Adjusting a dihedral torsion using the “Dihedral Angle Editor” tool available on chemTool. 281

Figure 107. Screenshot of the “QM | QM/MM”, “Fixed Atoms”, and “Charges” tabs of the main “Model” section of the chemTool. 283

Figure 108. Screenshot of the “Input” main section of the chemTool. 284

Figure 109. chemTool guide to save an ORCA input file storing the atomic coordinated in a PDB file. 285

Figure 110. Screenshot of the “Energy” tab of the chemTool. 285

Figure 111. Screenshot of the analysis of vibrational frequencies using chemTool. ..286

Figure 112. Screenshot of the energy profile analysis conducted using chemTool. ...287

Figure 113. 3D representation of the structure of a DNA sequence in augmented reality using BioSIM^{AR}. 295

Figure 114. Example of a printed card showing the pattern code to observe the 3D structure of an RNA molecule using the BioSIM^{AR}. 296

Figure 115. Overview of the workflow involved in the (yellow) development of BioSIM^{AR}, and the processes taken when users point their devices' cameras to a 2D pattern (blue). 298

Figure 116. Tukey test results for students regarding the use of BioSIM^{AR} (n=26). Scores on the scale have a range of 1-5. The boxes correspond to the first quartile (left), median (center), and third quartile (right). The whiskers correspond to the lowest (left) and maximum (right) values observed. The blue dots represent outliers. 301

Figure 117. Comparing the structure of (left) D-Alanine and (right) L-Alanine, highlighting the chiral α-carbon. 303

Figure 118. Photography of the Protein Wars arcade machine at "Mostra da Universidade do Porto". 305

Figure 119. Screenshot of Protein Wars at level 1. 308

Figure 120. Screenshot of part of the initial message displayed for level 1 of Protein Wars. 309

Figure 121. Photos of the process of building the arcade-style game machine for Protein Wars. 310

Figure 122. Screenshots of the first three web pages that are shown on Criminal Investigation. The third page provides details about the actual crime, including an interactive map with key locations with valuable information. 315

Figure 123. Screenshot of the web platform showing the evidence bar code associated with each evidence. 316

Figure 124. Screenshots of all the pages showed on the web platform with instructions about how to perform a chromatography test to acquit one of the suspects. 317

Figure 125. Screenshots of all the pages showed on the web platform with instructions about how to test the pH of different soil samples to find the felon of this "crime". 318

Figure 126. The Biofilms Structural Database (BSD). (A) BSD main page. (B) List of database entries in BSD. (C) Entries by structure for each Protein Data Bank (PDB) code with an indication of the main information available for each protein/structure and interactive 3D representation of the protein. (D) Entries by ligand with main information and interactive 3D representation. 327

Index of Equations

Equation 1. Total energy function for a system treated with MM. 31

Equation 2. General energy function for bonded terms of AMBER force field. 31

Equation 3. Energy function for bond terms. k_b: force constant; l: bond length; l_0: bond length at the equilibrium 32

Equation 4. Energy function for angle terms. k_a: force constant; θ: bond length; θ_0: bond length at the equilibrium 32

Equation 5. Energy function for dihedral angle terms. k_d: force constant; ρ: dihedral angle; γ: dihedral angle for which the energy is minimum; n: number of minima of the energy function.....32

Equation 6. Energy function for improper torsion terms. ki: force constant; φ: improper angle torsion; φ0: equilibrium value for the improper torsion 32

Equation 7. Energy function for non-bonded interaction. ϵ: dieletric constant; qi and qj: charge of atom i and j, respectively; ri,j: distance between atoms i and j; $\epsilon i,j$ and $\sigma i,j$: parameters for the interaction between the atoms i and j 33

Equation 8. Determination of $\epsilon i,j$ parameter for atoms i and j. 33

Equation 9. Determination of $\sigma i,j$ parameter for atoms i and j.....33

Equation 10. Second order derivative of the Newton's Second Law. 36

Equation 11. Mathematical relation between potential energy and atomic positions. ...36

Equation 12. Function to determine the position of an atom i at time $t1$. 37

Equation 13. Function to determine the velocity of an atom i at time $t1$. 37

Equation 14. Pressure of a system according to the Berendsen barostat. Pt: pressure of the system; $nparticles$**:** total number of particles; $Vcell$: volume of the unit cell; kB: Boltzmann constant; Tt: temperature of the system; $\boldsymbol{Fij}t$: force that atom j exerts on atom i; $\boldsymbol{rij}(t)$: distance between atoms i and j 39

Equation 15. Friction coefficient (μP) as a function of pressure variation. τP: relaxation time constant; Δt: integration time step.....39

Equation 16. Langevin temperature control added to the Newton equation. γ: average number of collisions required to damp the highest vibration in about $e-1$ 39

Equation 17. Relation between the frequency and the wavelength of a wave in vacuum. 44

Equation 18. Energy of a photon. h: Plank's constant; v: radiation's frequency 44

Equation 19. Relation between the energy of an electron after the incidence of a photon. 45

Equation 20. L. de Broglie's equation. 45

Equation 21. The uncertainty principle. 46

Equation 22. Schrödinger equation for one particle and one dimension. m: mass of the particle; $V(x,t)$: potential-energy function of the system 46

Equation 23. General Schrödinger equation. 47

Equation 24. General time-independent Schrödinger equation. 47

Equation 25. Born-Oppenheimer approximation. 48

Equation 26. Born-Oppenheimer approximation as a sum of Hamiltonians. 48

Equation 27. The electronic Hamiltonian. 48

Equation 28. The nuclear Hamiltonian. 48

Equation 29. STO equation. N: normalization constant; r, θ, φ: spherical coordinates .49

Equation 30. GTO equation. N: normalization constant; rx, ry, rz: cartesian coordinates; i, j, k: non-negative integers; α: positive exponent; r: distance from the electron to the nucleus 49

Equation 31. Hartree-Fock method. me: electron's mass; Z: effective nuclear charge; ri: distance of electron *i* to the nucleus; rij: distance between electrons *i* and *j* 53

Equation 32. Wave function as the product of one-electron wave functions. 53

Equation 33. One-electron spin orbital function. 54

Equation 34. HF function as a product of n spin orbital functions. 54

Equation 35. Slater determinant. 54

Equation 36. Variational principle. 55

Equation 37. Exact ground-state wave function. cX: expansion coefficients; a, b, c ...: limits in the summation indices 58

Equation 38. Zero-order Hamiltonian of the HF-SCF method. i: electron; Ne: number of electrons of the system 61

Equation 39. Perturbation contribution. 61

Equation 40. Perturbation contribution for electron i. j and i: electrons; m: occupied molecular orbitals; Jm: Coulomb operator; Km: exchange operator 61

Equation 41. Wave function as a function of the HF wave function. 62

Equation 42. Exponential operator. 62

Equation 43. Excitation operator. 62

Equation 44. Electron density at the nucleus. $\rho(rA)$: spherically averaged density.....64

Equation 45. Born-Oppenheimer approximation applied to DFT method. 65

Equation 46. Potential energy nucleus-electron according to DFT method. M: total number of atoms of the system; ZA: atomic charge of atom A 65

Equation 47. Potential energy electron-electron according to DFT method. 65

Equation 48. Hamiltonian for the reference system. 66

Equation 49. Kohn-Sham orbitals. 66

Equation 50. Exchange-correlation energy. 67

Equation 51. Exchange energy. 68

Equation 52. Correlation energy. 68

Equation 53. Spin polarization. 68

Equation 54. Volume containing one electron. 68

Equation 55. Exchange energy based on LDA. $\varepsilon XLDA$: exchange density for a given particle 69

Equation 56. Exchange energy based on LSDA. 69

Equation 57. Correlation energy based on PW method. $x = r$; r: electron position; $a, \alpha, \beta 1, \beta 2, \beta 3, \beta 4$: empirical parameters 69

Equation 58. Exchange energy according to B method. β: fitting parameter 70

Equation 59. Correlation energy according LYP method. a, b, c, d: fitting parameters ..70

Equation 60. Exchange-correlation energy of a H+H method. 71

Equation 61. Exchange-correlation energy of the B3LYP functional. a, b, c: empirical parameters obtained by fitting experimental data 71

Equation 62. ZPE correction. 73

Equation 63. Energy at temperature T. 73

Equation 64. Frequency of a harmonic oscillator. m: mass of the particle; k: force constant. 74

Equation 65. Enthalpy contribution estimated by thermal corrections. kB: Boltzmann's constant; T: temperature 75

Equation 66. Gibbs free energy. 75

Equation 67. Entropy contribution estimated by thermal corrections. 75

Equation 68. Energy function for a hybrid additive method. $EQMModel$: energy of the model system calculated using QM; $EMMSurround.$: energy of the surrounding system calculated using MM; $EInteraction$: energy of interaction between the two layers 77

Equation 69. Energy function for a hybrid substrative method (ONIOM). 77

Equation 70. Arrhenius equation. 83

Equation 71. Eyring equation. κ: transmission coefficient; kB: Boltzmann's constant; h: Plank's constant 83

Equation 72. Energy of the system according to AMBER force field. 264

Equation 73. Distance between two atoms ij. 264

Equation 74. Angle between two vectors ji and jk. 264

Equation 75. Calculation of the Aij parameter. 266

Equation 76. Calculation of the Bij parameter. 266

Equation 77. Calculation of the εij parameter. 266

Equation 78. Calculation of the Dij parameter. 266

Equation 79. Total energy of the system according to Energy Split. 268

Index of Tables

Table 1. Experimental kinetic data available for the reactions catalyzed by SHMT. 141

Table 2. Compatibility Analysis for BioSIM^{AR} running in the different web browsers and operating systems. 297

This page was intentionally left blank.

List of Abbreviations

2D	Bidimensional
3D	Tridimensional
AADC	Aromatic L-Amino Acid Decarboxylase
ACS	American Chemical Society
ADI	Arginine Deiminase
adoMet	*S*-Adenosylmethionine
AIDS	Acquired Immunodeficiency Syndrome
API	Application Programming Interface
AR	Augmented Reality
ARGase	Arginase
ASNase	Asparaginase
BRENDA	Braunschweig Enzyme Database
BSD	Biofilms Structural Database
CASSCF	Complete Active-Space Self-Consistent Field
CBS	Complete Basis Set
CC	Coupled-Cluster
CCD	Double Excitation Determinants
CCSD	Single and Double Excitation Determinants
CCSD(T)	Single, Double, and Triplet Excitation Determinants
CGL	Cystathionine-γ-Lyase
cGTO	Contracted Gaussian Type Orbitals
CI	Configuration Interaction
CPU	Central Processing Unit
cryo-EM	Cryogenic Electron Microscopy
CSF	Configuration State Function
cSHMT	Cytosolic Serine Hydroxymethyltransferase

CSS	Cascading Style Sheets
CSV	Comma-Separated Values
DCI	Doubly Excited States
DFT	Density Functional Theory
DLPNO	Domain-based Local Pair Natural Orbital
DNA	Deoxyribonucleic Acid
DOI	Digital Object Identifier
DOPA	Levodopa or L-3,4-Dihydroxyphenylalanine
dTMP	Deoxythymidine Monophosphate
dUMP	Deoxy-Uridine Monophosphate
EA	External Aldimine
EC	Enzyme Commission
ECP	Effective Core Potential
EPR	Electron Paramagnetic Resonance
EPS	Extracellular Polymeric Substance
ExPASy	Expert Protein Analysis System
FAD	Flavin Adenine Dinucleotide
FDA	Food and Drug Administration
FMN	Flavin Mononucleotide
GGA	Generalized Gradient Approximation
GPCR	G-Coupled Protein Receptor
GPU	Graphics Processing Unit
GTO	Gaussian Type Orbitals
GUI	Graphical-User-Interface
HDC	Histidine Decarboxylase
HF	Hartree-Fock
HIV	Human Immunodeficiency Virus
HL	High-Level

HTML	HyperText Markup Language
IA	Internal Aldimine
IRC	Intrinsic Reaction Coordinate
ISI	Institute for Scientific Information
IUBMB	Union of Biochemistry and Molecular Biology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
LL	Low-Level
LSDA	Local Spin Density Approximation
MBPT	Many-Body Perturbation Theory
MCSCF	Multiconfigurational Self-Consistent Field
MD	Molecular Dynamics
MDF	Medium-Density Fiberboard
Met Synthase	Met Synthase
MGL	Methionine-γ-Lyase
MM	Molecular Mechanics
MP	Møller-Plesset
MPPT	Møller-Plesset Perturbation Theory
M^{pro}	Main Protease of SARS-CoV-2
mSHMT	Mitochondrial Serine Hydroxymethyltransferase
MTHFR	Methylenetetrahydrofolatereductase
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NMR	Nuclear Magnetic Resonance
NO	Nitric Oxide
ONIOM	Own N-layer Integrated molecular Orbital molecular Mechanics
OpenGL	Open Graphics Library

PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PES	Potential Energy Surface
PLK	Pyridoxal/Pyridoxine Kinase
PLP	Pyridoxal-5'-Phosphate
PNPOx	Pyridoxine/Pyridoxamine-5′-Phosphate 5′-Phosphate Oxidase
Post-HF	Post-Hartree-Fock
PP	Pseudo-Potential
PT	Perturbation Theory
PW	Perdew and Wang method
QM	Quantum Mechanics
QR	Quick Response
QS	Quorum Sensing
QSAR	Quantitative Structure-Activity Relationship
RCSB	Research Collaboratory for Structural Bioinformatics
RNA	Ribonucleic Acid
RTP	Rádio Televisão Portuguesa
RubisCO	Ribulose-1,5-Bisphosphate Carboxylase-Oxygenase
SCF	Self-Consistent Field
SDCI	Singly and Doubly Excited States
SHMT	Serine Hydroxymethyltransferase
SMILES	Simplified Molecular-Input Line-Entry System
SP	Single Point
STO	Slater Type Orbitals
Tcl	Tool Command Language
THF	Tetrahydrofolate
TPP	Thiamine Pyrophosphate
TS	Transition-state

URC	Uric Acid
VMD	Visual Molecular Dynamics
VR	Virtual Reality
XAN	Xanthine
XD	Xanthine dehydrogenase
XO	Xanthine Oxidase
ZPE	Zero-Point Energy

This page was intentionally left blank.

A. Introduction

So many of the chemical reactions occurring in living systems have been shown to be catalytic processes occurring isothermally on the surface of specific proteins, referred to as enzymes, that it seems fairly safe to assume that all are of this nature and that the proteins are the necessary basis for carrying out the processes that we call life.

John Desmond Bernal

This page was intentionally left blank.

This first chapter provides an overview of enzymes and their importance in different fields. Moreover, it also highlights the importance of using computational simulations to obtain valuable insights into the enzyme's structural and functional features. Since the main goal of this Ph.D. project was studying the chemistry associated with pyridoxal-5'-phosphate (PLP) dependent enzymes, a particular attention is given to this cofactor and respective enzymes.

1. Enzymes

Life is possible because of two fundamental conditions. Every living organism must be able to self-replicate, and they also need to catalyze chemical reactions efficiently and selectively [1]. All forms of life use enzymes to increase the rate of chemical reactions to make them fast enough to be compatible with life. We can think about digestion as a good example of the importance of enzymes in humans. When we eat some kind of food, our digestive system provides different types of enzymes that will digest our food into different components. This means that enzymes will efficiently break the structure of such molecules into smaller pieces that will eventually reach our bloodstream. Without enzymes, most of the essential reactions would take days, months, years, or even millions of years to occur [2].

Since enzymes are proteins capable of catalyzing a chemical reaction, they are considered biological catalysts. A catalyst is, by definition, a chemical/biochemical species that increases the rate of a chemical reaction, but it is not consumed [2]. Therefore, the catalyst can be used over and over again to make reactions faster.

The evolution of living creatures over millions of years originated high specific and efficient enzymatic catalysts, which are among the most incredible catalysts available in nature. Enzymes can greatly increase the rate of a panoply of different reaction types in aqueous solution under mild pH and temperature conditions. There are very few non-biological catalysts capable of achieving such perform in similar conditions [1].

The importance of enzymes in life is corroborated by the fact that more than 99% of all biochemical reactions are catalyzed by them. Only less than 1% are reactions non-catalyzed or catalyzed by RNA entities (ribozymes) [3].

Ribozymes are considered enzymes, which means that not all enzymes are proteins. Only few classes of RNA molecules possess the capability to catalyze chemical reactions. Most part of enzymes consists of one or more peptide chains.

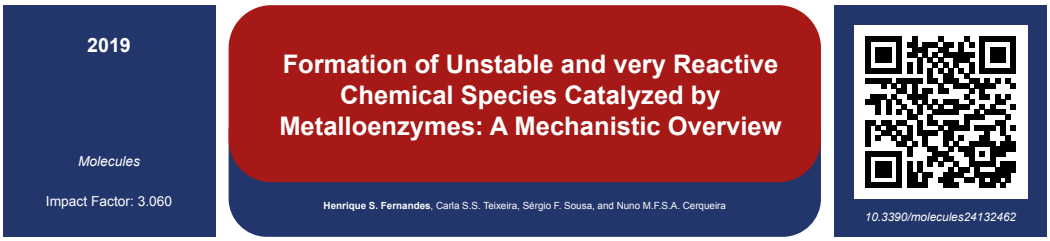
The molecular weights of an enzymatic system can range from a few kDa to one million kDa [1]. Among all the amino acid residues of an enzyme, only some of them are directly

related to its specificity and catalytic activity. However, a major part of them are essential for the correct protein folding and the enzymatic activity depends on the protein's primary, secondary, tertiary, and quaternary structural organizations.

Even though the catalytic activity of some enzymes relies on their own amino acid residues, there are several enzymes that need "assistance". In such cases, enzymes can incorporate a cofactor to assist them in their catalytic function.

A cofactor is a non-protein molecule that interacts with enzymes to help them in the transport of substrates, intermediates, and/or products or even step in the catalytic process. The protein structure of an enzyme that requires a cofactor is called apoenzyme since it corresponds to the enzyme without the cofactor. On the other hand, the complete enzyme with the required cofactor(s) is designated holoenzyme [1, 4].

Enzyme cofactors can be classified in metal ions (inorganic cofactors) or organic cofactors (also known as coenzymes) in terms of their chemical nature. Several enzymes employ metal ions, including transition metals, as cofactors. Mn, Mg, Na, K, Co, Cu, Mo, Ca, Zn, Ni and Fe are examples of the most commonly found metal ions in natural enzymes [5, 6]. In the course of this thesis, a review about how metalloenzymes catalyze the formation of very reactive chemical species was written (Appendix: page 393).



[1]

Regarding organic cofactors, there is also a panoply of different molecules playing a crucial role as cofactors, such as pyridoxal-5'-phosphate (PLP), tetrahydrofolate (THF), flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), thiamine pyrophosphate (TPP), coenzyme A, nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADPH), biotin, heme, cobamide, lipoate, among others [1, 4].

Beyond this classification of the cofactors, these molecular entities can also be classified as prosthetic groups. A prosthetic group is a cofactor that is tightly bonded to the enzyme.

[1] This publication was chosen for the cover of the Molecules issue where it was published. (Appendix: page 391)

In some cases, the prosthetic group can be even covalently bonded to at least one amino acid residue. [1]

Beyond the enzyme classification in terms of the presence or absence of a cofactor and its respective type, the most used classification system for enzymes regards their activities.

The International Union of Biochemistry and Molecular Biology (IUBMB) developed a classification system based on the enzymatic activity, and it is called “Enzyme Commission”, commonly referred as EC number. The EC number consists of a sequence of 4 digits, separated by points, which organize enzymes from the most general group (first digit) to the most specific one (fourth digit). This classification system considers six major groups of enzymes [1, 4]:

- EC 1: Oxidoreductases catalyze oxidation/reduction reactions;
- EC 2: Transferases catalyze the transfer of a functional group between two chemical species;
- EC 3: Hydrolases catalyze the cleavage of different types of covalent bonds through the attack of a water molecule;
- EC 4: Lyases catalyze the cleavage of other bonds by processes distinct from oxidation or hydrolysis;
- EC 5: Isomerases catalyze the isomerization of a single molecule;
- EC 6: Ligases catalyze the join of two molecules by one or more covalent bonds.

Although this classification system for enzymes is not perfect [4], it has been widely used to provide an identification number for all known enzymes. This unique identification for each enzyme helps to overcome the issues caused by enzymes that are known by different names. For example, aromatic L-amino acid decarboxylase (AADC) is also known as levodopa (DOPA) decarboxylase, tryptophan decarboxylase, and 5-hydroxytryptophan decarboxylase, but the EC number is 4.1.1.28.

During the course of this Ph.D. project, the catalytic mechanism of enzymes belonging to the four first EC groups were studied: xanthine oxidase (XO) (EC 1.17.3.2), serine hydroxymethyltransferase (SHMT) (EC 2.1.2.1), main protease of SARS-CoV-2 (M^{pro}) (EC 3.4.22.69), histidine decarboxylase (HDC) (EC 4.1.1.22), and cystathionine-γ-lyase (CGL) (EC 4.4.1.1).

As previously described, enzymes act as biological catalysts lowering the activation Gibbs free energy ($\Delta G^{\ddagger}$) of the global reaction. Consequently, the reaction rate increases,

and the number of enzymatic turnovers per unit of time increases. In addition, specificity is also one important feature of enzymes. While in the laboratory, reactions can be conducted in separated reactors, in cells, a lot of different reactions occur simultaneously in the same compartment. Therefore, enzymes evolved to become highly specific for their target substrates when required. Glucose oxidase [7] and glucokinase are two examples of highly specific enzymes that can only catalyze reactions involving glucose as substrate. [4] In some cases, enzymes evolved to be specific enough to distinguish two molecules that differ only on one atom or two isomers. Ribulose-1,5-bisphosphate carboxylase-oxygenase (RubisCO) can be three times more specific to carbon dioxide than dioxygen. Even though this enzyme tends to be promiscuous for both substrates, RubisCO evolved to become more selective (3:1) to carbon dioxide even though it has compromised its catalytic efficiency [8].

On the other hand, other enzymes can accept a variety of substrates. The AADC can catalyze the decarboxylation of different substrates under biological conditions, namely DOPA, phenylalanine, tyrosine, histidine, tryptophan, and 5-hydroxytryptophan [9, 10].

In sum, enzymes can present different levels of selectivity, from discriminating between different stereoisomers of a certain substrate to accept different molecules as substrate [4]. In addition, the promiscuity of certain enzymes can lead them to catalyze totally different reactions under different conditions or acting on different substrates [1, 4].

1.1. Coenzyme-dependent enzymes

As described above, enzymes can employ organic (coenzymes) and/or inorganic (metal ions) cofactors in their active sites. Cofactors can simply play a structural role where they help to correctly place other chemical species within the active site. In addition, they can act directly in the chemical reaction. The PLP and THF coenzymes and the Mo ion are examples of cofactors that are directly involved in the catalysis of some enzymes, namely the ones studied in this project.

Therefore, the following sections of this chapter are focused on these three cofactors: PLP, THF, and Mo ion.

1.1.1. Pyridoxal-5'-Phosphate (PLP)

PLP is one of the catalytically active forms of vitamin B_6 [11] (Figure 1). It is employed as a cofactor by several distinct enzymes accounting for more than 140 distinct enzymatic activities [12].

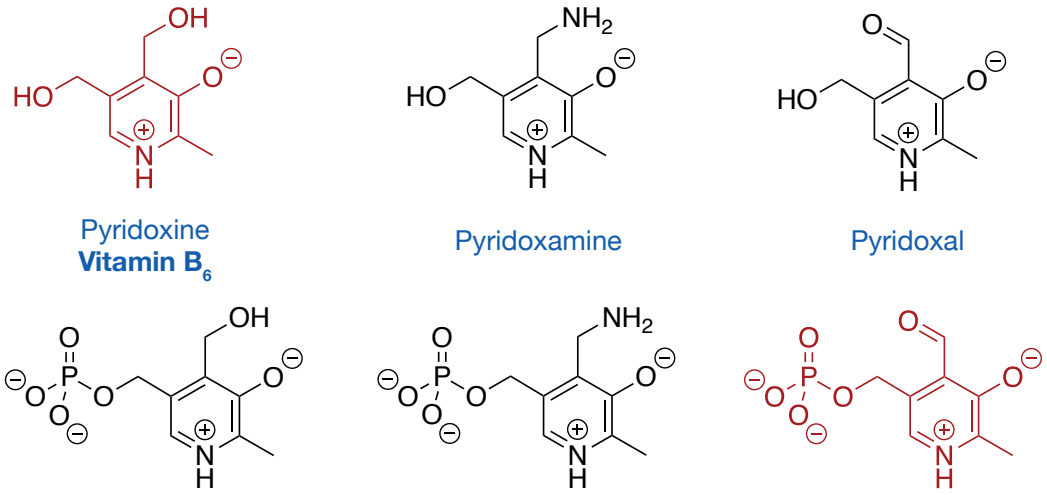


Figure 1. Structures of pyridoxine (vitamin B_6), pyridoxamine, pyridoxal, pyridoxine-5'-phosphate, pyridoxamine-5'-phosphate, and pyridoxal-5'-phosphate.

The BRENDA database [13] has 8083 entries[2] corresponding to all already discovered enzymes (and included in the database). Among all those enzymes, 224 enzymes use PLP as a cofactor, accounting for almost 3% of all known enzymes. Considering the EC

[2] Access date: 2021, January 14th

classification, there are PLP-dependent enzymes belonging to the first five of the six enzyme classes (Figure 2). So far, there are no PLP-dependent enzymes that naturally catalyze a ligase reaction (EC 6).

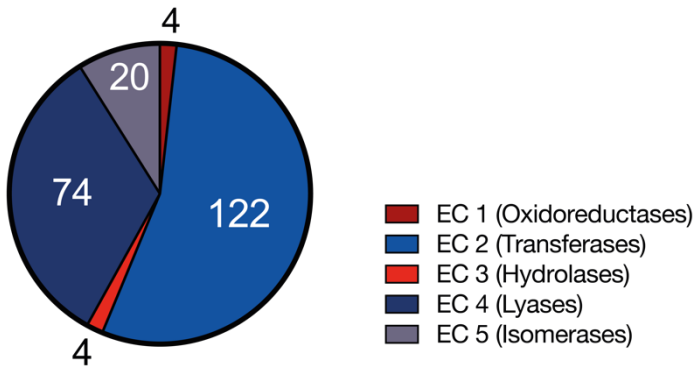


Figure 2. Distribution of all already known PLP-dependent enzymes over the different enzyme classes. [13]

Although PLP is essential for the life of all organisms, only microorganisms and plants are able to synthesize PLP *de novo* [11]. Animals are auxotroph for this cofactor and need to obtain vitamin B_6 from their diet. Then, the pyridoxal/pyridoxine kinase (PLK) and pyridoxine/pyridoxamine-5′-phosphate 5′-phosphate oxidase (PNPOx) enzymes catalyze the conversion of vitamin B_6 into PLP (Figure 3) [11, 14].

Most PLP-dependent enzymes are involved in the synthesis, interconversion, and degradation of amino acids. Nonetheless, some of them play key roles in the metabolism of neurotransmitters, one-carbon units, biogenic amines, tetrapyrrolic compounds, amino sugars, modulation of steroid receptor-mediated gene expression, and regulation of immune function [11].

The wide range of action of PLP-dependent enzymes makes living beings highly reliant on a reasonable pool of PLP within their cells. Fortunately, cases of PLP deficiency are very rare since most of the dietary sources contain vitamin B_6. Consequently, the reported cases of diseases caused by PLP deficiency are usually associated with the malfunction of PLK and PNPOx enzymes. These two enzymes are involved in the conversion of vitamin B_6 into its catalytic active form (PLP) [15-18]. In addition, there are reports of some cases of inhibition of PLP-dependent enzymes by the consumption of certain drugs [19-24].

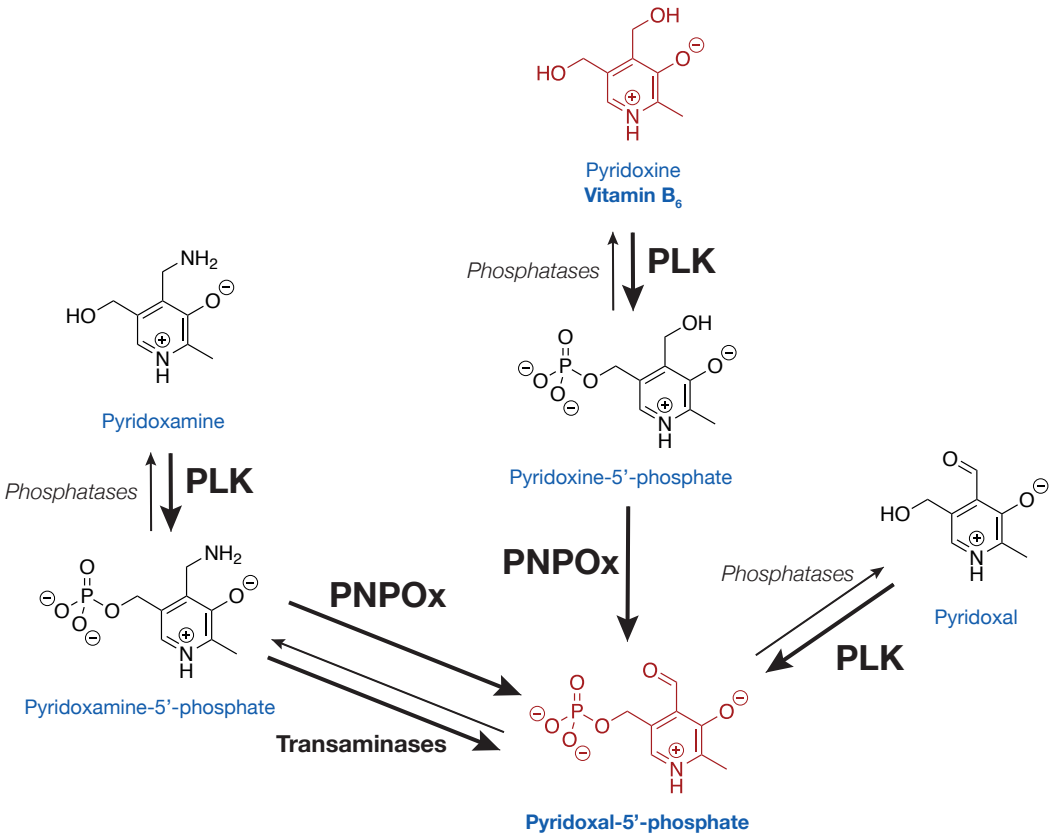


Figure 3. PLP salvage pathway present in all living organisms. (Adapted from [11])

PLP is absolutely crucial for all living organisms, but the pool of free PLP is maintained at a very low level in organisms. Since PLP has a very reactive aldehyde group, it can easily react with primary and secondary amines causing toxicity. For this reason, the pool of free PLP is kept low, and most PLP molecules are within the active site of several enzymes all over the organism [11, 25].

a. PLP-dependent enzymes

PLP is one of the most common organic cofactors in enzymes, and it is associated with the catalysis of a panoply of different types of reactions. As already mentioned, the reactive aldehyde group of PLP can easily react with primary and secondary amines. Therefore, PLP can easily react with amino acids through the amine group, and the enzymes that use it as a cofactor catalyze a considerable number of different reactions involving amino acids as substrate.

PLP-dependent enzymes are known to be able to catalyze six different types of reactions involving amino acids (Figure 4).

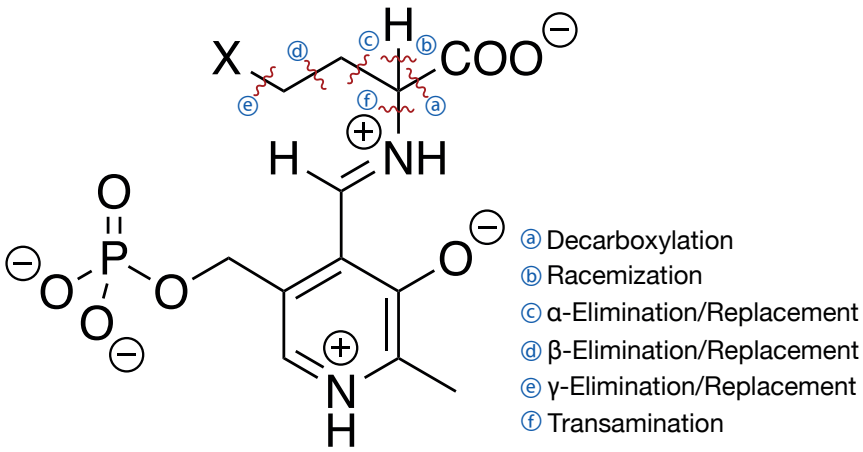


Figure 4. Overview of all types of reactions catalyzed by PLP-dependent enzymes.

The decarboxylation involves the cleavage between the α-carbon and the substrate's carboxyl carbon (Figure 4 a). This reaction involves the release of one carbon dioxide molecule, and then, the resulting intermediate is protonated in the α-carbon. Currently, 23 different PLP-dependent decarboxylases are known, including HDC and AADC. In this work, the catalytic mechanism of HDC was studied using a computational approach (Chapter C – 1.1. page 89), aiming an atomistic detailed description of the mechanisms of PLP-dependent decarboxylases.

PLP-dependent enzymes like alanine or aspartate racemases can invert the chiral center of L-alanine and L-aspartate, converting them into D-alanine and D-aspartate, respectively. The reaction is accomplished through the abstraction of the proton bonded to α-carbon, followed by the re-protonation from the opposite side of the molecule. Consequently, the chiral center at the α-carbon is inverted (Figure 4 b).

PLP-dependent enzymes can also catalyze the elimination or substitution of chemical groups at the α, β, and γ positions of the amino acids' sidechain (Figure 4 c d e). These reactions are catalyzed following different mechanisms depending on the substrate and the elimination site. In this Ph.D. project, two different mechanisms were studied for the α-elimination and α-addition catalyzed by SHMT (Chapter C – 1.2. page 111 and Chapter C – 1.3. page 139). It was also studied the catalytic mechanism of the γ-elimination catalyzed by CGL (Chapter C – 1.4. page 153).

Finally, some PLP-dependent enzymes are capable of catalyzing transamination reactions (Figure 4 f). In this reaction, the amine group of one amino acid is transferred to an α-keto acid producing a different amino acid. The amino acid substrate undergoes a set of reactions until being released as a keto acid and left its amino group bonded to

the PLP cofactor. Then, a different keto acid enters and carries the opposite reaction to form the respective amino acid [26].

Previously, it was presented a panoply of specific reactions catalyzed by PLP-dependent enzymes. Interestingly, different enzymes have evolved to take advantage of PLP as a cofactor and modulate its chemistry to catalyze so much different reactions. One of the main goals of this project is to disclose how enzymes use PLP with such a chemical variety.

Although PLP-dependent enzymes can catalyze so many different reactions, there are two initial stages that are common to all enzymes that use PLP as a cofactor.

The first stage is required to activate the enzyme because it allows the fixation of the PLP cofactor in the apoenzyme's active site. Most of the PLP-dependent enzymes have a Lys residue at the active site that reacts with the PLP cofactor anchoring it. After this reaction, the PLP cofactor becomes covalently bonded to the enzyme, through Lys residue, forming the internal aldimine (IA) [27]. This reaction requires an activation Gibbs free energy of 13.5 kcal/mol and it is exergonic by 9.1 kcal/mol [27]. The enzyme can hold the PLP in the active site, waiting for a substrate molecule to initiate a catalytic cycle. Although this process has been shown to be reversible, the enzyme releases the PLP molecule in a thermo and kinetically less favorable mechanism [27].

The catalytic mechanism for the activation of a PLP-dependent enzyme was already computationally studied by E.F. Oliveira *et al*. [27]. The activation mechanism consists of three sequential and reversible steps (**Figure 5**). The PLP cofactor enters the active site of the enzyme, and its reactive carbonyl group undergoes a fast ($\Delta G^{\ddagger}$ = 0.2 kcal/mol) nucleophilic attack by the key Lys residue (**Figure 5** A). Then, a water molecule mediates a proton transfer from the Lys amine group to the hydroxyl group of the PLP (**Figure 5** B). In some PLP-dependent enzymes, this proton transfer can be mediated by other species as a Cys residue [27]. Finally, the imine formation occurs, triggering the release of a water molecule (**Figure 5** C). At this moment, the IA is fully formed, and the enzyme waits for the substrate to proceed with the catalysis. In any case, the IA can be disrupted and the PLP cofactor released, following the opposite mechanism as illustrated in **Figure 5** (D E F).

After the activation of the enzyme through the covalent binding of the PLP cofactor by a Lys residue, a molecule of the substrate can enter the active site (**Figure 5** G). The substrate needs to react with the IA to become covalently bonded to the PLP cofactor, forming the external aldimine (EA). The mechanism for the EA formation was proposed to be the same across all PLP-dependent enzymes [28-32]. The mechanism was

computationally studied by N.M.F.S.A. Cerqueira *et al.* [28], and then confirmed with other studies on different PLP-dependent enzymes [29, 32].

The formation of the EA starts with the nucleophilic attack of the amino acid substrate to the carbon of the IA imine bond (**Figure 5** H). This step results in an intermediate where both the Lys residue and the substrate are covalently bonded to the PLP. At this stage, a proton is transferred from the substrate's amine group to the amine group of the Lys residue (Figure 5 I). The proton transfer has been proposed to occur through a Tyr or Cys residue, a water molecule, or even without the interference of an additional species [28, 29, 32]. This reaction triggers the release of the Lys residue, and solely the substrate becomes covalently bonded to the PLP cofactor (EA).

Since the EA is formed, the enzyme is ready to catalyze the respective specific reaction (**Figure 5**). In this case, distinct catalytic mechanisms are recognized for different enzymes. Even though different enzymes catalyze the same type of reaction for different substrates, the catalytic mechanisms can be different depending on the substrate. In some cases, the high specificity requirement for a certain substrate can lead to considerable alterations in the active site conformation. Consequently, the catalytic mechanism for a certain type of reaction can be different when similar enzymes are compared.

When the enzyme finishes the set of specific reactions, usually the product is still bonded to the PLP cofactor as a different EA. Consequently, the enzyme proceeds to release the product and restore the initial IA conformation. At this moment, the imine bond undergoes a nucleophilic attack by the Lys residue (**Figure 5** K). Similarly to the mechanism where the EA is formed (**Figure 5** H), an intermediate is formed with both Lys residue and the product bonded to the PLP. Finally, the product of the reaction is released through the transfer of a proton from the amine group of the Lys residue to the amine group of the abandoning product (**Figure 5** L). At this stage, the IA is restored, and the enzyme is ready to receive another molecule of the substrate to undergo a new catalytic cycle. Alternatively, the enzyme can be "deactivated" (**Figure 5** D), which involves the release of the PLP cofactor to go back to the apoenzyme form.

The activation of PLP-dependent enzymes and the formation of the EA are common to all PLP-dependent enzymes. Therefore, only the specific reactions were studied for each PLP-dependent enzyme studied in this project.

For all the PLP-dependent enzymes studied in this project, a more detailed contextualization is provided in the respective section of chapter C (page **87**).

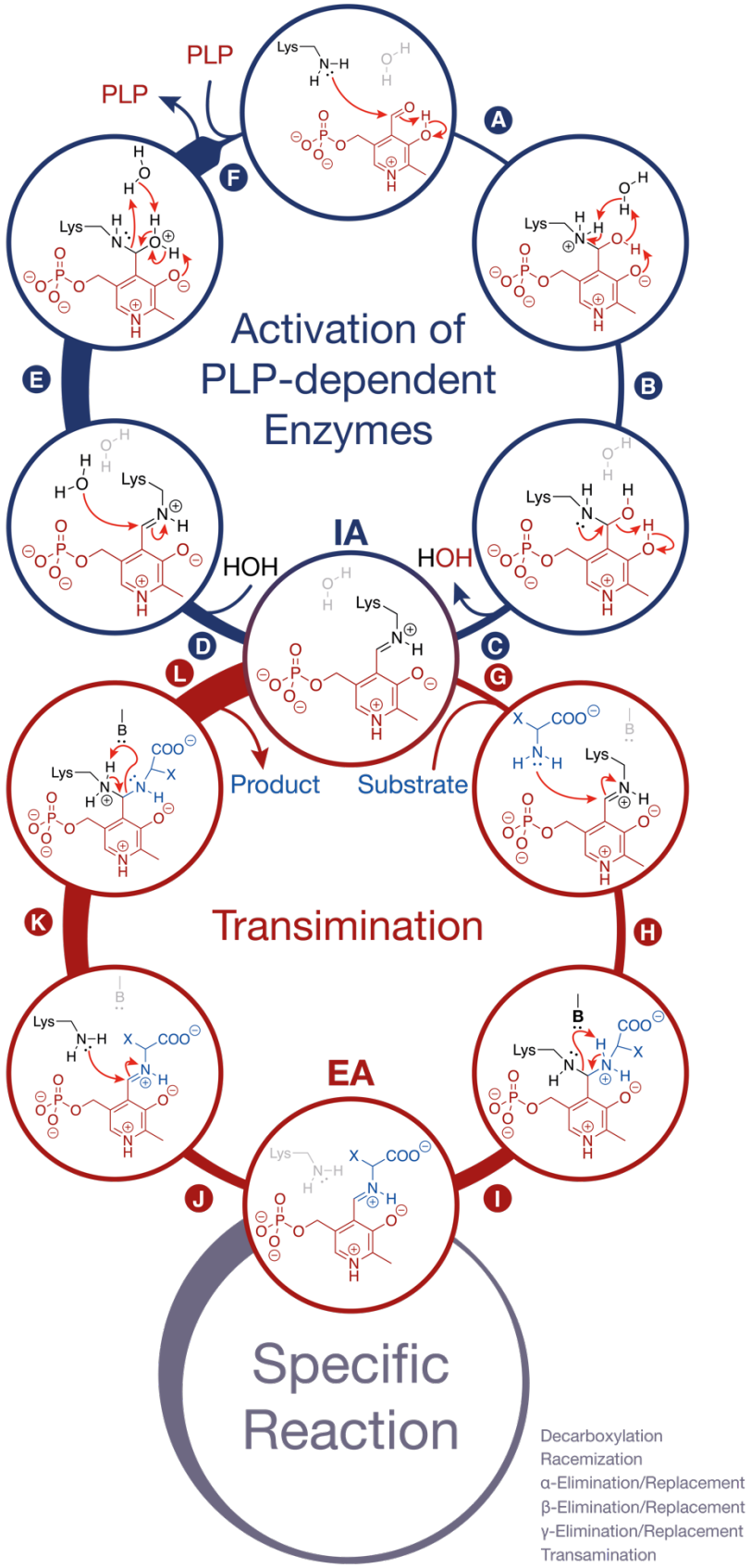


Figure 5. Representation of the general catalytic mechanism for the formation of the internal (IA) and external (EA) aldimines in PLP-dependent enzymes.

1.1.2. Tetrahydrofolate (THF)

The PLP cofactor mentioned before is an example of a very important mediator of biological events in living organisms. THF is another cofactor that plays an important role in the one-carbon transfer reactions along with biotin and *S*-adenosylmethionine (adoMet) [1].

THF (Figure 6) is a derivative of the vitamin B_9 (also known as folic acid or pteroyl-L-glutamic acid) (Figure 6) and participates in key reactions associated with the synthesis of amino acids and nucleic acids [33].

All plants and fungi, as well as some protozoa, bacteria, and archaea, can produce vitamin B_9 *de novo*. The remaining organisms, as humans, need to obtain vitamin B_9 from diet [34]. Vitamin B_9 is then converted in its active form through a four electrons reduction catalyzed, in two catalytic cycles, by the dihydrofolate reductase enzyme. These two cycles oxidized two molecules of NADPH [35].

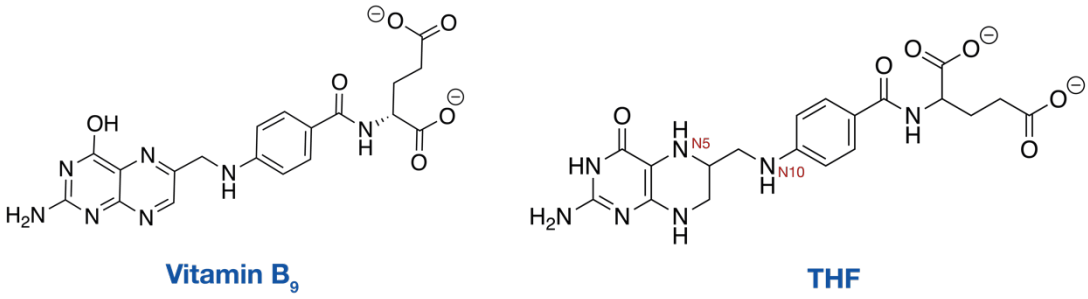


Figure 6. Structures of vitamin B_9 and THF.

As stated before, THF is a single-carbon carrier that is bonded to position N5, N10, or both (Figure 6). Serine, glycine, tryptophan, and histidine are possible sources of carbon to THF, but serine is undoubtably the major source. SHMT (Chapter C – 1.2. page 111) catalyzes the transfer of a one-carbon unit from serine to THF, originating 5,10-methylene THF [36].

The 5,10-methylene THF molecule can then transfer the one-carbon unit to a deoxy-uridine monophosphate (dUMP) molecule to convert it into deoxythymidine monophosphate (dTMP). This reaction is catalyzed by the thymidylate synthase, and it is crucial for supplying the dTMP needed for the DNA synthesis (Figure 7) [37].

The 5,10-methylene THF also plays an important role in the maintenance of the methylation pathways in cell. This intermediate is converted into 5-methyl THF by the methylenetetrahydrofolatereductase (MTHFR) enzyme. The resulting 5-methyl THF

molecule is required for the conversion of homocysteine into methionine by the methionine synthase (Met synthase) (Figure 7) [38, 39].

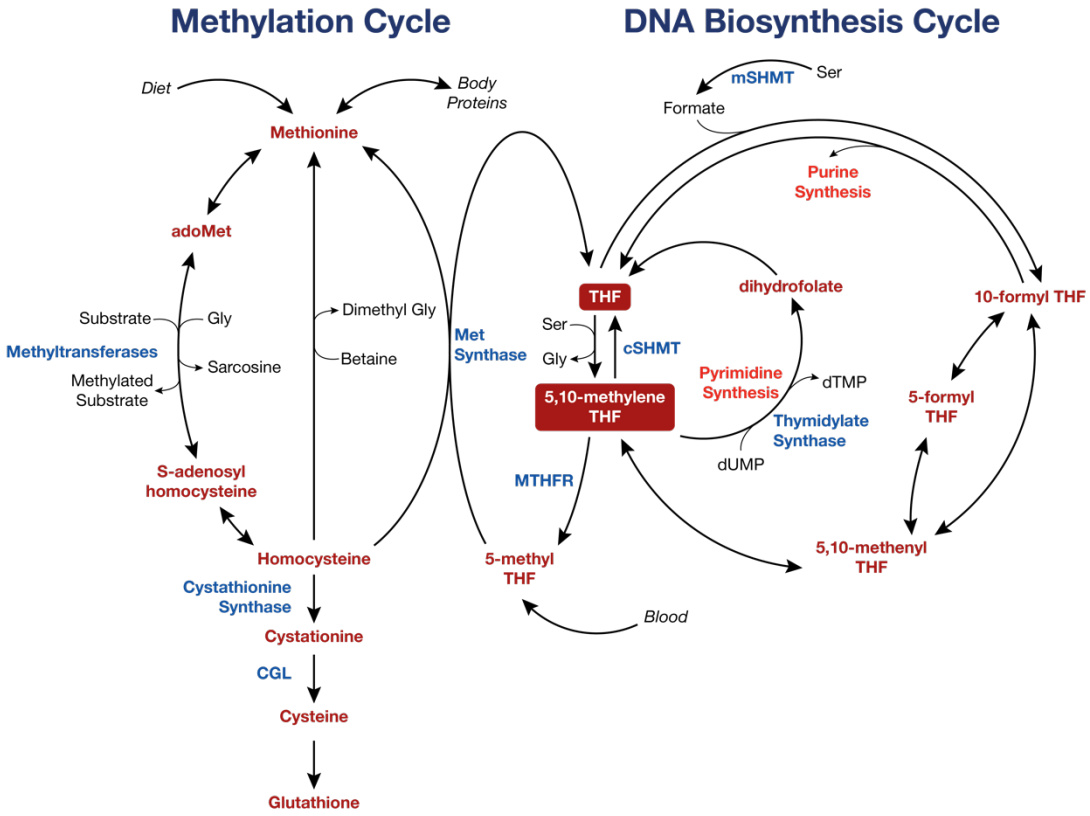


Figure 7. Intracellular folate metabolic pathways. cSHMT: cytosolic SHMT, mSHMT: mitochondrial SHMT. [38, 39]

1.2. Metalloenzymes

Metalloenzymes are a subgroup of enzymes that employ metal ions as cofactors. As already mentioned, metalloenzymes can use different metals belonging to the alkali, alkali earth, and first-row transition metals groups of the periodic table [5]. In addition, some enzymes can also use Mo from the second-row transition metals group [40].

Although the main goal of this Ph.D. project is the study of PLP-dependent enzymes, there was the opportunity to study the catalytic mechanism of the XO enzyme with different substrates. Therefore, the catalytic mechanism of the xanthine oxidation (Chapter C – 2.1. page 181) and nitrate and nitrite reduction (Chapter C – 2.2. page 199) by the XO was studied. The following section provides a quick overview of Mo as a cofactor and Mo-dependent enzymes.

1.2.1. Molybdenum

Molybdenum (Mo) is a heavy metallic element belonging to the sixth group of the *d*-block of the periodic table. Mo is a trace element on Earth and also in the human body, where it is about 17 million times less abundant than carbon [40].

Despite that scarcity, most organisms rely on Mo for their survival [41], being found in the active site of enzymes that catalyze oxidation-reduction reactions [42-46]. There are an important set of reactions where Mo-dependent enzymes are crucial for the natural cycles of carbon, nitrogen, sulfur, and oxygen. For example, the nitrogenase enzyme, a Mo-dependent enzyme, catalyzes the important fixation of atmospheric dinitrogen into organic ammonium [47].

a. Mo-dependent enzymes

The interest in Mo-dependent enzymes started in the early 50s when electron paramagnetic resonance (EPR) spectroscopy was used to demonstrate Mo reduction to Mo(V). These studies were mainly focused on the XO and sulfite oxidase enzymes [48-54]. Currently, more than 50 Mo-containing enzymes are known, from which the greatest amount was found in prokaryote [55]. Among all the Mo-dependent enzymes already known, most of them employ Mo in the mononuclear form: Moco centers. In these centers, one Mo ion is coordinated by the *cis*-dithiolene group of one or two pterin molecules (Figure 8 A) and/or an oxygen/sulfur/selenium atom(s) in a panoply of possible arrangements (Figure 8 B). Based on those structures of Moco, three large families of Mo-dependent enzymes were established: the xanthine oxidase family, the sulfite oxidase family, and the dimethylsulfoxide reductase family.

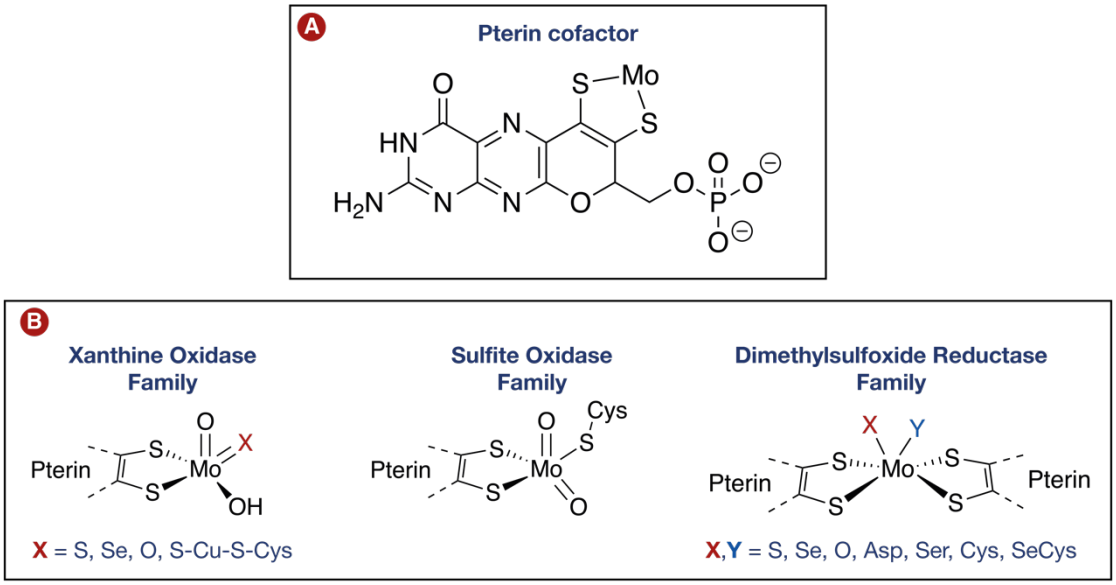


Figure 8. Schematic representation of the pterin cofactor structure. Schematic representation of the general Moco cofactor structure for enzymes belonging to each of the three main families of Mo-dependent enzymes.

As already mentioned, the XO enzyme was the target of two computational studies in the course of this project. Therefore, particular focus will be given to the xanthine oxidase family of Mo-dependent enzymes.

As shown in Figure 8, the active site of a xanthine oxidase family enzyme has a Mo ion coordinated in a square-pyramidal geometry. An oxo group occupies the apical position, where the equatorial plane is occupied by the two sulfur atoms of the pterin, a catalytic OH group, and a terminal sulfo, seleno, or oxo group (X position). In the case of XO, the X position is occupied by a sulfur atom.

The studies of the XO started with the observation of a methylene blue reduction by formaldehyde in the presence of fresh cow milk [56]. Later on, it was shown that milk was capable of oxidizing hypoxanthine and xanthine (XAN) to urate (URC) [57]. Since then, several studies have been conducted to understand the structure and catalytic mechanism of XO.

In mammals, XO is a cytoplasmatic enzyme that act as NAD^+-dependent dehydrogenase (xanthine dehydrogenase – XD). Although XD can exist under normal physiological conditions [58-60], XD can easily be converted in the oxidase form (XO), where it reduced O_2 instead of NAD^+. This process can occur reversibly through the oxidation of key Cys residues or irreversible by proteolysis [61-68]. Despite the different final electron acceptor, these differences do not impact the catalytic mechanism.

XO has three additional redox-active centers essential for its plenty enzymatic activity. There are two [2Fe-2S] clusters that establish an “electron chain” between the Moco and

the FAD cofactors. When the XO catalyzes the oxidation of hypoxanthine or xanthine, the Mo ion becomes reduced (Mo(VI) to Mo(IV)). In order to complete the catalytic cycle, two electrons are transferred throughout the two [2Fe-2S] clusters until they reach the FAD cofactor. Then, FAD is reduced to $FADH_2$. Finally, an O_2 or NAD^+ molecule (depending on the XO form) oxidizes the $FADH_2$ molecule to FAD, restoring its initial oxidation state (**Figure 69**). This process is absolutely essential for the enzymatic turnover. Otherwise, the XO will be stuck in the reduced form and could not oxidize more hypoxanthine or xanthine molecules [55].

The enzymes belonging to the XO family typically catalyze the hydroxylation of the C-H bond in aromatic heterocyclic compounds and aldehydes [58-60]. In particular, XO mainly catalyzes the oxidation of hypoxanthine into XAN, and XAN into URC [55].

The catalytic mechanism of XO has been proposed based on the available structural and kinetic studies [58-60, 69-71]. Moreover, there were already some computational studies devoted to the atomistic and electronic description of this mechanism [72-74]. Unfortunately, those computational studies were not capable of providing a full description for the mechanism. Therefore, the catalytic mechanism of XO was studied in the project, considering the reaction of XAN oxidation into URC (Chapter C – 2.1. page **181**).

More recently, it has been shown, under certain stress conditions, that XO can show nitrate and nitrite reductase activity, catalyzing the reduction of nitrate into nitrite and nitrite into nitric oxide (NO), respectively [75-78]. Accordingly, the produced NO provides an important protective role against the ischemic damage [79-83].

These studies showed that XO has certain catalytic plasticity to catalyze this type of reactions in situations where O_2 is absent. In this way, XO overcomes the lack of its final electrons-acceptor, oxidizing the Moco cofactor and producing NO [77, 84-92].

Until then, there was no atomistic description of the catalytic mechanism behind the nitrate/nitrite reductase activity of XO. Hence, computational methods were employed, in this work, to achieve a description of the mechanism associated with this reductase activity of XO (Chapter C – 2.2. page **199**).

2. The Clinical and Industrial Relevance of Enzymes

Enzymes are powerful biocatalysts that assume important roles in several biological systems required by all living organisms. The specificity and efficiency of enzymes are a consequence of millions of years of evolution where these structures have been progressively enhanced.

Since the earliest stages of humanity, humans have benefited from nature, collecting and learning from it. Enzymes are no exception, and they can also be exceptional tools to use as therapeutic agents or even as industrial catalysts.

Despite all the incredible functions assigned to enzymes, they are not all about perfection. There are several diseases caused by malfunction, excessive or deficient activity. Additionally, in the case of infections, enzymes from infectious agents can be excellent targets to control and/or treat associated infection.

The following three sections present an outline of how enzymes can be seen as drug targets against several diseases, therapeutic agents, or industrial catalysts.

2.1. Enzymes as Drug Targets

Pharmacologically speaking, a drug target is a biological entity to whom a drug can be directed to interact and change its function [93]. Commonly, a pathology is caused by a malfunction in a specific subcellular structure or a set of structures. Subsequently, drugs can be administrated to target those structures aiming a behavior modulation of their activities. Considering enzymes, there are numerous diseases caused by an excessive activity of a certain enzyme or set of enzymes such as gout and uric acid stones formation caused by XO [94-99], atopic dermatitis [100], chronic allergic contact dermatitis [101], allergic rhinitis [102], gastric ulcer [103, 104], diabetes [105], chronic heart failure [106], and several types of cancer caused by HDC [107-111], among many others.

Enzymes belong to the topmost studied groups of drug targets along with nuclear hormone receptors, G-coupled protein receptors (GPCRs), transporters, ion channels, and nucleic acids [112]. This interest in enzymes as therapeutic drug targets, especially through inhibition, culminates in almost half (47 %) of all commercially available drugs being enzyme inhibitors in 2005 [113, 114]. Recent publications have also highlighted an increasing interest in the enzyme inhibitors development [115-117], showing that, in 2018, 53% of the clinical candidates were drugs directed against enzymes [116].

The increasing interest in enzymes as drug targets are a consequence of their intrinsic features that allow the development of highly specific inhibitors, taking advantage from the natural specificity of enzymes to certain substrates. Additionally, the huge improvement in the scientific techniques and technology of the past decades has contributed to have a more detailed description of the structure and function of enzymes. Computational studies on the catalytic mechanism of enzymes, molecular dynamics simulations, free-energy calculations, virtual screening, among others have greatly contributed to these advances [118-122].

In 2005, 317 FDA (Food and Drug Administration) approved drugs acted as inhibitors of 71 different enzymes, and about 65% of those drugs react in the active site of the enzyme. Among the 71 enzymes, four of them are inhibited by transition state (TS) analogs [112, 123]. The major part of the enzyme inhibitors react in the active site of their targets, and some of them were designed as TS analogs. The complete electronic and atomistic description of the mechanisms behind those reactions and characterization of all intermediates and TSs are only possible using highly accurate computational methodologies coupled with valuable experimental data [118].

The study of enzymatic mechanisms by computational means provides valuable information for the rational drug design of new inhibitors. In addition, the increasing knowledge about the functional and structural features of enzymes deeply improves the understanding of general enzymatic catalysis.

2.2. Enzymes as Therapeutic Agents

In terms of the clinic, enzymes are not always seen as targets to be inhibited. Indeed, enzymes can also be used, in a modified or original form, as therapeutic agents.

Enzymes can be seen as “nanomachines” capable of efficiently convert different substrates in their respective products. They are also fully compatible with biological systems and highly efficient under normal biological conditions. These are some of the features that turn enzymes into strong candidates to be used as therapeutic agents.

One of the most studied approaches is using enzymes to deplete or decrease the levels of a certain molecule in the human body. The success of this approach is corroborated by several patents addressing this strategy [124], and more importantly some of them already approved by the FDA to be used clinically [124, 125], which is the case of L-asparaginase (ASNase) [126, 127] commercialized as Elspar®, Oncaspar®, and Erwinaze®. The use of depleting enzymes has been mainly studied and applied to reduce the blood levels of certain amino acids. In this way, the amino acid depletion causes a selective amino acid “starvation” in certain cancer cells or infectious agents, triggering a slowdown in growth and eventually their death. A review about patents on amino acid deprivation strategies for cancer and viral infections therapies was published, and it is available in the appendix: page 393.

Most mammals and humans ceased the capability to produce 9 of the 20 amino acids that are needed for the biosynthesis of proteins. Therefore, the so-called essential amino acids are obtained from the diet and/or microbiota. The list of essential amino acids is composed of valine, phenylalanine, threonine, methionine, leucine, isoleucine, lysine, and histidine.

Six other amino acids are denominated conditionally essential, meaning that their synthesis can be carried out by humans in certain conditions. The conditionally essential amino acids are arginine, cysteine, glycine, glutamine, proline, and tyrosine.

The remaining five nonessential amino acids are dispensable because human cells can synthesize them: aspartate, asparagine, glutamate, alanine, and serine.

The ideal conditions for amino acid depletion therapies occur when the tumors are sensitive to one of the nonessential or conditionally essential amino acids. Tumor cells tend to adjust their metabolic program to prioritize cell growth and proliferation. Consequently, synthesis pathways are commonly downregulated, and the required "building blocks" are obtained by an external supply. In this scenario, the healthy cells are feeding the cancer cells with the required supplies for their carcinogenic activity. Therefore, depleting these compounds from the bloodstream will have a greater impact on cancer cells than normal cells that can still produce these amino acids.

However, the depletion of nonessential and conditionally essential amino acids is not always possible. Some tumors or infectious agents depend on an external supply of one or more essential amino acids. In these cases, the strategy involves using depletion enzymes for essential amino acids, even with expectedly severe side effects.

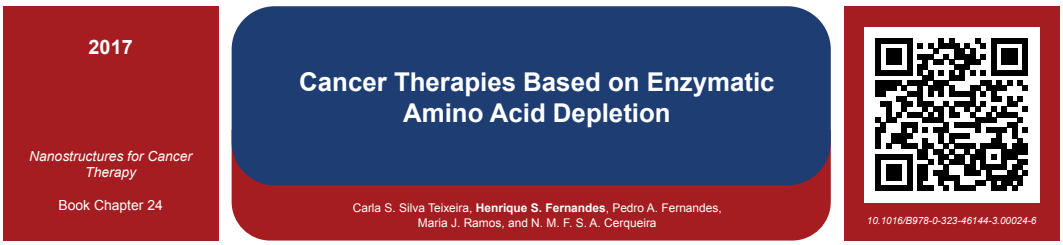
Methionine-γ-lyase (MGL), arginase (ARGase), arginine deiminase (ADI), and ASNase are the most commonly used enzymes for treating certain types of cancer and infections by bacteria, fungi, and viruses.

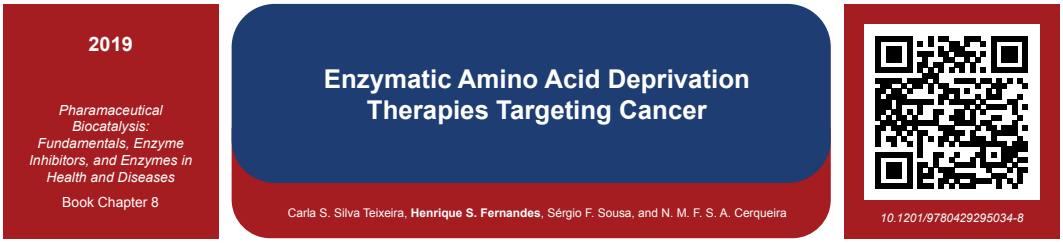
MGL has been used to treat some types of cancers, such as nervous system-related ones [124, 128-132], and also in situations of infection by fungus [124, 133] or virus [134].

ARGase and ADI are two enzymes used to deplete arginine from the bloodstream. An extensive list of cancer types, namely metastatic melanoma, prostate, cervical, squamous cell, breast, ovarian, lung, and colon carcinomas [135, 136], are auxotroph for arginine, which turns them susceptible to its depletion. Moreover, the arginine depletion by enzyme administration has been investigated for viral infections in cases of hepatitis and ocular lesions [124].

As already referred, ASNase is already commercialized for the treatments of certain types of cancers [126, 127]. Nonetheless, ASNase was also suggested for the combined treatment of human immunodeficiency virus (HIV) infection [137].

Regarding the application of amino acid depletion in cancer therapies, two book chapters were written and published (Appendix: page 393).





The knowledge about the structure and function of enzymes provides valuable insights into how we can take benefit from these powerful “nanomachines”. Currently present enzymes in all living organisms result from million years of evolution on improving efficiency and specificity. Therefore, a complete understanding of enzymes could improve the way we use them in a panoply of new application fields, reducing costs, environmental impact, and improving the well-being of the society and entire ecosystem.

2.3. Enzymes as Industrial Catalysts

Up to now, it has been described the application of enzymes in some areas of medicine and pharmacology, either as the direct use of enzymes to treat pathologies or the development of inhibitors for certain enzymes. It is also worth mentioning the growing relevance of using enzymes as catalysts in industry.

Industrial biocatalysts are gaining popularity for the purpose of cost-effectiveness, eco-friendliness, and long-term sustainability [138]. However, the process of finding a suitable enzyme to catalyze a reaction with a direct industrial application is not an easy task. The process involves isolating or modifying an existing enzyme to efficiently catalyze a certain chemical reaction of interest. Nevertheless, the process also needs to be scalable for industrial application, and the biocatalyst (enzyme) needs to be easily separated to be reused [138, 139].

The use of biocatalysts has already been applied in several industrial applications like the food and pharmaceutical industry [138]. Although enzymes are commonly used in several industrial processes, there is plenty of space for exploring new enzymes to replace expensive and hazardous classical chemical processes.

In the last decade, different techniques have been used to improve the applicability of enzymes in industrial scenarios, namely metabolic and genome mining, directed evolution, *in silico* studies, and computational calculations [138, 139]. In 2018, Frances Arnold was awarded the Nobel Prize in Chemistry for her pioneer studies on directed evolution [140, 141]. Directed evolution is a technique where certain enzymes are randomly mutated, and the best ones are selected, discarding the enzymes with non-interesting properties. This approach intends to mimic the natural evolution *in vivo* or *in vitro* and in a much faster way [142]. This promisor strategy has some drawbacks concerning the number of screened mutations required to obtain better enzymes, which could also be expensive and time-consuming.

In this sense, computational simulations can provide an important assistance in restricting the exploration of enzyme-derivatives to a small number of candidates. The understanding of the catalytic mechanisms of different enzymes and different reactions provide valuable insights to propose key mutations with potential enhancing capabilities.

In this project, different mutants of SHMT were tested to improve the enzymatic efficiency to catalyze a reaction with industrial interest (Chapter C – 1.3. page 139). SHMT has shown to be capable of catalyzing the addition of aldehydes to glycine in a stereospecific manner, producing α,α-dialkyl-α-amino acids. The synthesis of this type of compounds is very challenging due to the presence of a stereogenic tetrasubstituted carbon on drugs

like myriocin, sphingofungines E and F, lactacystin, and AADC inhibitors [143]. Different mutations at the active of site of SHMT are being tested to look for a better SHMT mutant to carry on this kind of reactions.

B. Computational Methodologies

A theory can be proved by experiment;
but no path leads from experiment to the birth of a theory.
Albert Einstein

This page was intentionally left blank.

The 20th century was absolutely astonishing in terms of the development of physics, mathematics, and chemistry that provided a better and deeper knowledge about all the matter that surrounds us. In particular, a deep investigation was devoted to understanding sub-atomic particles and phenomena.

All of these advances originated new tools and methods to theoretically study several properties of atomic particles, like thermodynamics and kinetics. In addition, this knowledge was also essential for the invention and development of the computer as we know it today. When these two paths merged, computational chemistry appeared to answer some questions about how chemical systems behave when looking at the atomistic and electronic levels.

Nowadays, we are living in a silver age of computational chemistry since we are able to study a panoply of very interesting chemical and biochemical systems with enough accuracy, albeit with a reasonable number of approximations. These developments were only possible due to the hardware developments that allowed to significantly shorten time spent on computational calculations.

For example, the introduction of GPU-accelerated software in computational chemistry lets molecular dynamics simulation being accelerated hundreds of times. The continuous improvement of such technology and the development of new algorithms will inevitability allow faster and more accurate results from computational simulations.

1. Molecular Mechanics

A reasonable part of the chemistry problems can be studied using high accurate methods like quantum mechanics (QM) since their chemistry is limited to about a few hundred atoms. However, biomolecular systems with thousands of atoms are also extremely appealing to be studied using computational methodologies. For these cases, computational studies provide singular information about, for example, how enzymes catalyze reactions, how we can learn from Nature to industrially produce some chemical species, or how proteins behave in different conditions. Unfortunately, the available computational power and algorithms are not suitable for computing those large biomolecular systems using the QM in a practical amount of time.

Therefore, molecular mechanics (MM) appears like an excellent inexpensive alternative to treat those large systems with the counterpart of some limitations regarding the accuracy and the type of phenomena that can be studied.

In MM, the atom is the smallest entity that interacts with the other atoms by simple potential-energy functions as harmonic oscillators and Coulombic potentials. Consequently, all phenomena involving electrons, such as chemical reactions where chemical bonds are broken and/or formed, cannot be studied in this way. Consequently, MM is suitable to describe the spatial rearrangement of thousands to millions of atoms within the timescale of the femtosecond to second [144].

In the computational field, the temporal evolution of a set of atomic positions is named molecular dynamics (MD). MD simulations are extensively used to sample the conformation space occupied by biomolecular systems through the mathematical integration of the classical equations of movement. The implementation of MD just requires a force field that can accurately describe the physics of the molecular system.

1.1. Classic Force Fields

The classic force fields are designed to apply simple classical mechanical equations to threat molecules as a set of charged spheres bonded together through springs threated as harmonic potentials. Accordingly, the total energy (*U*(r)) of the system depends on two different types of interactions: bonded and non-bonded interactions (Equation 1) [145, 146].

$$U(r) = U_{bonded}(r) + U_{non-bonded}(r)$$

Equation 1. Total energy function for a system treated with MM.

The two addictive terms of the Equation 1 assume different forms depending on the force field that is used. For biomolecular systems, there are a reasonable amount of different force fields, such as GAFF[3] [147, 148], ff99SB [149], ff14SB [150], CHARMM[4] [151], OPLS[5] [152, 153], and GROMOS[6] [154].

In this work, the AMBER[7] software [155], together with the GAFF, ff99SB, and ff14SB force fields, was used. In the following sections, the general potential energy function is explained in detail, even though each force field applies slightly different variations of this general formalism.

1.1.1. Bonded Interactions

The potential-energy term for bonded interactions includes the sum of the energy involved in the 1-2 bond stretch, the 1-3 angle bend, and the 1-4 dihedral and improper torsions (Equation 2) [145].

$$U_{bonded}(r) = \sum_{bonds} U_{bond}(l) + \sum_{angles} U_{angle}(\theta) + \sum_{dihedrals} U_{dihedral}(\rho) + \sum_{impropers} U_{improper}(\varphi)$$

Equation 2. General energy function for bonded terms of AMBER force field.

Under the MM approach, the bond stretches are described considering the harmonic oscillator approximation. Consequently, the potential-energy for the bond length

[3] General AMBER Force Field
[4] Chemistry at Harvard Macromolecular Mechanics
[5] Optimized Potentials for Liquid Simulations
[6] Groningen Molecular Simulation package
[7] Assisted Model Building with Energy Refinement

displacement (l) relative to the equilibrium position (l_0) is given by Equation 3 following Hooke's law [144, 145].

$$U_{bond}(l) = \frac{1}{2}k_b(l - l_0)^2$$

Equation 3. Energy function for bond terms.
k_b: force constant; l: bond length; l_0: bond length at the equilibrium

An identical approach is used to compute the potential-energy associated with the angle amplitude fluctuations established by three covalently bonded atoms. The energy increases depending on how far the angle displacement (θ) is with respect to the equilibrium angle (θ_0) [144, 145].

$$U_{angle}(\theta) = \frac{1}{2}k_a(\theta - \theta_0)^2$$

Equation 4. Energy function for angle terms.
k_a: force constant; θ: bond length; θ_0: bond length at the equilibrium

Both the bond stretching and the angle bend energies require two parameters for each pair or trio of atoms: the force constant and the equilibrium value for which the energy expression is null.

The last two types of bonded interactions are calculated differently depending on the manner the four atoms are linked. For the dihedral interaction, it is considered a series of four atoms A, B, C, and D linked together in this order (Figure 9). The dihedral angle corresponds to the angle between the ABC and BCD planes (ρ). The relative movement of one of the planes with respect to the other one causes a variation of the energy as a function of Fourier series of the dihedral angle (Equation 5) [144, 145].

$$U_{dihedral}(\rho) = \frac{1}{2}k_d(1 + \cos(n\rho + \gamma))$$

Equation 5. Energy function for dihedral angle terms.
k_d: force constant; ρ: dihedral angle; γ: dihedral angle for which the energy is minimum; n: number of minima of the energy function

Lastly, the improper angles are also defined by four atoms, where A, B, and C are linked together in this order, and the D atom is bonded to B (Figure 9). The improper angle corresponds to the angle between the ABC and ACD planes. The expression for this contribution accounts for the angular deviation (φ) from the equilibrium position (φ_0) (Equation 6) [144, 145].

$$U_{improper}(\varphi) = k_i(\varphi - \varphi_0)^2$$

Equation 6. Energy function for improper torsion terms.
ki: force constant; φ: improper angle torsion; φ0: equilibrium value for the improper torsion

In Figure 9, it is represented all types of bonded movements accounted for the potential-energy expression of a MM force field.

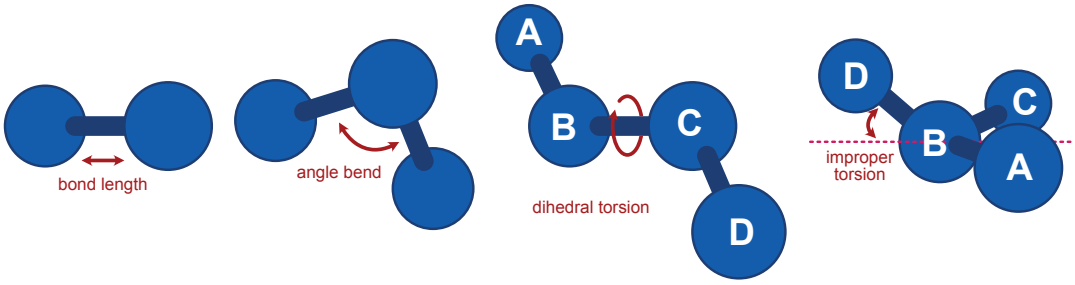


Figure 9. Normal modes commonly applied in MM biomolecular force fields.

1.1.2. Non-Bonded Interactions

The non-bonded interactions complete the energy expression for a system treated using MM. These interactions are calculated using the combination of Coulomb's potential and a 12-6 Lennard-Jones potential for short-range interactions (Equation 7) [144, 145].

$$U_{non-bonded} = \sum_{i}^{N} \sum_{\substack{j \\ i<j}}^{N} \left(\frac{1}{4\pi\epsilon} \frac{q_i q_j}{r_{i,j}} + \epsilon_{i,j} \left(\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right) \right)$$

Equation 7. Energy function for non-bonded interaction.
ϵ: dieletric constant; q_i and q_j: charge of atom i and j, respectively; $r_{i,j}$: distance between atoms i and j; $\epsilon_{i,j}$ and $\sigma_{i,j}$: parameters for the interaction between the atoms i and j

The parameters for the long-range interactions, $\epsilon_{i,j}$ and $\sigma_{i,j}$, are calculated based on the atomic parameters for each atom, ϵ_i and ϵ_j, and σ_i and σ_j, respectively, according to the Lorentz-Berthelot rules (Equation 8 and Equation 9) [156, 157].

$$\epsilon_{i,j} = \sqrt{\epsilon_i \, \epsilon_j}$$

Equation 8. Determination of $\epsilon_{i,j}$ parameter for atoms i and j.

$$\sigma_{i,j} = \frac{\sigma_i + \sigma_j}{2}$$

Equation 9. Determination of $\sigma_{i,j}$ parameter for atoms i and j.

Based on this potential-energy expression, the total energy of a biomolecular system can be calculated in fractions of seconds. The calculation requires the atomic position of all atoms of the system, the information about how the atoms are bonded to each other, and the empirical parameters for all the energy terms on the expression (available in the force field).

1.2. Molecular Energy Minimization

Usually, the study of a biomolecular system starts with an X-Ray crystallographic or Nuclear Magnetic Resonance (NMR) structure, which means that some modeling is required before starting the calculations. One of the most common required modifications is the addition of missing hydrogen atoms on the original structure, the inclusion of solvent and counterions, and/or reversion of mutations that were included, e.g., to help in the crystallographic process.

Consequently, the overall molecular system is not settled in a local minimum of the potential-energy surface that is described by the specific employed force field. Therefore, it is common to perform a set of energy minimizations using a MM force field like the one mentioned before. Energy minimization is a process where it is needed to find the coordinates of all atoms ($\boldsymbol{r}_{min}$) of the system that minimizes the energy of the entire system (U_{min}). However, for a multidimensional system as a biomolecular complex, neither $\boldsymbol{r}_{min}$ or U_{min} are known. Then, the best solution is to conduct an iteratively approach to test several geometry conformations ($\boldsymbol{r}_n$) until we reach a certain energy variation threshold. Exemplifying, if we start from a $\boldsymbol{r}_0$ conformation whose potential energy is U_0, a new set of coordinates are extrapolated $\boldsymbol{r}_1$ as well as its potential-energy U_1. Considering the difference between U_1 and U_0, it is possible to determine the potential-energy gradient along the $\boldsymbol{r}_0$ to $\boldsymbol{r_1}$ direction. Based on that, successive minimization cycles (n) are followed to achieve a situation where the difference between U_{n+1} and U_n is below a custom threshold that is very close to zero.

In a typical MM minimization, two different types of methods can be applied: steepest descent [158, 159] or conjugate gradients [160].[161] In the MM energy minimizations performed in this work, both methods were employed. The steepest descent is considerable faster to calculate each iteration but requires more iterations to achieve the desired threshold. On the other hand, the conjugate gradients method takes more time in each iteration but reaches the minimized structure in a lower number of iterations. Therefore, a potential-energy minimization starts with the steepest descent to rapidly approximate the structure to the minimum well. Then, the latest iterations are conducted using the conjugate gradients method (**Figure 10**). This strategy allows a good balance between performance and accuracy.

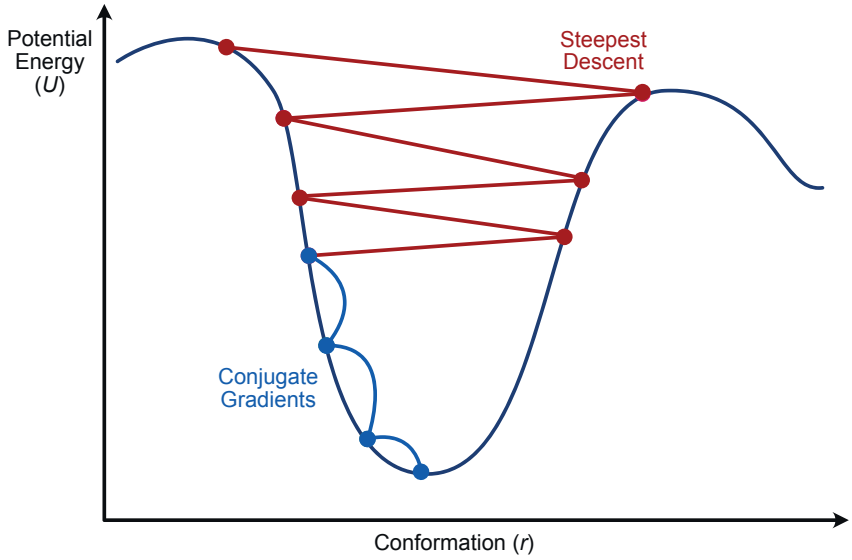


Figure 10. Schematic representation of the behavior of steepest descent and conjugate gradients algorithms used sequentially for potential-energy minimization.

In practical terms, the minimization of the initial molecular complex occurs in separated stages to progressively adjust the position of the different parts of the complex without cause clashes or aberrations. Generally, the hydrogens are the first atoms to be minimized, followed by the water molecules, the protein sidechains, and the entire system.

1.3. Molecular Dynamics Simulations

Although the energy minimization is an inexpensive manner to get a structure of a local minimum, this is only a static representation of one of the infinite conformational possibilities that would occur in a certain time frame. In order to increase the sample of structures that occur over time, an MD simulation can be performed.

An MD simulation is a common computational approach to predict the temporal evolution of a certain system. In this particular case, an MD simulation can be used to predict how the atoms of a molecular system behave during a certain simulation's time range.

This kind of simulation provides relevant information about the conformational space, thermodynamics, and structural flexibility properties.

The minimum energy of a macroscopic system, at a certain temperature (*T*), pressure (*P*), or chemical potential (μ), is given by an ensemble of different molecular conformations. Consequently, a macroscopic property can be calculated through the statistical treatment of an ensemble of conformations.

A relationship between the force ($\boldsymbol{F_i}$) that each atom i of mass m_i feels and its position $\boldsymbol{r_i}$ can be obtained through the second order derivative of the Newton's Second Law of motion (**Equation 10**) [144, 145].

$$\boldsymbol{F_i} = m_i \frac{d^2 \boldsymbol{r_i}}{dt^2}$$

Equation 10. Second order derivative of the Newton's Second Law.

Since the force $\boldsymbol{F_i}$ can be given by $-\boldsymbol{\nabla} U_i(\boldsymbol{r_i})$ where the U_i is the potential-energy of the atom i and can be calculated using the energy equation of the force field as explained above. Consequently, the potential energy of the system can be correlated with the position of all atoms of the system ($\boldsymbol{r}$) (**Equation 11**).

$$-\boldsymbol{\nabla} U(\boldsymbol{r}) = m \frac{d^2}{dt^2} \boldsymbol{r}(t)$$

Equation 11. Mathematical relation between potential energy and atomic positions.

However, **Equation 11** cannot be analytically solved because it is, in fact, a system of coupled polyatomic equations. Consequently, this problem only has a numeric solution, where the next position of each atom can be calculated from its actual position, its velocity, and its acceleration [144, 145].

Considering a simple case where atom i is in the position $\boldsymbol{r_i}$ at the time t_0, and its potential-energy U_i is calculated from the energy expression of the force field. Since the

mass of the atom is known (m_i), the acceleration ($\boldsymbol{a_i}$) that the atom i undergoes can be calculated from the U_i. Then, the position of atom i at the time t_1 ($t_1 > t_0$) can be calculated according to the following equation [144, 145]:

$$\boldsymbol{r_i}(t_1) = \boldsymbol{r_i}(t_0) + \boldsymbol{v_i}(t_0) \times (t_1 - t_0) + \frac{1}{2}\,\boldsymbol{a_0}(t_0) \times (t_1 - t_0)^2$$

Equation 12. Function to determine the position of an atom i at time t_1.

The Equation 12 also requires the velocity ($\boldsymbol{v_i}$) of atom i on time t_0. For the subsequent time frames (t_1, t_2, t_3…), the velocity of the atom can be easily calculated based on the Equation 13 [144, 145].

$$\boldsymbol{v_i}(t_1) = \boldsymbol{v_i}(t_0) + \boldsymbol{a_0}(t_0) \times (t_1 - t_0)$$

Equation 13. Function to determine the velocity of an atom i at time t_1.

The main question that remains to be solved is how to obtain the initial velocity ($\boldsymbol{v_i}(t_0)$) for all the atoms of the system. Usually, these velocities ($\boldsymbol{v_i}(t_0)$) are derived from a Maxwell-Boltzmann distribution at a given temperature, in a way that the center of mass of the system has a velocity equal to zero [144, 145].

In sum, three main variables need to be known to predict the conformation evolution over a certain time range: initial atomic positions, initial velocities for each atom, and the acceleration that each atom undergoes. However, this approach has several limitations that must be taken into account and rationalized for each particular biomolecular system. The main limitations and approaches to vanish them are going to be discussed in the following sections.

1.3.1. Integration Time Step

In order to access the conformation space of a protein over a time range of 100 ns, for example, that time needs to be fractionated into small increments of time (Δt) called the integration time step. During the Δt, all the atoms of the system move linearly, undergoing the same acceleration. Consequently, the Δt needs to be small enough to avoid aberrant movements. In the limit, the smaller the integration time step, the more accurate the results are, becoming closer to the analytic solution. However, a small integration step requires more calculations and becomes more time demanding. Here, a commitment needs to be assumed to achieve a good balance between computational cost and accuracy.

One of the most used strategies is to use an integration time step that is 10 times smaller than the fastest phenomenon that occurs in the system. This allows a good compromise

between computational efficiency and accuracy. Considering a biomolecular system, the fastest event corresponds to the vibration of hydrogen-involving bonds (e.g., C-H, N-H, and O-H). Each vibration takes approximately 10 fs to occur at normal temperature, so it is reasonable to use an integration time of 1 fs. This integration step allows calculating 10 iterations for each vibration, providing a good description of the movement and avoids misbehaviors.

Bearing in mind the 100 ns simulation example, the calculations for all the atoms of the system needed to be made 1×10^8 times if a 1 fs integration time was used. This becomes high costly if bigger systems are used or greater simulation times are required. To overcome this issue, some algorithms were developed to allow the use of bigger integration steps [144, 145].

The SHAKE algorithm is one of the most used strategies to increase the integration step. When this algorithm is used, some of the fastest vibrations are constrained. This means that the increase in the integration step does not cause aberrant description of these fastest movements [144, 145].

1.3.2. Ensembles

The statistical mechanics describes how the molecular events and the macroscopic system properties are related. By definition, a macroscopic system is defined by $\alpha + 2$ functions of state, where α if the number of system's components. Temperature, volume, number of particles, and pressure are examples of state functions.

The macrostate can be described as an ensemble of microstates under the same thermodynamic conditions. For the study of biomolecular systems, the most popular ensembles are the canonic (number of particles (*N*), *V*, and *T* constants), the isothermal-isobaric (*N*, *P*, and *T* constants), the microcanonic (*N*, *V*, and total energy constants), and the grand canonic (*V*, *T*, and *μ* constants). Among these ensembles, the canonic and isothermal-isobaric were the ones chosen for the MD simulations, regarding the equilibration and production stages, respectively [144, 145].

1.3.3. Pressure and Temperature Control

Since the production stage of an MD simulation is conducted using the isothermal-isobaric ensemble, the three state functions (*N*, *P*, and *T*) that characterize this ensemble need to be kept constant.

The number of particles in the system (*N*) is immutable, so this condition is assured across the entire simulation. However, the potential-energy expression of the force field

does not include any term to control the temperature and pressure of the system. Therefore, both the temperature and the pressure of the system need to be evaluated in each iteration to iteratively add a correction to keep these values constant. There are different types of algorithms to do so, but the ones used in this work were the Berendsen barostat [162] and the Langevin dynamics [163, 164] to control pressure and temperature, respectively.

According to the Berendsen barostat, the pressure of the system is given by the sum of the average velocities of the atoms, considering an ideal gas, and the consequent potential-energy raised because of the interaction of all the atoms with each other (Equation 14) [144, 162].

$$P(t) = \frac{n_{particles}}{V_{cell}} k_B T(t) + \frac{\sum_{i,i<j}^{j} \boldsymbol{F_{ij}}(t) \cdot \boldsymbol{r_{ij}}(t)}{3V_{cell}}$$

Equation 14. Pressure of a system according to the Berendsen barostat.
$P(t)$: pressure of the system; $n_{particles}$: total number of particles; V_{cell}: volume of the unit cell; k_B: Boltzmann constant; $T(t)$: temperature of the system; $\boldsymbol{F_{ij}}(t)$: force that atom j exerts on atom i; $\boldsymbol{r_{ij}}(t)$: distance between atoms i and j

The pressure of the system needs to be constantly corrected to the constant pressure (P_0), so a friction coefficient ($\mu(P)$) is used to scale the velocities of the system's atoms. For a cubic cell, the friction coefficient is given by Equation 15. [162]

$$\mu(P) = \sqrt[3]{1 + \frac{\Delta t}{\tau_P}(P - P_0)}$$

Equation 15. Friction coefficient ($\mu(P)$) as a function of pressure variation.
τ_P: relaxation time constant; Δt: integration time step

The temperature is kept constant (T_0) recurring to the Langevin dynamics [163, 164] that adds two terms to the Newtonian's motion equation (Equation 11): the friction term and the stochastic term (Equation 16).

$$\left[-\gamma \frac{d}{dt}\boldsymbol{r}(t) + \sqrt{\frac{2\gamma k_B T_0}{m}}\delta(t)\right] - \nabla U(\boldsymbol{r}) = m\frac{d^2}{dt^2}\boldsymbol{r}(t)$$

Equation 16. Langevin temperature control added to the Newton equation.
γ: average number of collisions required to damp the highest vibration in about e^{-1}

The friction term ($-\gamma \frac{d}{dt}\boldsymbol{r}(t)$) applies a friction coefficient to correct the friction of the solvent molecules that cannot be described by the force field equations. The stochastic term ($\sqrt{\frac{2\gamma k_B T_0}{m}}\delta(t)$) corrects the statistical error the comes from the deterministic potential-energy functions [144, 161, 165, 166].

1.3.4. Periodic Boundary Conditions

Any molecular system that is study by MD simulations cannot be dissociated by the environment that surrounds it. In the case of proteins, the environment (solvent) is commonly an aqueous solution.

The inclusion of the solvent in these simulations can be accomplished through the use of implicit models of solvation to simulate the presence of the solvent and counterions. This approach adds empirical coefficients to simulate the viscosity of the solvent and also adjust the dielectric constant to mimic the effect of the solvent into the electrostatic interactions. However, the results obtained with implicit solvent simulations have shown issues providing accurate thermodynamics values and also poor realistic trajectories for a major part of the biomolecular systems. Consequently, explicit solvent models are commonly used to study this type of molecular systems. In this case, the target molecular system is placed, typically, inside a cubic water box where the water molecules and the counterions are considered explicitly. However, this approach creates a new challenge regarding the description of the intermolecular interactions on the box's boundary.

The Periodic Boundary Conditions (PBC) was created to overcome this issue, providing a simple way to describe this type of interaction without a significant impact on the computational cost. According to the PBC approach, the molecular system is a central spot in an infinite cubic network of cell replicas. Accordingly, the motion equations are only calculated for the central unit, and the movements of each atom are replicated among all the other replica cells. Despite this is not a perfect approach, it is a good approximation to overcome this pitfall [144, 145].

When PBC is used, some considerations need to be taken into account to avoid some abnormal behavior. In particular, it should be avoided the interaction of the biomolecular system with its image in a neighbor replica cell. Therefore, a minimum distance between the biomolecular system and the face of the water box needs to be kept, and it needs to be bigger than a custom non-bond interactions cutoff. Considering a non-bonded interactions cutoff of 10 Å, it would be reasonable to set a 12 Å minimum distance between the biomolecular system and the boundary of the solvent box. In the case of highly charged molecular systems, the size of the box should be increased to keep the cohesion of the model.

1.3.5. Non-Bonded Interactions Cutoff

As mentioned before, the non-bonded interactions are calculated as a contribution of short and long-range interactions. The short-range interactions are normally described

by a 12-6 Lennard-Jones potential. According to the mathematic expression (Equation 7), we can see that the highest term varies according to the r_{ij}^{6}, which means that its energy contribution is rapidly damped at shorted distances. On the other hand, the Coulomb's term is proportional to r_{ij}^{-1}, making its contribution more significant for greater distances. Consequently, more atoms have a non-neglectable influence on the total energy of the system [144, 145].

The perfect approach would be the calculation of all the intermolecular interactions of the system, but the PBC approach makes the number of atoms in the system infinite. Consequently, the number of intermolecular interactions accounted for each atom of the system needs to be reduced to a finite number. The common approach is to establish a radial cutoff that limits the accounted interactions for each atom. For example, if a 10 Å cutoff is set, only the atoms whose distance is ≤ 10 Å are included in the calculation.

This approximation allows a good commitment between accuracy and computational cost, but the cutoff needs to be chosen wisely to avoid aberrant results and misinterpretations.

This section summarizes the fundaments and general protocol for a biomolecular study using MM, in particular, the steps involved in the energy minimization and molecular dynamics. Figure 11 summarizes the common workflow for the preparation, energy minimization, and MD simulation of a biomolecular system, based on the theory discussed so far. The MD simulation stage is usually divided into two main parts: the equilibration and the production. During the equilibration phase, the system's temperature is gradually increased using the canonic ensemble. Afterward, the simulation is prolonged using the isothermal-isobaric ensemble at constant temperature (production phase).

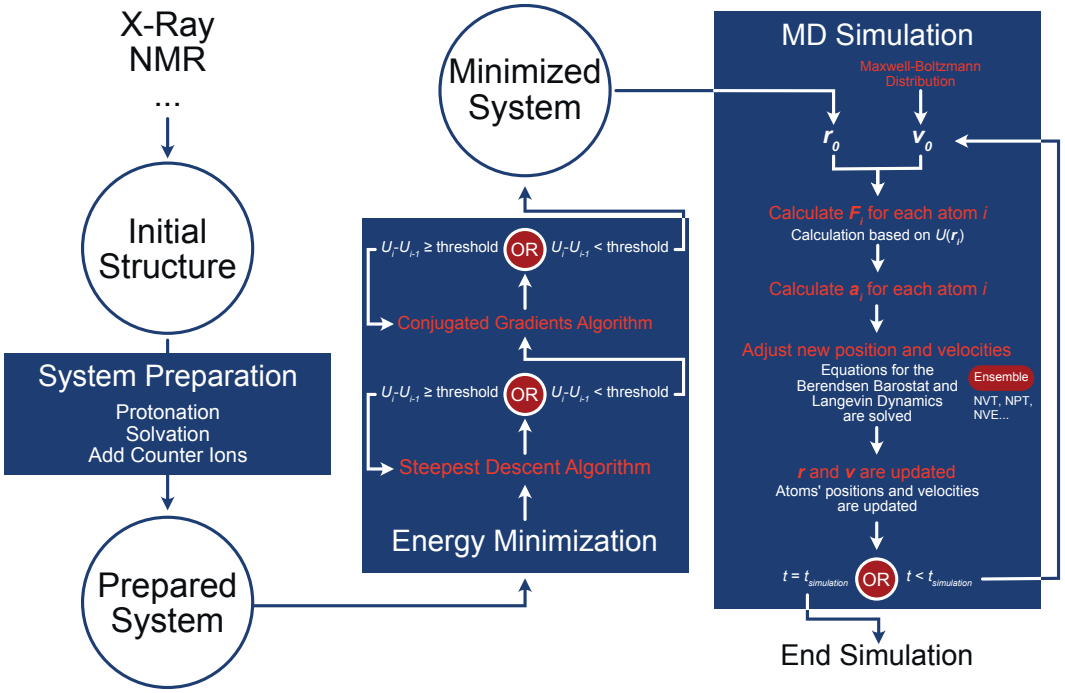


Figure 11. Flowchart for the preparation, energy minimization, and MD simulation of a biomolecular system.

2. Quantum Mechanics

Classical MD simulations are a powerful methodology to understand the time-evolution of complex molecular systems using MM. Although classical MD simulations are able to provide an overview of conformational rearrangements of big molecular systems at a low computational cost, they are not able to describe important chemical processes that explicitly involve electrons, such as chemical reactivity and excited states. Processes that require an electronic description need to be studied using QM calculations, namely *ab initio* methods [144, 167-170].

Although QM methods provide a huge step forward in terms of accuracy, this advantage entails a counterpart. Due to the current hardware and algorithm limitations, the QM methodologies can only be applied to molecular systems with no more than a few hundred atoms [171].

The application of the QM to study chemical processes involves the application of the wave function to the electronic configuration of the system. However, the wave function equation has a critical limitation since it cannot be solved analytically for many-electron systems [168, 169]. Consequently, through the years, several methods have been developed to overcome this limitation, searching for the best approximations that provide a good commitment between accuracy and computational cost.

Hartree-Fock (HF) was one of the first wave function-based methods developed to be applied in many-electrons systems. However, HF methods are not capable of accurately account for electron correlation [144, 167-170]. Several post-HF methods were proposed to overcome this limitation, in particular, Møller-Plesset (MP) Perturbation Theory and Coupled-Cluster (CC) Theory [144, 167-170]. These last methods provide an incredible advance in terms of the accuracy of QM calculations, but they are computational very demanding. Therefore, their application is limited to significantly smaller molecular systems.

In the 70s, Density Functional Theory (DFT) method arose to ally the accuracy of the QM description and calculation speed. This method is not based on the wave function equation but on the electron density of the system. Although the perfect exchange-correlation density function is not known yet, several DFT functionals are available using different approaches to include this contribution.

Each one of these QM methodologies is going to be discussed in more detail in the next sections of this chapter.

2.1. Wave Function

The discovery of subatomic particles and consequent study of their properties were the main driving force for the development of the QM. The experiments targeting these particles, particularly electrons and photons, showed that classical equations are not suitable to describe the observed phenomena.

The work of several scientists as Einstein, De Broglie, Planck, Bohr, Schrödinger, among others, postulated theories about QM, which were essential to describe the observations of subatomic phenomena. The beginning of the QM is assigned to Plank's study of the light emitted by heated solids in 1900. However, other previous studies were also essential for this development, namely when Thomas Young described the wave nature of light, James Maxwell unified the electricity and magnetism laws, and Hertz detected radio waves produced by accelerated electric charges [172].

An important mark was set when it was discovered that all electromagnetic waves travel at a constant speed in vacuum, 2.998×10^8 m/s. Consequently, a relation between the frequency (υ) and the wavelength (λ) was established based on **Equation 17**.

$$c = \upsilon\lambda$$

Equation 17. Relation between the frequency and the wavelength of a wave in vacuum.

As a consequence of the blackbody studies, Plank postulated a relation between the energy of electric charges of the blackbody and the frequency of the emitted radiation (υ) [173]. By fitting several experimental blackbody curves, Plank found that this relation depends on a constant h = 6.6×10^{-34} J.s (Plank's constant). Moreover, he also showed that the energy is quantized in discrete values rather than a continuum spectrum.

This concept was, later on, applied by Einstein to explain the photoelectric effect [174]. Einstein proposed that light is composed of particle-like entities, called photons. When a photon is absorbed by an electron in the metal, part of the photon energy is used to free the electron from the nuclear attraction, and the remaining energy appears as the kinetic energy of the ejected electron [168]. Considering that the energy of a photon is given by the following equation:

$$E_{photon} = hv$$

Equation 18. Energy of a photon.
h: Plank's constant; v: radiation's frequency

The energy of the photon (E_{photon}) is equal to the sum of the minimum energy required by an electron to escape from the ground state ($\Phi_{electron}$) and the maximum kinetic

energy of that emitted electron ($T_{electron}$) (Equation 19). Therefore, an increase in the light's frequency (v_{light}) increases the kinetic energy of the emitted electron.

$$hv_{light} = \Phi_{electron} + T_{electron}$$

Equation 19. Relation between the energy of an electron after the incidence of a photon.

On the other hand, the increase in the light's intensity (number of photons per unit of time) increases the number of ejected electrons from the metal.

These developments led to further theorization about the structure of the atom. It was known that the atoms are mainly composed of negative particles (electrons), positive particles (protons), and neutrons with a neutral charge. Since protons and neutrons are considerably more massive than electrons, the electron motion is significantly greater than the atomic nuclei. Therefore, the chemical properties of atoms and molecules are predominantly determined by their electronic configuration.

Over the years, some atomic models were proposed, in particular the one suggested by Niels Bohr in 1913 [168, 175, 176]. Bohr postulated that the electron in a hydrogen atom can only assume certain energy levels – the electron's energy is quantized. Therefore, the transitions between different energy levels are associated with the absorption of a discrete amount of energy in the form of light, according to Equation 18. Although the model was able to describe the transition in the hydrogen atom, it failed for slightly heavier atoms like helium. Moreover, the model did not account for chemical bonds.

A decade later, L. de Broglie makes an analogy between the wave behavior of light and the electrons [177]. According to his equation (Equation 20), the motion of an electron of mass m and speed v have a wavelength λ.

$$\lambda = \frac{h}{mv}$$

Equation 20. L. de Broglie's equation.

In 1927, de Broglie's theory was corroborated by experimental results where the reflection and diffraction of electros were observed [168]. Later, the same effect was also observed for helium and hydrogen atoms, suggestion that the wave behavior is general to all microscopic particles like large molecules as $C_{48}H_{26}F_{24}N_8O_8$ [178].

The wave-particle duality of matter/light is not easily perceived in the light of classical physics, and it seems to be quite counter-intuitive to the human thinking that is largely "shaped" by macroscopic observations.

In 1927, Heisenberg discovered a limitation imposed by the wave-particle duality principle [179]. The fact that microscopic entities behave as particles and waves makes

it impossible to simultaneously and accurately measure the position and momentum (velocity) of such entities. This limitation was called the uncertainty principle. The uncertainty principle postulates that the observation/measurement interferes with the system, making it impossible to know the exact position and speed of a particle at a given time. According to the principle, the product of the uncertainties of position (Δx) and momentum (Δp) are of the order of magnitude of Plank's constant (Equation 21).

$$\Delta x \Delta p \approx h$$

Equation 21. The uncertainty principle.

All of these works showed that classical mechanics could only be applied to macroscopic particles. Newton's equations failed when applied to describe microscopic phenomena. This means that we cannot predict the time-evolving of a certain microscopic state over time using classical equations. Moreover, according to the Heisenberg uncertainty principle, the impossibility to know the position and speed of a microscopic particle also makes impossible the prediction of future states.

Therefore, to describe the state of a certain system using quantum mechanics, it was postulated a function (Ψ) of the coordinates of the particles called the state function or wave function. Since the wave function evolves with time, it is also a function of time ($\Psi(x,t)$). The wave function contains all possible information about a certain system. The concept of wave function and related equations were introduced, for the first time, by Erwin Schrödinger in 1926 [180]. In a simple example, if we consider one-particle and one-dimensional system, the evolution of the state is given by the following equation:

$$-\frac{\hbar}{i}\frac{\delta\Psi(x,t)}{\delta t} = \frac{\hbar^2}{2m}\frac{\delta^2\Psi(x,t)}{\delta x^2} + V(x,t)\Psi(x,t)$$

Equation 22. Schrödinger equation for one particle and one dimension.
m: mass of the particle; $V(x,t)$: potential-energy function of the system

Equation 22 is the time-dependent Schrödinger equation applied to a particle in a single dimension and corresponds to the first derivative of the wave function with respect to time. The solution of this equation reveals the future state (wave function) at any time *t*, only requiring the knowledge about wave function at time t_0. Thus, the main challenge is to know the wave function that characterizes a certain system at a certain time.

The solution for the Schrödinger equation gives information about the system in the form of a probability since, according to the uncertainty principle, it is impossible to know the exact position and velocity of particles. Consequently, the solution of the equation is a range of probable positions of the particle at a given time *t*. The probability associated with a particle's position is given by the probability density function $|\Psi(x,t)|^2$.

In a more general way, the Schrödinger equation can be written as a product of a quantum operator (Hamiltonian) and the wave function that results in the wave function itself multiplied by the energy value, the eigenvalue. In other words, the $-\frac{\hbar^2}{2m}\frac{\delta^2}{\delta x^2}+V(x,t)$ is the quantum operator or Hamiltonian ($\widehat{H}$) that acts in the eigenfunction ($\Psi(r,t)$) that is the wave function of the system and results in the eigenvalue (energy, *E*) multiplied by the wave function (**Equation 23**) [144, 167-169].

$$\widehat{H}\Psi(r,t) = E\Psi(r,t)$$

Equation 23. General Schrödinger equation.

In this case, the operator $\widehat{H}$ is the sum of two operators for the potential ($\widehat{V}$) and kinetic ($\widehat{T}$) energy terms of the system.

In certain applications, like in quantum chemistry calculations, it is common to use a time-independent derivative of the Schrödinger equation (**Equation 24**). Nonetheless, there are computational methods to address problems where time-dependent equations are absolutely required, e.g., the study of excitation processes and photoabsorption spectra.

This simplification can be applied if no external forces are applied to the system. In this way, the equation can be written depending only on the position of the particle [144, 167-169].

$$\widehat{H}\Psi(r) = E\Psi(r)$$

Equation 24. General time-independent Schrödinger equation.

As a consequence, the probability of finding the particle in a certain position ($|\Psi(r)|^2$) does not change over time. This does not mean that the particle is at rest, but rather it means that the probability density is stationary. When the Schrödinger equation is calculated in a time-independent manner, the result is a stationary state.

Although the Schrödinger equation allows a description of the system and predicts its properties, the equation has two unknown terms: the wave function and the energy. Thus, further considerations are required to solve this problem.

2.1.1. Born-Oppenheimer Approximation

In 1927, the Born-Oppenheimer approximation [181] was introduced to alleviate the problem and allow the application of the Schrödinger equation to complex systems. Since the electron's mass is about 1836 smaller than the atomic nuclei [169], it can be

assumed that the electrons' motion is instantaneously adapted to the nuclei's movement. Consequently, the movement of electrons is (almost) totally uncorrelated with the nuclei's motion. Thus, the Born-Oppenheimer approximation postulated that the wave function of the system (Ψ_{system}) could be given by the product of the wave function of atomic nuclei (Ψ_{nuclei}) and electrons ($\Psi_{electrons}$):

$$\Psi_{system} = \Psi_{nuclei} \times \Psi_{electrons}$$

Equation 25. Born-Oppenheimer approximation.

Considering the application of the operator $\hat{H}$ to the wave functions, the equation can be written as a sum of the nuclear ($\hat{H}_{nuclear}$) and electronic ($\hat{H}_{electronic}$) Hamiltonians:

$$\hat{H}_{system} = \hat{H}_{nuclear} + \hat{H}_{electronic}$$

Equation 26. Born-Oppenheimer approximation as a sum of Hamiltonians.

Both nuclear and electronic Hamiltonians are a sum of potential and kinetic operators for the species involved. Therefore, the electronic Hamiltonian is given by the sum of the potential energy associated with the interelectronic repulsion ($\hat{V}_{electron-electron}$) and nucleus-electron attraction ($\hat{T}_{electron-nucleus}$) as well as the intrinsic kinetic energy of the electrons ($\hat{T}_{electrons}$) (Equation 27).

$$\hat{H}_{electronic} = \hat{T}_{electrons} + \hat{V}_{electron-electron} + \hat{T}_{electron-nucleus}$$

Equation 27. The electronic Hamiltonian.

On the other hand, the nuclear Hamiltonian depends on the kinetics of the nuclei ($\hat{T}_{nuclei}$) and the interatomic repulsion ($\hat{V}_{nucleus-nucleus}$):

$$\hat{H}_{nuclear} = \hat{T}_{nuclei} + \hat{V}_{nucleus-nucleus}$$

Equation 28. The nuclear Hamiltonian.

In the study of most of the chemical reactions, e.g., in biological systems, the electronic rearrangement has the most important role since the chemical reactions are a consequence of the electronic rearrangement. Therefore, the contribution of the nuclei to the total energy of the system can be computed, later on, using a classical approach.

Although the problem has become simpler with the application of the Born-Oppenheimer approximation, the electronic wave function is still unknown.

2.1.2. Linear Combination of Atomic Orbitals

One of the simplest ways to think about the wave function of a multi-electron molecular system is considering the combination of all the atomic orbitals involved. The Linear Combination of Atomic Orbitals (LCAO) is a quantum chemistry method to obtain the

molecular orbitals of a system by simply combine all the individual atomic orbitals – basis sets [144, 167-170]. The only requirement here is to know the wave function for all the atomic orbitals of the system, which is also challenger. The Schrödinger equation (both time-dependent and time-independent) only has an exact solution for hydrogen and hydrogen-like systems, so the equation for the remaining atoms needs to be determined somehow.

In order to overcome this problem, different families of functions arose to reproduce the properties that an atomic orbital must follow. Different basis sets provide different levels of accuracy regarding the description of the orbitals and, therefore, have different computational costs associated.

In principle, an infinite number of basis sets would result in the most accurate wave function of the system. However, the application of an infinite number of functions is not computationally feasible. Therefore, only a finite number of functions are employed, contributing to an additional error associated with the truncation of the basis set. Thus, the main goal here is to keep the number of basis functions low (minimizing the number of two-electron integrals) but in enough amount to keep a reasonable accuracy.

Two of the first basis sets developed to study molecular systems were the Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO). [168]

An STO for an atom with the atomic number Z is given as a function of the effective nuclear charge Z_e and the quantum numbers associated with the orbital: n, l, and m_l:

$$\phi(r,\theta,\varphi) = Nr^{n_e-1}e^{\left(\frac{-Z_e\frac{r}{a_0}}{n_e}\right)}Y_{lm_l}(\theta,\varphi)$$

Equation 29. STO equation.
N: normalization constant; r,θ,φ: spherical coordinates

However, STOs basis sets lack an appropriate description of the orthogonal orbitals with different n. Moreover, the application of STO for large molecules is extremely demanding in terms of computational effort.

Therefore, the GTOs appeared to make *ab initio* calculation in large systems computationally feasible:

$$\phi(r_x,r_y,r_z) = Nx^i y^j z^k e^{-\alpha r^2}$$

Equation 30. GTO equation.
N: normalization constant; r_x,r_y,r_z: cartesian coordinates; i,j,k: non-negative integers; α: positive exponent; r: distance from the electron to the nucleus

The main advantage of GTOs is the mathematical simplicity that allows the calculations to become significantly faster, mainly because a product of two Gaussians at different centers is equivalent to a single Gaussian function centered at a point between those two centers. Consequently, a two-electron integral on three or four different atomic centers can be reduced to integrals of only two centers, which is computationally much more efficient.

However, the GTOs have an important disadvantage regarding the description of electrons near the nucleus. For example, the orbital 1s has a cusp at the atomic center that is well described by the STOs, but GTOs cannot accurately consider it. Since GTOs cannot describe well the atomic orbitals near the nucleus, large basis sets are required to achieve a good level of accuracy.

This problem was also alleviated by the combination of several GTOs to form what is called contracted Gaussian type orbitals (cGTOs). Several cGTOs have been developed through the years, in particular the ones developed by J. Pople [182], R. Ahlrich [183, 184], and T. Dunning [185-187]. Among all of these basis sets, the Pople's 6-31G is still one of the most used ones nowadays.

The 6-31G basis set is a type of cGTO that uses six Gaussian primitives to treat the core electron orbitals, three for the inner valence electrons and one for outer valence electrons [168, 169].

The Pople's basis sets are commonly employed in a more complete form, including polarization functions for heavy or heavy and hydrogen atoms. For example, the 6-31G(d,p) basis set includes five d-type Gaussian orbitals to polarize the heavy atoms and three p-type orbitals for the hydrogen atoms. In addition, diffuse functions can also be included to improve the description of the electron density of ions. Like the polarization functions, the diffuse functions can also be applied to solely heavier atoms (6-31+G) or hydrogen atoms too (6-31++G).

Pople's basis set is one of the most used basis sets, and this happens because a particular combination with some methods, like B3LYP functional (it will be discussed later in this chapter), has shown very good results for a panoply of molecular systems [188-196]. However, the fundamentals that support this family of basis sets have some problems, for example, the absence of functions to describe the electronic correlation. These phenomena and some other relativistic effects are not considered in this type of basis sets. However, they are usually important for accurately describe the electronic behavior of heavier atoms, like transition metals.

Since several enzymes employ transition metals in their active sites, the lack of an appropriate description of these atoms in this type of basis sets compromises their applicability. Moreover, the transition metal atoms have a lot more electrons that require an expansion of the basis functions for the internal core electrons. Otherwise, the valence orbitals are deeply affected by the poor description of the interelectronic repulsion, relativistic effects, and the inner electrons' shield effect.

Fortunately, the two drawbacks mentioned above can be, in some extension, overcame with the application of a function for the core electrons and made only the valence electrons explicitly considered during the calculations. The function that simulates the effect of the core electrons is called Effective Core Potential (ECP) or Pseudo-Potential (PP) [168, 169]. The main advantages of this approximation are:

- the core electrons are treated implicitly, cutting the computational cost drastically when compared with the explicit approach;
- part of the relativistic effects is taken into account without full relativistic calculations, which are computationally demanding.

The available ECPs correspond to a set of functions that are adjusted to fit the results obtained with accurate calculations where relativistic effects are accounted, and large basis sets are employed.

The LanL2DZ ECP [197-201] is one example of this type of potentials to describe the core electrons and it is commonly used to study of the catalytic mechanism of enzymes employing metals in their active site [195, 202-208]. A combination of LanL2DZ and Pople's basis sets can be made to study such type of mechanisms.

Although the mentioned approaches provide accurate results, particularly in terms of geometries when combined with certain functionals, there is a panoply of other basis sets that include a better description of the electrons.

Currently, cc-pVNZ basis sets are widely used to provide a better description of the orbitals and also an affordable computational cost. These basis sets were initially developed by Dunning and then improved to include polarization (cc-**p**VNZ) and electronic correlation description (**cc**-pVNZ). The N can assume a value of D, T, Q, 5, 6…, corresponding to double, triple, quadruple, quintuple, sextuple-zeta…, respectively [185].

The diffuse functions can also be added to cc-pVNZ functionals to allow an ameliorated description of anions and long-range interactions, originating the aug-cc-pVNZ basis sets sub-family.

Unfortunately, cc-pVNZ functionals are considerable computationally more expensive than Pople's one making its application almost forbidden for geometry optimization in big molecular systems. Therefore, it is common, as in this work, that cc-pVNZ basis sets are used to refine the energies through single-point energy calculations. It means that the geometries are optimized using a Pople's basis set, like 6-31G(d,p), and then the energy of the resulting geometry is calculated using this more accurate basis set.

The basis sets provide the required mathematical tools to build initial molecular orbitals for a chemical system through the combination of individual atomic orbitals.

2.2. Hartree-Fock

The exact wave function is only known for the hydrogen atom (one proton and one electron system). For other light atoms as helium and lithium, there are very accurate wave functions that were obtained, including interelectronic distances in the mathematical expressions. However, for higher atomic number atoms, the approach needs to be different. An accurate wave function is obtained through an initial guess of the wave function using the Hartree-Fock procedure [144, 167-169].

The Hartree-Fock method is a fundamental approach to use atomic and molecular orbitals in many-electrons systems (**Equation 31**).

$$\widehat{H} = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2 - \sum_{i=1}^{n}\frac{Z^2}{4\pi\epsilon_0 r_i} + \sum_{i=1}^{n-1}\sum_{j>i}^{n}\frac{e^2}{4\pi\epsilon_0 r_{ij}}$$

Equation 31. Hartree-Fock method.
m_e: electron's mass; Z: effective nuclear charge; r_i: distance of electron *i* to the nucleus; r_{ij}: distance between electrons *i* and *j*

Equation 31 omits some aspects related to the electrons as the interaction spin-orbital, which can be important for atoms with higher atomic numbers. Nonetheless, those terms will be discussed later in this chapter.

2.2.1. Many-Electrons System

Since **Equation 31** has a term for the repulsion of electrons, a many-electrons system can only be treated as a simple sum of n one-electron equations if these repulsions are neglected. In that case, a wave function (ψ) can be written as a product of n equations for one-electron orbitals:

$$\psi = f_1(r_1, \theta_1, \phi_1) \times f_2(r_2, \theta_2, \phi_2) \times \ldots \times f_n(r_n, \theta_n, \phi_n)$$

Equation 32. Wave function as the product of one-electron wave functions.

For the ground state of a given atom, each of the lowest orbitals is filled with two electrons with opposite spin according to Pauli's exclusions principle. However, this approximation has several pitfalls. One of them is the fact that all electrons "feel" the same atomic attraction from the nucleus, discarding the screening effect of the other electrons.

2.2.2. Hartree-Fock Theory

The HF theory was the first method used with success in computational chemistry to treat many-electrons systems. The HF theory describes the orbitals where the electrons move around, and the total wave function is given as a product of all the electron orbitals.

However, the HF theory introduces an approximation regarding the interelectronic repulsion, where the movement of the electrons depends only on the electrostatic field created by the nuclei and the average field of the $n-1$ electrons (other electrons). Consequently, the wave function of each electron of the system can be written as a function of the electrostatic field created by all other electrons [144, 167-169].

As mentioned before, each electron must obey the Pauli exclusion principle. Therefore, each one-electron function ($\phi(r_n)$) needs to be multiplied by the corresponding spin function (α or β):

$$\chi(r_n) = \phi(r_n) \times \alpha$$

Equation 33. One-electron spin orbital function.

Consequently, the HF function (Ψ_{HF}) for a N-electrons systems is given by the product of n spin orbital functions ($\chi(r_n)$):

$$\Psi_{HF} = \chi(r_1) \times \chi(r_2) \times \ldots \times \chi(r_n)$$

Equation 34. HF function as a product of n spin orbital functions.

2.2.3. Slater Determinant

Further consideration must also be taken into account to allow the interchange of electrons in a molecular system considering the HF equation (Equation 34).

According to the Pauli exclusion principle [209], two electrons cannot share the same spin in an orbital. Consequently, the wave function must be antisymmetric in order to allow only the interchange of electrons that would not share the same spin-state in the same final orbital.

In order to achieve such a requirement, the antisymmetric of a wave function can be built from a Slater determinant [210] (Equation 35).

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(r_1) & \chi_2(r_1) & \cdots & \chi_N(r_1) \\ \chi_1(r_2) & \chi_2(r_2) & \cdots & \chi_N(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(r_N) & \chi_2(r_N) & \cdots & \chi_N(r_N) \end{vmatrix}; \langle \chi_i | \chi_j \rangle = \delta_{ij}$$

Equation 35. Slater determinant.

In the Slater determinant, each column corresponds to the spatial single-electron spin wave function while the rows correspond to the coordinates of each electron of the system. Therefore, the Slater determinant represents every possible combination for a N-electrons system.

2.2.4. Variational Principle

Equation 34 provides a good initial guess for a wave function that can define the molecular system. However, it is far from being accurate enough and needs to be improved.

To achieve a more accurate wave function for the system, the variational principle is employed. This principle is based on the fact that every wave function has an energy higher or equal to the exact energy of the system [144, 167-169]. This principle can be mathematically proved and acts as an important tool to improve the wave function of a given system [168]. This means that the wave function can be modified iteratively until we reach an accurate one that fits a pre-defined threshold. In each step, the energy variation can be tested to understand if the new wave function is better or worse than the previous one. In other words, if the energy increases, it means that the new wave function is less accurate than the previous one. If the energy decreases, the new wave function is closer to the exact solution of the hypothetical electronic Schrödinger equation.

The energy of an approximate wave function can be calculated as the expectation value of the Hamiltonian divided by the norm of the wave function:

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

Equation 36. Variational principle.

In terms of the applicability of this method to real systems, Hartree introduces, in 1927, the Self-Consistent Field (SCF) (Figure 12) [211]. The SCF method takes a set of tridimensional atomic nuclei coordinates as input to create an initial guess of molecular orbitals, as mentioned before. Then, the Fock matrix is generated and diagonalized to be tested in terms of SCF convergence. If the SCF test fails to obey the pre-defined thresholds, a new Fock matrix is generated and tested again. Otherwise, the Fock matrix is used to calculate the properties of the system, and the calculations end. The SCF cycle occurs as many times as required to fulfill the criteria or until the software imposed a maximum number of cycles is reached.

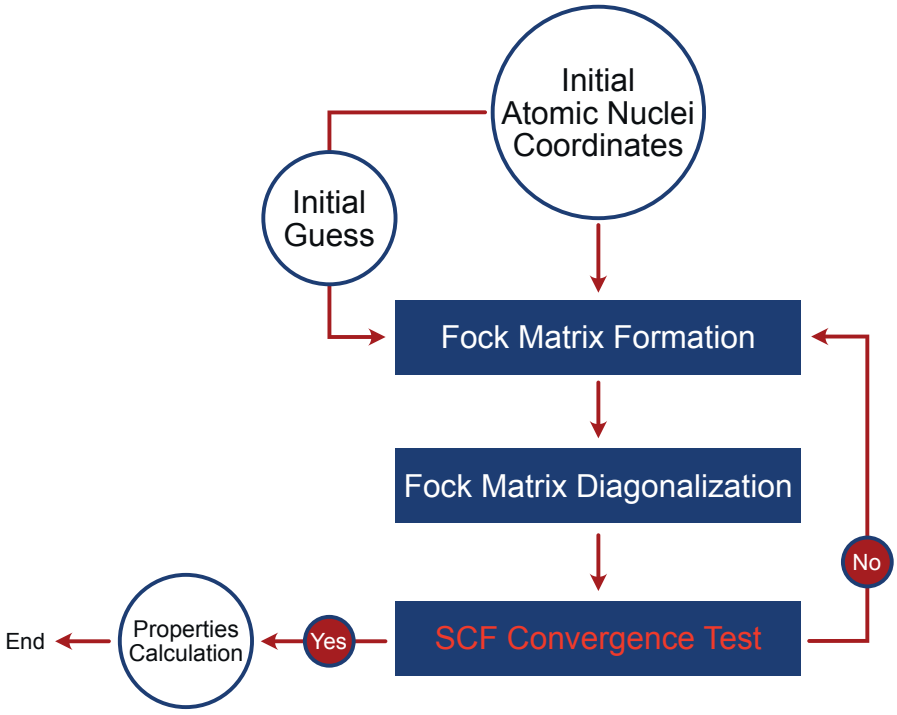


Figure 12. Flowchart of HF SCF method.

2.2.5. Hartree-Fock Limit: Electron Correlation

Based on the assumptions discussed above, the HF equations can be used with additional determinants to account for more electronic phenomena or with additional approximations culminating into semi-empirical methods. The first case leads to the convergence of the wave function to a more accurate solution that becomes closer to the exact solution of the electronic Schrödinger equation. In the second case, the semi-empirical methods are considerably cheaper (from the computational point of view) and resort to additional empirical parameters from experimental data [144, 167-169].

Although the HF method provides a simple way to test and achieve accurate geometries and properties due to the SCF approach, this is only true for “well-behaved” molecular systems. This means that the good results obtained with HF are limited to closed-shell electronic configurations and systems with low degeneracy of molecular orbitals. Moreover, HF treats electrons as non-interacting particles, which is not real since the electrons interact repulsively with each other. Therefore, there is an energy difference between the one obtained with HF and the real one that includes this type of interactions – the correlation energy. The correlation energy is a term that allows compensating the overestimation of the electron repulsion by the Coulomb integrals employed by the HF methods [168].

For close-shell systems, the correlation energy assumes a very small contribution, justifying the good results obtained with HF. However, when reactions involve open-shell systems or electron transfers, the correlation effect adopts a more significant role in the chemical reactivity. In this case, the electron correlation is essential for an accurate description of the chemical system.

2.3. Post-Hartree-Fock Methods

HF methods were the first attempt to try applying the Schrödinger equation to more complex molecular situations that are not based on the simplest hydrogen-like systems. However, the required approximations included in the methods make them unsuitable to accurately address the study of most of the important chemical phenomena. The main problem of HF is the fact that it relies on the average contribution of electrons. Therefore, the interactions between the electrons are not treated explicitly, neglecting the electron correlation effect.

Consequently, the post-Hartree-Fock (post-HF) methods were developed to refine the HF approach, especially giving attention to better techniques to describe the electron correlation.

2.3.1. Configuration State Methods

When a finite basis set is used with the HF method, a finite set of molecular orbitals is defined. This means that a basis set with N_0 members have N_0 molecular orbitals. If we order the N_0 molecular orbitals from lowest to highest energy, half of them will be doubly occupied with the electrons of the system. This state corresponds to the ground-state wave function. However, there is still a half of virtual unoccupied antibonding orbitals that are empty [168].

In the case of one electron migrates from one occupied spin-orbital to a virtual unoccupied spin-orbital, a singly excited state is obtained. In the same way, if a second electron follows the same behavior but for a different virtual unoccupied spin-orbital, a double excited state is achieved. This means that there are multiple possibilities for excited states that respect the required electronic symmetry. All or part of these states, also known as determinants, can be linearly combined to form a configuration state function (CSF). [168]

The exact ground-state wave function (Ψ) can be expressed as a linear combination of all possible Slater determinants arising from a complete set of spin orbitals:

$$\Psi = c_0\Psi_0 + \sum_{a,p} c_a^p \Psi_a^p + \sum_{\substack{a<b \\ p<q}} c_{ab}^{pq} \Psi_{ab}^{pq} + \sum_{\substack{a<b<c \\ p<q<r}} c_{abc}^{pqr} \Psi_{abc}^{pqr} + \cdots$$

Equation 37. Exact ground-state wave function.
c_X: expansion coefficients; a, b, c ...: limits in the summation indices

The summation indices limit the application only for single excited determinants. The *ab initio* methods based on this approach are called configuration interaction (CI) methods.

The energy calculated with the exact ground-state wave function computed using a CI method is the exact non-relativistic ground-state energy. Accordingly, the difference between this exact energy and the one computed with HF and an infinite basis set corresponds to the correlation energy.

Although the CSF provides an exact wave function for a specific system, it is computationally impossible to solve it considering an infinite basis set of Slater determinants constructed with an infinite number of spin-orbitals. This cumbersome can be alleviated if only the determinants for a certain symmetry are included in the calculation. Then, considering all CSFs for a certain symmetry and a finite basis set, the full CI is obtained, which corresponds to the best CI calculation that can be performed for that specific basis set.

Therefore, for a CI calculation at a given basis set, the correlation energy can be calculated through the difference between the HF SCF and full CI calculations using the same basis set. If the basis set is enlarged in both calculations, the resulting basis set correlation energy becomes closer to the exact value [168].

Despite the promising results obtained with the CI method, its application is computationally too demanding for application in interesting chemical problems. In order to turn the calculations “lighter”, truncated versions of CI methods were developed to account only the ground-state and doubly excited states (DCI) or ground-state, singly and doubly excited states (SDCI), respectively.

Notwithstanding, the lack of size consistency is also pointed out as an important defect of CI calculations. This means that the energies computed with CI methods are not proportional to the number of electrons in the system. As a consequence, the energy of two particles infinitely separated is not equal to the sum of the energy of each particle individually. The magnitude of the size consistency error increases as the size of the molecular system increases [144, 167-169].

2.3.2. Multiconfiguration SCF Methods

In the CI methods, the expansion coefficients c_X (Equation 37) are initially calculated using HF and then are kept settled through the entire CI calculation. In a refined variation of the CI method, the multiconfigurational self-consistent field (MCSCF) method, the expansion coefficients are also optimized during the calculation. The inclusion of such additional calculations makes MCSCF computationally arduous but allows obtaining good results with the inclusion of a smaller number of CSFs.

The complete active space self-consistent field (CASSCF) method is an example of a method based on the MCSCF approach, and it is particularly important to study excited states. In this method, the spatial wave functions are split into three classes: inactive, virtual, and active orbitals.

The inactive orbitals are composed of the lowest energy wave functions doubly occupied, whereas the virtual orbitals are the highest energy ones that are fully unoccupied. A set of active orbitals are also set to include orbitals with intermediate energy and do not fall in one of the previous two categories. In this way, the CSFs included in a CASSCF calculation depend only on the possible states of active electrons for the considered active orbitals.

Therefore, the success of the application of CASSCF highly depends on the choice of active orbitals. One possible approach is to support the choice of CSFs on the molecular orbital theory, considering all the bonding, non-bonding, and antibonding orbitals as active orbitals in the CASSCF calculation [168].

Despite the improvement of MCSCF methods in terms of accuracy using few CSFs compared with CI calculations, the problem of size consistency was not solved.

2.3.3. Møller-Plesset Perturbation Theory

As referred, the main disadvantage of the CI methods is the lack of size consistency. The Perturbation Theory (PT) was developed to provide an alternative approach to find the correlation energy in a consistent size manner [168].

However, the PT method has the drawback of not being based on the variational principle. Consequently, there are no guarantees that the calculated energies are always higher than the exact energy of the system. Since the energies could be higher or lower than the exact energy, there is no systematic way to ensure if the result is evolving in the direction of the exact solution.

The application of the PT method to a system with many interacting particles is called Many-Body Perturbation Theory (MBPT). This approach was firstly introduced by C. Møller and M.S. Plesset in a method called Møller–Plesset Perturbation Theory (MPPT) [212].

In the MPPT methods, the zero-order Hamiltonian (H_{HF}) from the HF-SCF method is used, and it can be written as a sum of one-electron Fock operators (f_i):

$$H_{HF} = \sum_{i=1}^{N_e} f_i$$

Equation 38. Zero-order Hamiltonian of the HF-SCF method.
i: electron; N_e: number of electrons of the system

The perturbation (H_p) is calculated by the difference of the electronic Hamiltonian (H) and the zero-order Hamiltonian (H_{HF}):

$$H_p = H - H_{HF}$$

Equation 39. Perturbation contribution.

Thus, for electron i, the perturbation ($H_p(i)$) is given by:

$$H_p(i) = j_o \sum_j \frac{1}{r_{ij}} - \sum_m (2J_m(i) - K_m(i))$$

Equation 40. Perturbation contribution for electron i.
j and i: electrons; m: occupied molecular orbitals; J_m: Coulomb operator; K_m: exchange operator

In sum, the MPPT method takes the advantages of HF methods and try to include a technique to introduce the electron correlation contribution through a set of perturbations to the Fock operator.

The perturbation equations can be employed considering different integer orders from zero to n, but the zeroth and first-order equations (MP0 and MP1) give a null contribution for the energy calculation [168]. However, for higher-order perturbation equations (MP2, MP3, MP4…), the contribution of the electron correlation start to become properly accounted for the Hamiltonian. The computational cost associated with MP calculations increased with the number of basis functions employed, which makes some high order MP methods prohibited for certain systems due to their sizes.

Despite the size consistency of the MP methods, the fact that the method does not rely on variational principle makes unpredictable the effect of the application of more complete basis sets[8]. Moreover, higher-order MP methods do not necessarily provide better results, even though the accounted electron correlation contribution increases as well.

[8] In other methods, the computed results are closer to the exact solution as more complete is the basis set employed.

2.3.4. Coupled-Cluster Theory

The Coupled-Cluster (CC) theory is another alternative to CI and MMPT approaches presented so far [213]. Similarly to MMPT, the CC method is size consistent, but it does not follow the variational principle.

The CC method introduces the cluster operator (C) that allows the establishment of a relation between the exact electronic wave function (Ψ) to the HF wave function (Ψ_0):

$$\Psi = e^{C}\Psi_0$$

Equation 41. Wave function as a function of the HF wave function.

The exponential operator (e^C) of Equation 41 is defined by a series expansion (Equation 42) that corresponds to the sum of the effects of a one-electron, two-electron, three-electron, …, N-electron excitation operators (Equation 43) [168].

$$e^{c} = 1 + C + \frac{1}{2!}C^2 + \frac{1}{3!}C^3 + \cdots + \frac{1}{N!}C^N$$

Equation 42. Exponential operator.

$$C = C_1 + C_2 + C_3 + \cdots + C_N$$

Equation 43. Excitation operator.

The application of the CC operator to the wave function of each excitation state follows the same type of approach described for CI methods (Equation 37). Recalling Equation 41, the effect of the exponential operator can be understood as $C_1\Psi_0$, $C_2\Psi_0$, $C_3\Psi_0$, …, $C_N\Psi_0$ components but also $C_1C_1\Psi_0$, $C_1C_2\Psi_0$, and $C_1C_2C_3\Psi_0$, for example. Taking the $C_2\Psi_0$ and $C_1C_1\Psi_0$ cases as example, both represent double excited determinants. However, $C_2\Psi_0$ represent a “connected” double excitation determinant whereas the $C_1C_1\Psi_0$ corresponds to a “disconnected” contribution.

Since the application of the CC method is computationally demanding, it is common the truncation of the cluster operator. Consequently, the method can include only the double excitation determinants (CCD), the single and double excitation determinants (CCSD), or the single, double, and triplet excitation determinants (CCSD(T)) [144, 167-169].

The mathematic behind the cluster operator allows performing more accurate calculations with the same computational effort required for methods like CI and MPPT. In particular, the CC methods allow the inclusion of triplet excitation determinants (CCSD(T)) at a relatively affordable computational cost. Consequently, the CCSD(T) methods is, nowadays, considered the “gold standard” for accurate computational calculations, with an error lower than 1 kcal/mol [214-219]. Therefore, it is common to use CCSD(T) as a reference in benchmark studies of other functionals and/or methods.

In parallel with the MP methods, the main disadvantage of the CC method is the fact that it is a non-variational method with the consequences mentioned above.

2.3.5. Domain Based Local Pair Natural Orbital Coupled-Cluster Method

Although the CC methods, as CCSD(T), provides a good commitment between accuracy and computational cost, they are still highly demanding for systems with more than few dozens of atoms. Consequently, the application of CCSD(T) requires an enormous computational effort when biological systems are studied.

Over the years, several different approaches were developed to alleviate the computational cost associated with the use of the CCSD(T) method, using local correlation approaches, reduction of the slow convergence of the electron correlation energy, and hybrid methods of explicit and local correlation [220-231]. Recently, Frank Neese and co-workers developed a new approach called domain-based local pair natural orbital (DLPNO) that benefits of the locality of the electron correlation component of CCSD(T) method. The DLPNO-CCSD(T) method estimates the pair correlation energy and classifies electron pairs as weakly or strongly correlated, according to an energy threshold value. Then, the strongly correlated electrons are computed using the CC method, and the weakly ones are computed using the MP2 approach. The DLPNO approach also includes a set of controllable thresholds to adjust the accuracy of the method. Studies have reported that the correlation energy computed by DLPNO-CCSD(T) method is 99.9 % of the canonical CCSD(T) calculation with a much more affordable computational cost. In addition, the DLPNO-CCSD(T) method scales almost linearly the size of the system [232-236].

A set of benchmarks studies have shown that DLPNO-CCSD(T) is able to calculate energies with differences below 1 kcal/mol when compared with the CCSD(T) method [235-237]. Consequently, the DLPNO-CCSD(T) has been applied in several studies of catalytic mechanisms of enzymes, with more than one hundred atoms quantum treated [232-236].

2.4. Density Functional Theory

Although several *ab initio* methods have been developed based on the Schrödinger equation, the truth is that their range of applicability is very narrow due to the limited computational power. These limitations impose restrictions in the size of the molecular systems studies that should not surpass some dozens of atoms. Thus, the application of such accurate methods is not suitable for the study of large system, like the biologically relevant ones.

One of the first tries to study bigger systems was to add empirical parameters able to reduce the number of mathematical operations. This implementation originates several semi-empirical methods that are very sensitive to the systems composition and provide very poor results [144, 167-169].

The Density-Functional Theory (DFT) [238-240] completely revolutionized the computational chemistry. This method was developed by L.H. Thomas and E. Fermi in 1927 [238, 239], and it is not based on the Schrödinger equation. Contrariwise, DFT is based on a different property, the electron density ($\rho(r)$), which is a function based on the position and atomic number of the nuclei and the total number of electrons of the system. In other words, a N-electrons systems can be described as a single continuous electron distribution under the influence of the external potential of all positively charged atomic nuclei. This means that the kinetic and potential energies of electrons can be directly related to the electron density.

Since nuclei are point charges, they could be described as local maxima in the electron density ($\rho(r)$). Thus, for an atom A with an atomic charge Z_A, the electron density at the nucleus region $\rho(r_A)$ is given by:

$$\left.\frac{\delta\bar{\rho}(r_A)}{\delta r_A}\right|_{r_A=0} = -2Z_A\rho(r_A)$$

Equation 44. Electron density at the nucleus.
$\bar{\rho}(r_A)$: spherically averaged density

2.4.1. Hohenberg-Kohn Theorems

The applicability of DFT is possible in terms of quantum mechanics because of two fundamental theorems by Hohenberg and Kohn [240].

The first theorem, also called the existence theorem, asserts that in a non-degenerated system, the ground state energy can be solely defined by the electron functional. The theorem can be proven as giving the nuclei's positions and the number of electrons, the Hamiltonian is known, and consequently the ground-state wave function.

The second theorem, also called the variation theorem, states that the electron density follows the variational principle. In line with what was already mentioned, it is possible to prove that the energy of a candidate density function is always higher or equal to the real ground-state energy [168]. Like other variational-based methods, DFT can be applied in an iterative manner, searching for density functions that lower the energy in each step.

2.4.2. Born-Oppenheimer

The two theorems mentioned before support the application of DFT in quantum mechanics. The DFT method allows to significantly reduce the number of variables that need to be computed when compared with other wave function-based methods. Consequently, the calculation speed is one of the main advantages of DFT compared to other *ab initio* methods. Since DFT does not require four variables for each electron (x, y, and z coordinates and spin), the mathematical operations are considerably simple to compute.

Moreover, the Born-Oppenheimer approximation can also be applied to DFT equations. Thus, the electronic DFT energy (E_{DFT}) is computed as the sum of the kinetic energy of electrons ($T_{electrons}[\rho(r)]$), the potential energies associated with the nucleus-electron attraction ($V_{nucleus-electron}[\rho(r)]$), and electron-electron repulsion ($V_{electron-electron}[\rho(r)]$):

$$E_{DFT\ electronic} = T_{electrons}[\rho(r)] + V_{nucleus-electron}[\rho(r)] + V_{electron-electron}[\rho(r)]$$

Equation 45. Born-Oppenheimer approximation applied to DFT method.

Both nucleus-electron and electron-electron potential energies are calculated using classical approaches as described by Equation 46 and Equation 47.

$$V_{nucleus-electron}[\rho(r)] = \sum_{A=1}^{M} \int \frac{Z_A}{|r - r_A|} \rho(r)\, dr$$

Equation 46. Potential energy nucleus-electron according to DFT method.
M: total number of atoms of the system; Z_A: atomic charge of atom A

$$V_{electron-electron}[\rho(r)] = \frac{1}{2} \int \int \frac{\rho(r_i)\rho(r_j)}{|r_i - r_j|}\, dr_i\, dr_j$$

Equation 47. Potential energy electron-electron according to DFT method.

2.4.3. Kohn-Sham Equations

Until now, only the formalism behind DFT has been presented. However, no guidance has been described to explain how DFT could be really applied. The first approach was introduced by W. Kohn and L.J. Sham in 1965 [241, 242]. They proposed equations to

obtain the electron density of a system from a set of one-electron equations that do not interact with each other. In this hypothetical reference system, the non-interacting electrons (N_{nie}) "feel" an external potential $v_{ref}(\boldsymbol{r})$ that is chosen in order to make the reference electron density ($\rho_{ref}(\boldsymbol{r})$) identical to the real electron density ($\rho(\boldsymbol{r})$). Based on that, the Hamiltonian for the reference system (h_{ref}) is given by

$$h_{ref} = \sum_{i=1}^{N_{nie}} h_i^{KS}\,;\; h_i^{KS} = -\frac{\hbar^2}{2m_e}\nabla_i^2 + v_{ref}(\boldsymbol{r_i})$$

Equation 48. Hamiltonian for the reference system.

The one-electron Kohn-Sham orbitals (Ψ_m^{KS}) are eigenfunctions of the one-electron Kohn-Sham Hamiltonians (h_i^{KS}):

$$h_i^{KS}\Psi_m^{KS}(i) = e_m^{KS}\Psi_m^{KS}(i)$$

Equation 49. Kohn-Sham orbitals.

Although the equations for DFT seem similar to the ones applied in HF methods, there are some differences regarding the calculation of electron-electron interactions. While in HF methods, the Fock operator is applied to all electron-electron pairs using two operators, in DFT, the interaction is calculated using a one-body operator. In DFT, the electron-electron interactions (Coulomb repulsion and exchange-correlation) are calculated from the interaction of each electron with the electron density. This approach dramatically decreased the number of integrals that need to be solved, making DFT calculations much faster than HF-based methods (for the same accuracy).

2.4.4. Exchange-Correlation Density Functionals

Although the DFT Hamiltonian is exact, Equation 48 is not accurate enough to provide good results because it does not account for the interaction between electrons. Consequently, the exchange and correlation effects functions ($v_{XC}[\rho]$) need to be added to the Hamiltonian in some way. Throughout history, several different approximate schemes have been developed to address this DFT issue. These approximations were necessary to apply the DFT method, but its implementation became almost semi-empirical, even though the formalism is not.

Nowadays, there are several different functionals that tried to minimize the main source of error in DFT, which is the calculations of the exchange-correlation energy ($E_{XC}[\rho]$). This contribution can be calculated as a sum of contributions of both exchange energy ($E_X[\rho]$) and correlation energy ($E_C[\rho]$):

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho]$$

Equation 50. Exchange-correlation energy.

The exchange term assumes a more preponderant role than the correlation since it usually has a bigger contribution for the energy [168].

The required usage of approximations to account for the exchange-correlation effect arises an important drawback of the DFT method. The variational method can still be applied, but the resulting energies can no longer be assumed as always higher or equal than the real energy of the system. This means that the quality of the results can no longer be compared by how lower the energy of the system is.

As mentioned, there are a lot of DFT functionals that were adjusted to reproduce accurate results. However, the adjustments included in the exchange-correlation functions unviable the development of perfect functionals. Notwithstanding, there are some properties that a "perfect" functional should obey:

1. Self-interaction free: the exchange energy ($E_X[\rho]$) for a one-electron system, like the hydrogen atom, should cancel the Coulomb energy and the correlation energy ($E_C[\rho]$) should be zero;
2. The coordinate scaling of the exchange energy ($E_X[\rho]$) should be linear;
3. No direct scaling rule applies for correlation energy ($E_C[\rho]$) but it should increase when the electron coordinates are scaled by a factor larger than 1;
4. The exchange potential should be asymptotic ($-\frac{1}{r}$) as $r \to \infty$;
5. The correlation potential should be asymptotic ($-\frac{\alpha}{2r^4}$) where α is the polarizability of the $N_e - 1$ system.

Hence, the development of DFT functionals requires adjusting a set of parameters in order to fulfill the above criteria and/or to fit experimental data. This kind of approach causes variability in terms of the accuracy of the DFT functionals for different molecular systems. In addition, there is no "gold" DFT functional like, e.g., the CCSD(T) in *ab initio* methods.

Hereupon, the remaining topic will be focused on some DFT functionals that have been developed and generally used by the scientific community. Most of the methods are used based on their popularity in terms of the number of studies that made use of them. The philosophy behind this is that they were tested several times and, therefore, they should be better [168].

By definition, the exchange energy ($E_X[\rho]$) is given by the sum of the contributions of all electron from α ($E_X^{\alpha}[\rho_{\alpha}]$) and β ($E_X^{\beta}[\rho_{\beta}]$) spin densities:

$$E_X[\rho] = E_X^{\alpha}[\rho_{\alpha}] + E_X^{\beta}[\rho_{\beta}]$$

Equation 51. Exchange energy.

On the other hand, the correlation term takes into account the interaction between all electrons in a spin-independent way:

$$E_C[\rho] = E_C^{\alpha\alpha}[\rho_{\alpha}] + E_C^{\beta\beta}[\rho_{\beta}] + E_C^{\alpha\beta}[\rho_{\alpha}, \rho_{\beta}]$$

Equation 52. Correlation energy.

In both Equation 51 and Equation 52, the total density function of the system is given by the sum of ρ_{α} and ρ_{β} contributions. In a closed-shell singlet system (when spin multiplicity[9] is 1), the ρ_{α} and ρ_{β} contribution is identical.

Commonly, the equations are re-written as functions of the spin polarization (ς) (Equation 53) and the radius of the volume containing one electron (r) (Equation 54).

$$\varsigma = \frac{\rho_{\alpha} - \rho_{\beta}}{\rho_{\alpha} + \rho_{\beta}}$$

Equation 53. Spin polarization.

$$\rho = \frac{3}{4\pi r^3}$$

Equation 54. Volume containing one electron.

Contrary to *ab initio* methods, where the quality of the methods depends on how the electron correlation is accounted, DFT does not follow the same principle because most of the exchange-correlation energy is accounted empirically. Among the methods developed to calculate the exchange-correlation energy in DFT, Local Density Approximation (LDA), Generalized Gradient Approximation (GGA), and meta-GGA methods are the most popular ones [144, 167-169].

[9] The split multiplicity is given by 2S+1, where S is total spin of the system.

Local Density Approximation

The LDA assumes that the density can be locally treated as a uniform electron gas whose function slowly varies:

$$\varepsilon_X^{LDA} = -C_X \rho^{\frac{1}{3}}$$

Equation 55. Exchange energy based on LDA.
ε_X^{LDA}: exchange density for a given particle

Equation 55 is applicable for closed-shell systems, but a more general formula can be written taking into account the electron's spin, originating the Local Spin Density Approximation (LSDA):

$$\varepsilon_X^{LSDA} = -C_X \left(\frac{1}{2}\left[(1+\varsigma)^{\frac{4}{3}} + (1-\varsigma)^{\frac{4}{3}}\right]\right)\rho^{\frac{1}{3}}$$

Equation 56. Exchange energy based on LSDA.

For the correlation term, Perdew and Wang (PW) [243] have proposed a mathematical expression to calculate its contribution, including six fitting parameters:

$$\varepsilon_C^{PW}(x) = -2a\beta(1+\alpha x^2)\ln\left(1+\frac{1}{2a(\beta_1 x+\beta_2 x^2+\beta_3 x^3+\beta_4 x^4)}\right)$$

Equation 57. Correlation energy based on PW method.
$x = \sqrt{r}$; r: electron position; $a, \alpha, \beta_1, \beta_2, \beta_3, \beta_4$: empirical parameters

Although the LSDA and PW methods provide an alternative to overcome the problem of electron exchange-correlation energy drawback of DFT, they are far from being perfect. Usually, LSDA computed energies are approximately 10% underestimated, which is enough to completely hide the effect of electron correlation.

Generalized Gradient Approximation

One of the first attempts to improve the quality of the LSDA method was to consider the system as a non-uniform electron gas. In this assumption, the electron exchange and correlation functions will take into consideration not only the electron density but also their derivatives.

The preliminary tests on this attempt have shown that the exchange term benefits from the inclusion of the first-order derivative of the electron density.

However, this is not true for the correlation correction. In sum, the direct inclusion of first-order derivatives of the electron density resulted in worst results than the original LSDA method.

Consequently, the first-order derivative needs to be included in a different way, as it occurs in Generalized Gradient Approximation (GGA) methods. In GGA, the derivate is added as a function in a way that the Fermi and Coulomb holes integrate to -1 and 0, respectively [168].

Axel D. Beck [244] proposed one of the first GGA exchange methods to correct the LSDA exchange energy – the B or B88 method:

$$\varepsilon_X^B = \varepsilon_X^{LSDA} + \Delta\varepsilon_X^B;\ \Delta\varepsilon_X^B = -\beta\rho^{\frac{1}{3}}\frac{x^2}{1+6\beta x\sinh^{-1}x};\ x = \frac{|\nabla\rho|}{\rho^{4/3}}$$

Equation 58. Exchange energy according to B method.
β: fitting parameter

The B exchange functional allows reduce the LSDA exchange in almost 2 orders of magnitude, leading to a huge improvement.

Similarly, some GGA functional have been proposed to improve the correlation energy. One of the most popular ones is the LYP function from Lee, Yang, and Parr [245]:

$$\varepsilon_C^{LIP} = -4a\frac{\rho_\alpha\rho_\beta}{\rho^2\left(1+d\rho^{\frac{1}{3}}\right)}$$

$$-ab\omega\left[\frac{\rho_\alpha\rho_\beta}{18}\left[144\left(2^{\frac{2}{3}}\right)C_F\left(\rho_\alpha^{\frac{8}{3}}+\rho_\beta^{\frac{8}{3}}\right)+(47-7\delta)|\nabla\rho|^2\right.\right.$$

$$\left.-(45-\delta)\left(|\nabla\rho_\alpha|^2+\left|\nabla\rho_\beta\right|^2\right)+2\rho^{-1}(11-\delta)\left(\rho_\alpha|\nabla\rho_\alpha|^2+\rho_\beta\left|\nabla\rho_\beta\right|^2\right)\right]$$

$$\left.+\frac{2}{3}\rho^2\left(|\nabla\rho_\alpha|^2+\left|\nabla\rho_\beta\right|^2-|\nabla\rho|^2\right)-\left(\rho_\alpha|\nabla\rho_\alpha|^2+\rho_\beta\left|\nabla\rho_\beta\right|^2\right)\right];$$

$$\omega = \frac{e^{-c\rho^{-1/3}}}{\rho^{\frac{14}{3}}\left(1+d\rho^{-\frac{1}{3}}\right)};\ \delta = c\rho^{-1/3}+\frac{d\rho^{-1/3}}{\left(1+d\rho^{-\frac{1}{3}}\right)}$$

Equation 59. Correlation energy according LYP method.
a, b, c, d: fitting parameters

The equation for LIP correlation energy does not take into account the parallel spin correlation when all spins are aligned. Therefore, it is usually combined with an exchange functional like B originating BLYP.

Meta-GGA Methods

Naturally, the GGA methods evolve for functionals including higher-order derivatives of the electron density. The inclusion of the Laplacian ($\nabla^2\rho$), the second-order derivative of the electron density, gave rise to the so called meta-GGA functionals. Alternatively, the

Laplacian can be replaced by the orbital kinetic energy density, which also allows an improvement of the GGA functionals [144, 167-169].

However, few pure meta-GGA functionals were developed since a more successful approach was developed – the hybrid GGA methods.

Hybrid GGA Methods

In the so-called hybrid GGA functionals, the principle is to include an HF exchange contribution to the total DFT exchange energy. Considering the LSDA method and a *Half-and-Half* (H+H) contribution, the exchange-correlation energy would be given by

$$E_{XC}^{H+H} = 0.5\, E_X^{HF} + 0.5\left(E_X^{LSDA} + E_C^{LSDA}\right)$$

Equation 60. Exchange-correlation energy of a H+H method.

Since the GGA functionals are considerable better than the LSDA, this approach was also applied to GGA functionals. The B3LYP [245-247] functional is one of the most popular DFT functionals and obeys this approach according to the following general equation:

$$E_{XC}^{B3LYP} = \left[(1-a)E_X^{LSDA} + b\Delta\mathrm{E}_\mathrm{X}^{B} + \; aE_X^{HF}\right] + \left[(1-c)E_C^{LSDA} + cE_C^{LIP}\right]$$

Equation 61. Exchange-correlation energy of the B3LYP functional.
a, b, c: empirical parameters obtained by fitting experimental data

In the case of B3LYP, the empirical parameters a, b, and c are 0.20, 0.72, and 0.81, respectively.

The idea of the hybrid GGA functionals is creating several different combinations of exchange and correlation GGA functionals with different adjusted empirical parameters.

In sum, the inclusion of HF exchange considerably helps to provide more accurate results using DFT. However, the amount of HF included in the functional deeply affects the accuracy for different chemical systems and for calculating different properties. Thus, a careful choice of DFT functionals is required to ensure that the functional truthfully describes the target chemical system.

Finally, DFT appears as a powerful tool to study several interesting chemical processes with some hundreds of atoms. Although it is required a careful choice of functional, the computational effort associated with DFT calculations is considerably lower than wave function-based methods with the same accuracy.

In the study of reactional mechanisms, it is common to employ DFT to optimize the geometries and then wave functional-based methods, like CCSD(T) or DLPNO-CCSD(T), to refine the energies of the obtained structures.

2.5. Nuclear Hamiltonian

All the previous discussion about QM methods is devoted to approaches to calculate the electronic energy. Due to the application of the Born-Oppenheimer, the total energy of the system can be divided into the terms involving electrons and terms that do not involve electrons (internuclear repulsion and nuclei kinetics) (**Equation 26**). This is only true if the instantaneous adjustment of electrons to the nuclei movement is considered. As already mentioned, this approximation is reasonably accepted since the mass of electrons is considerably smaller than the nuclei, and the chemical processes are mainly driven by electronic phenomena.

However, the nuclear Hamiltonian (**Equation 26**) needs to be calculated and sum to the electronic Hamiltonian to obtain the total energy of a certain system, according to the Born-Oppenheimer approximation.

Usually, further calculations are required to access the nuclear Hamiltonian contribution. Even at 0 K, the molecule vibrates, so the internal energy of the system (E_0) is calculated as the sum of the electronic energy ($E_{electronic}$) and the zero-point energy correction (ZPE) [248]:

$$E_0 = E_{electronic} + ZPE$$

Equation 62. ZPE correction.

In addition, thermal corrections are also computed to the desired temperature to allow the comparison between the computed energies and the experimental data. The thermal corrections are taking into account through the consideration of three terms associated with the translation, rotation, vibration of the nuclei. Thus, for a certain temperature T, the energy (E_T) can be calculated as [248]:

$$E_T = E_0 + [E_{translational} + E_{rotational} + E_{vibrational}]_T$$

Equation 63. Energy at temperature T.

The three terms are commonly calculated using approximated models. In the case of the rotational component, the rigid rotor model is used, the particle in a box model is used to calculate the translational term, and harmonic oscillators are employed to assess the vibrational component. When the vibrational calculation is computed for a TS, solely one of the frequencies should be imaginary, corresponding to the vibration of the atoms involved in that particular reaction. Consequently, the frequency calculation arises as an important test to verify the nature of stationary points [168, 248]. In the study of a catalytic mechanism, each step of the mechanism is defined by a reactant, a TS, and a product.

After the optimization, the frequency calculation is essential to ensure that both reactant and product correspond to a minimum (no imaginary frequencies) and the TS to a saddle-point (one imaginary frequency).

In a potential energy surface (PES), a certain reactant and product are linked by a path that minimizes the conversion of one into another. This path has a maximum that corresponds to the TS. The structure of the TS is a maximum in that path, but it is the minimum-energy structure in all other orthogonal coordinates. In mathematics, this spot is called a saddle point (Figure 13).

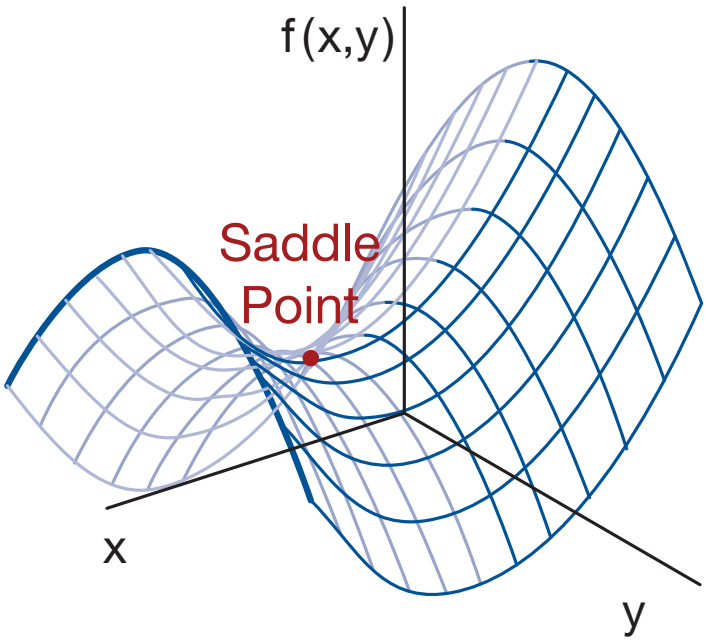


Figure 13. Saddle point at the surface of the function: $f(x,y) = \frac{1}{2}k(x^2 - y^2)$. (Adapted from [168])

Vibrational frequencies (v) are calculated considering the harmonic oscillator:

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

Equation 64. Frequency of a harmonic oscillator.
m: mass of the particle; k: force constant.

Therefore, the calculation of the frequency depends only on the determination of the force constants (k) associated with the displacement of the atoms. The harmonic force constants are calculated based on the second derivative of the energy with respect to the atomic coordinates. In a saddle point (TS), the TS structure can evolve into two opposite directions to decrease its energy (direction of reactants or products). Moreover, the movement across the orthogonal directions leads to an increase of the energy in

both directions. In this scenario, the second derivative (force constant) is negative, resulting in an imaginary frequency according to Equation 64.

As described, imaginary frequencies are specific to TS structures. Therefore, a truly optimized minimum should have no imaginary frequencies, whereas the TS must have only one. In the case of a chemical reaction, it must account for the atoms directly involved in the reaction.

During the thermal correction calculations, the enthalpy and the Gibbs free energy are also calculated. The enthalpy is calculated as the expansive work within an ideal gas model at a given temperature. Therefore, the enthalpy (H) is given by the following equation [249, 250]:

$$H = E_T + k_B T$$

Equation 65. Enthalpy contribution estimated by thermal corrections.
k_B: Boltzmann's constant; T: temperature

At a given temperature (T), the Gibbs free energy can be calculated according to the following equation:

$$G = H - TS$$

Equation 66. Gibbs free energy.

After the calculation of the enthalpy (H) as shown before, only the entropy (S) is missing. The entropy is calculated as a sum of all entropic contributions from translational, rotational, and vibrational components of the nuclei and also the electronic entropy:

$$S = S_{translational} + S_{rotational} + S_{vibrational} + S_{electronic}$$

Equation 67. Entropy contribution estimated by thermal corrections.

In this case, it is assumed that the electronic entropy ($S_{electronic}$) of a non-degenerated system is zero, and only the translational, rotational, and vibrational movements of nuclei affect the entropy of the system [248, 249].

In the study of an enzymatic mechanism, the nuclear Hamiltonian calculation is very important to refine the final energies and obtain the activation and reaction Gibbs free energies of each step. An accurate calculation of these quantities is essential to obtain a good estimation of the enzyme's catalytic activity that can be compared with the experimental data.

3. Hybrid Methods

Although DFT states an enormous advance in terms of the application of QM calculation in several systems, it is still limited to a few hundred atoms. While it is enough for a reasonable number of different chemical systems, it cannot be applied in systems with thousands or millions of atoms, like enzymes. Consequently, further approximations need to be taken into consideration when computational methods are applied in the study of enzymatic mechanisms.

The application of hybrid methods is one of the most common and successful approaches to overcome this drawback. The concept behind the hybrid methods is to split the molecular system into parts and treat those parts using different methods. The QM/MM is the most popular scheme of a hybrid method, where the site where the reaction occurs is accurately treated with a QM method, e.g., DFT, and the remaining part is treated using MM. However, this principle can be applied to other schemes such as QM/QM or QM/QM/MM. The development of this method worth the Nobel Prize in Chemistry for Warshel, Levitt, and Karplus, in 2013 [251-253].

This was an important milestone for computational chemistry, allowing the application of computational methods in the study of big and complex molecular systems like enzymes.

3.1. Quantum Mechanics/Molecular Mechanics

As mentioned, QM/MM is the most popular scheme among the hybrid methods, and it divides the system into two parts. One part (QM) is composed of a few hundred atoms and is usually treated using wave function or DFT methods. This set of atoms composes what is commonly called as the model system or high level. The remaining surrounding atoms are included in the calculation but using a lower theoretical level, such as the MM that can rapidly handle thousands of atoms.

Although the implementation of such methods seems to be simple, there are some considerations that need to be taken into account, in particular, the contributions of each Hamiltonian and the boundary region.

The simplest approach is to compute the total Hamiltonian of the system as a sum of the Hamiltonians of each layer and the boundary regions between each pair of layers. This approach is called an additive method and would assume the following form for the energy calculation of a QM/MM system ($E_{QM/MM}$):

$$E_{QM/MM} = E_{QM}^{Model} + E_{MM}^{Surround.} + E^{Interaction}$$

Equation 68. Energy function for a hybrid additive method.
E_{QM}^{Model}: energy of the model system calculated using QM; $E_{MM}^{Surround.}$: energy of the surrounding system calculated using MM; $E^{Interaction}$: energy of interaction between the two layers

However, the additive method raised a challenge regarding the calculation of the contribution of the boundary region where the two layers interact. Therefore, in 1995, Maseras and Morokuma proposed a so-called subtractive approach called ONIOM (our own N-layered integrated molecular orbital molecular mechanics) method [254].

In a QM/MM ONIOM calculation (**Figure 14**), the energy is calculated for the entire molecular system using the MM method, then the model system is calculated using the MM method as well, and finally, the model system contribution is added after being considered using a QM method.

$$E_{QM/MM} = E_{MM}^{All} - E_{MM}^{Model} + E_{QM}^{Model}$$

Equation 69. Energy function for a hybrid substrative method (ONIOM).

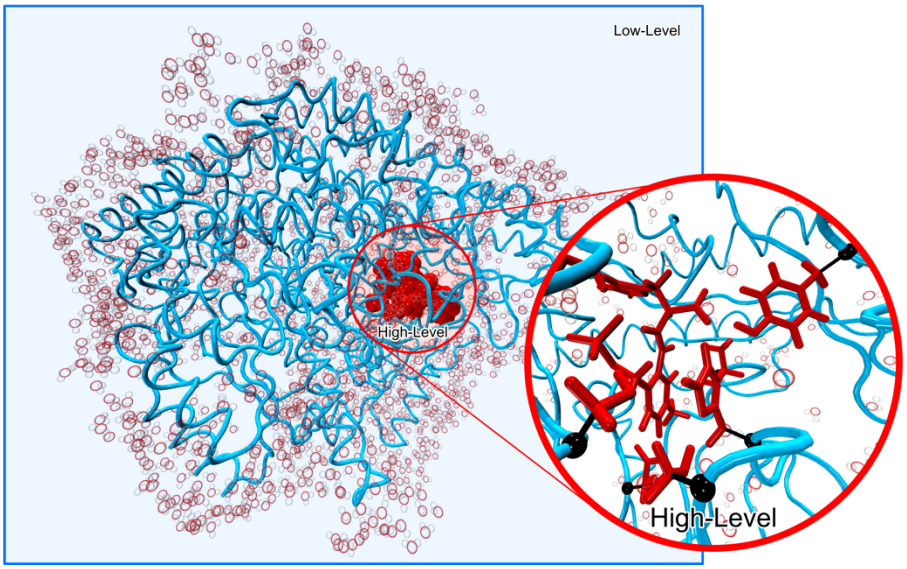


Figure 14. Representation of an ONIOM QM/MM Model. The red atoms are included in the model system/high level (treated with a QM method) and the blue representation corresponds to the surrounding system/low level (treated with a MM method). The black atoms correspond to the link atoms that span across the boundary between both QM/MM layers.

Although the aforementioned examples account for only two layers, which is the most common scenario, it can be applied for any number of layers treated with different theoretical methods.

Despite the fact that the ONIOM method makes the application simpler, it comes with a disadvantage: the interaction between the two layers is considered using MM. However, further improvements have been developed to improve the way this interaction is considered.

3.2. Mechanical, Electrostatic, and Polarizable Embedding

Regarding the interaction between two layers in a QM/MM calculation, three different approaches were developed to account for the contribution of the atomic charges and geometries between layers. In the ideal case, all electrons would be included in the QM calculations, but it would lead to the same problem since the size of the model is the main limitation of the QM methods. Therefore, the mechanical, electrostatic, and polarizable embedding schemes have been developed to mediate the interaction between layers.

3.2.1. Mechanical Embedding

Mechanical embedding is the simplest approach, and it is also the fastest one in terms of computational calculation. In this method, the QM region does not “see” the MM atoms during the calculations. So, it is equivalent to a calculation of the model system in vacuum. The surrounding atoms only provide geometric constraints that limit the possible movement of the atoms within the model system.

According to the mechanical embedding approach, in each optimization step, the surrounding system is optimized using MM, and then the model system is optimized using a QM method in the gas phase. This means that each layer is optimized under the effect of forces of the adjacent layer, but no electronic interaction is taken into consideration.

Due to the simplicity of this approach, it usually can be successfully applied if the model system is big enough to include all the important interactions.

3.2.2. Electrostatic Embedding

The electrostatic embedding scheme is an improved version of the previous method, allowing the polarization of the model system by the atoms of the surrounding layer. Partial atomic charges of the atoms in the surrounding layer (treated with MM) are included in the QM Hamiltonian as point charges. The inclusion of such charges in the QM Hamiltonian makes the calculation more time-demanding when compared with the mechanical embedding. The main problem of electrostatic embedding, in addition to the bigger computational effort that it requires, is some over polarization effect of the model system. Since the charges of the surrounding system are kept immutable during the simulation, they can generate artifacts when included in QM Hamiltonian.

3.2.3. Polarizable Embedding

The third method is called polarizable embedding, and it is an improved version of the electrostatic embedding approach. In a polarizable embedding calculation, the charges computed for an atom belonging to the model system (calculated with QM) are included in the MM calculations. However, this approach requires the use of polarizable MM force fields and becomes more exigent in terms of computational effort. An additional iterative cycle must be considered to allow the convergence of the model and surrounding regions that depend now on each other. This effort is even more demanding because the polarizable embedding tends to include partial charges, dipoles, and quadrupoles in the polarization of each atom.

The considerable increase in the computational effort associated with the use of polarizable embedding and the lack of support from commonly used MM force fields, e.g., AMBER, makes this approach less common than the previous two.

3.3. Boundary Conditions

So far, it has been discussed different methods to couple the calculations of different layers in a QM/MM system. However, those approaches only consider the non-bonded interactions, which does not raise any issue if no covalent bond spans through a pair of adjacent layers. This is true for systems where small molecules react in solvent, and no covalent bonds need to be cut between layers. Nonetheless, in the study of enzymatic mechanism, the enzyme is usually a single or multiple polypeptide chains that are too big to be considered as a whole in the model system. Thus, some covalent bonds are split when the layers are defined. If a chemical bond is cut, the cleavage can be homolytic or heterolytic, and in both cases, the atomic valence changes. Since the definition of layers is an approximation to allow the calculation using different methods, this cannot be considered in the calculation as a chemical phenomenon. Therefore, different methods were developed to address this question, in particular, the frozen localized orbitals and the link atom methods [144].

The frozen localized orbitals method allows keeping the orbitals of the atoms in the boundary regions as they are when the atoms are bonded to each other. This means that the orbitals are kept fixed during the calculations. Despite the good results obtained with this approach, its application is not straightforward in the sense that it requires a parameterization of the orbitals of each atom involved in bonds that span over the boundary of two layers [144].

The link atom method is extensively used to overcome this issue, in part because it is very simple to implement. The link atom is an atom that is placed in the direction of the truncated covalent bond and allows completion of the atomic valence of the involved atoms. Usually, the hydrogen atom is used because it is monovalent, small, and easy to compute. In some specific cases, it could be useful to use different atoms to mimic some polarization effect that has been compromised after truncation of layers. However, this is not recommended, and the inclusion of more atoms near that region should be considered. The limit between layers must be carefully chosen to avoid split non-single covalent bonds, highly polarized bonds, or bonds involved in chemical resonance. Moreover, it is also important to avoid cutting bonds close to the atoms actively involved in the chemical reactions of interest.

4. Enzymatic Kinetics

In the previous topics of this chapter, different methods have been presented and explained with respect to their applicability in the study of chemical reactions, in particular enzymatic mechanisms. One of the mentioned aspects was that usually DFT is used to obtain accurate geometries for the structures of reactant, transition states, reaction intermediates, and product of a chemical reaction. Then, more accurate methods, like DLPNO-CCSD(T), are used to refine the activation and reaction Gibbs free energies. The question that can be made is: why is the accuracy of the energy calculation so important?

The activation Gibbs free energies ($\Delta G^{\ddagger}$) are important checkpoints in the study of enzymatic mechanisms since they establish an important bridge between the computational results and the experimental kinetic data.

According to enzymatic kinetics, a mechanism involving two or more steps is limited by the step that has slower kinetics compared to all other ones. This means that the specific rate of that step is much lower when compared to the other steps. Consequently, the slowest step has the highest $\Delta G^{\ddagger}$ of the entire mechanism, and it would correspond to the rate-limiting step.

In the study of an enzymatic mechanism, using computational means, the final energy profile will include the $\Delta G^{\ddagger}$ of each step, and the highest barrier (or combination of sequential steps) will indicate the rate-limiting step. The proposal of mechanism obtained with the computational calculations can be validated, in some extension, comparing the calculated barrier for the rate-limiting step with the ones obtained experimentally.

Unfortunately, experimental scientists usually do not collect $\Delta G^{\ddagger}$. Instead, turnover numbers (k_{cat}), Michaelis constants (K_M), and/or reaction rates (v) are obtained. The K_M is expressed as molar concentration and provides insight about the affinity of the enzyme, which is not addressed in typical computational studies of enzymatic mechanisms. However, the k_{cat} is a measure of how fast an enzyme is. The k_{cat} is expressed as the number of catalytic cycles per unit of time, and it is mainly defined by the rate-limiting step of the mechanism.

Consequently, if a relationship between the k_{cat} and the $\Delta G^{\ddagger}$ could be established, it would be possible to compare the computed energies with the experimental kinetic data.

This relation was established in 1889 by Svante Arrhenius [255] and it is called the Arrhenius equation:

$$k_{cat} = Ae^{-\frac{E_a}{RT}}$$

Equation 70. Arrhenius equation.

A: pre-exponential factor; E_a: activation energy; R: universal gas constant; T: temperature

Equation 70 establishes a relation between the turnover number and the activation energy of a certain chemical reaction. In 1935, Eyring refined the Arrhenius equation to give statistical mechanical support to the empirical equation of Arrhenius:

$$k_{cat} = \frac{\kappa k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}}$$

Equation 71. Eyring equation.
κ: transmission coefficient; k_B: Boltzmann's constant; h: Plank's constant

The Eyring equation [256] allows establishing a relation between the turnover number of a chemical reaction and the $\Delta G^{\ddagger}$ of the rate-limiting step in a temperature-depending manner. The κ parameter (transmission coefficient) is normally equal to one, assuming that the reaction evolves in the direction of the product without recrossing.

In sum, an enzymatic turnover number, determined experimentally, can be converted in the maximum activation Gibbs free energy for the rate-limiting step of the mechanism. Comparing that value with the $\Delta G^{\ddagger}$ obtained with the computational simulations, it is possible to validate or refute a certain mechanistic proposal.

Moreover, the experimental kinetic data with mutant forms of enzymes offer excellent clues about the importance of certain residues for the catalytic or inhibitory processes. For example, a mutation of a specific residue can cause a significant decrease in the k_{cat} or completely abolish the enzymatic activity. This information elucidates that a specific amino acid residue is essential for the course of the reaction.

The other way around can also be used to produce biocatalysts through the enhancement of enzymes by direct mutagenesis. Different mutant enzymes can be studied computationally to assess their potential energy surfaces (PES) and look for the one that is more kinetically favored. Then, most promisor mutant enzymes are produced, in the wet lab, to test their real efficiency. Eventually, the best one could be scaled-up to an industrial application.

Notwithstanding, this is only one possible checkpoint to compare the computational data with experimental evidence. Other approaches, which might be synergically included (if

available), include a comparison of intermediate structures, electron paramagnetic resonance (EPR) spectroscopy, UV-Vis spectroscopy, among others.

C. Results and Discussion

Try to learn something about everything and everything about something.

Thomas Huxley

This page was intentionally left blank.

1. PLP-dependent Enzymes

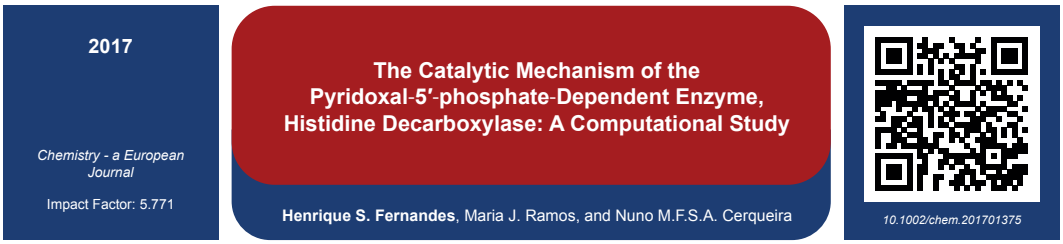
This first section of chapter C is focused on the four catalytic mechanisms of PLP-dependent enzymes studied during the course of this Ph.D. project. Afterward, a final conclusion compares the chemical features of the different studied enzymes, providing insights about how these enzymes shape the chemistry of PLP to catalyze different reactions and different substrates. Moreover, it also includes the main conclusions retrieved from the study of these PLP-dependent enzymes.

This page was intentionally left blank.

1.1. Decarboxylation

Histidine decarboxylase (HDC) was the enzyme selected to study the catalytic mechanism of a PLP-dependent decarboxylase. This enzyme was chosen because of its biological relevance in the synthesis of histamine and also taking into account the pathological situations related to the abnormal production of this amine. Moreover, the catalytic mechanism had not been completely described before, and its structure had been published recently.

All the calculations were performed by Henrique S. Fernandes, as well as the writing of the article resulting from this work, which was revised considering the contributions of all the authors. The work was published in the Chemistry – a European Journal, and the following content is almost a transcription of the published version. This manuscript was highlighted by the journal as a “Hot Paper”.



The catalytic mechanism of HDC, a PLP-dependent enzyme, was studied by using a computational QM/MM approach following the scheme M06-2X/6-311++G(3df,2pd):AMBER. The reaction involves two sequential steps: the decarboxylation of L-histidine and the protonation of the generated intermediate from which results histamine. The rate-limiting step is the first one ($\Delta G^{\ddagger}$ = 17.6 kcal/mol; ΔG_R = 13.7 kcal/mol) and agrees closely with the available experimental k_{cat} (1.73 s^{-1}), which corresponds to an activation barrier of 17.9 kcal/mol. In contrast, the second step is very fast ($\Delta G^{\ddagger}$ = 1.9 kcal/mol) and exergonic (ΔG_R = -33.2 kcal/mol). Our results agree with the available experimental data and allow us to explain the role played by several active site residues that are considered relevant according to site-directed mutagenesis studies, namely Tyr^{334B}, Asp^{273A}, Lys^{305A}, and Ser^{354B}. These results can provide insights regarding the catalytic mechanism of other enzymes belonging to family II of PLP-dependent decarboxylases.

This work was also presented in the following conferences:

- (Poster) *Encontro De Jovens Investigadores Em Biologia Computacional Estrutural*, **2015**, Coimbra, Portugal (Appendix: page 379)
- (Oral Communication) *IJUP'16*, **2016**, Porto, Portugal
- (Poster) *1st European Young Chemists Meeting* and *5º Encontro Português de Jovens Químicos*, **2016**, Guimarães, Portugal (Appendix: page 380)
- (Oral Communication) *11th European Scientist Meeting*, **2016**, Porto, Portugal

1.1.1. Introduction

Histamine is an endogenic biological amine that plays a role in several physiological events in the human body. Histamine has been reported to be involved in allergic reactions, central nervous system, stomach, cardiovascular system, and smooth muscle. In particular, histamine has been shown to act on cell cycle arrest [257], muscle endurance [258], hair growth [259], and histamine-containing granules maturation [260].

Due to the panoply of biological functions assigned to histamine, this amine is also associated with some illnesses, such as atopic dermatitis [100], chronic allergic contact dermatitis [101], allergic rhinitis [102], gastric ulcer [103, 104], diabetes [105], chronic heart failure [106], and several types of cancer [107-111].

In eukaryotes, histamine is produced from the decarboxylation of L-histidine into carbon dioxide and histamine by histidine decarboxylase (HDC, EC 4.1.1.22) [261-263]. HDC is an active homodimer that contains two active sites at the interface of both monomers, each one holding residues from each subunit [264, 265]. Each active site shelters one molecule of the PLP cofactor, that is essential to receive and covalently bind the substrate [12]. The PLP cofactor binds to the active site during the dimerization process where it becomes covalently bonded to an active site lysine (Lys^{305}) through an imine linkage, forming the internal aldimine. At this point, PLP cofactor is correctly positioned to receive and bind a histidine molecule (substrate) [266].

When the substrate becomes available in the active site of HDC, it binds to the PLP cofactor, forming the external aldimine. In this process, the imine bond formed between PLP and Lys^{305} is cleaved and a new one is formed between the PLP cofactor and the amino group of the substrate [266].

Since the formation of the internal and external aldimines have been shown to be common to all PLP-dependent enzymes, and their mechanism have been already studied, this work was devoted to the specific steps involved in the decarboxylation process catalyzed by the HDC.

The wild-type form of HDC has a turnover rate (k_{cat}) of 1.73 s^{-1} and an enzymatic efficiency (k_{cat}/K_M) of 17.3 s^{-1} mM^{-1} [261]. A set of mutagenesis studies have proposed some amino acid residues as crucial for the catalytic process, since their mutation have a deep impact in HDC's turnover rate.

The mutation of Lys^{305} by Ala [267], Gly [268] or Arg [269] completely inactivates the enzyme, corroborating the great importance of Lys^{305} for the acceptance and positioning of the PLP cofactor.

In addition, the mutation of Asp^{273} by Gly likewise culminates in a total inactivation of the enzyme. The Asp^{273} makes an important hydrogen bond with the pyridine ring of PLP and helps the stabilization of the cofactor inside the active site.

When, His^{194} or Ser^{304} were mutated by Gly, the enzymatic efficiency of the HDC reduces drastically and almost inactivates the enzyme. His^{194} seems to have an important role in the orientation and alignment of the PLP cofactor inside the active site, through a π-stacking interaction. On the other hand, Ser^{304} proximity to the Lys^{305} suggests a possible role in the orientation of Lys^{305} for the formation of the internal aldimine [268].

Finally, Tyr^{84}, Tyr^{80}, and Ala^{86} seem to be important for the substrate's stabilization, since their replacement by Gly have a negative impact in the enzymatic activity [268].

Fleming *et al.* have further reported that the mutation of some amino acid residues in the loop located in the top region of the active site interferes with the HDC activity. In particular, mutation of Tyr^{334} by Gly culminates in the complete loss of the enzymatic activity, whereas the mutation of the neighbor residues decreases its catalytic activity.

In 2008, Moya-Garcia *et al.* studied, by computational means, the first step of the catalytic mechanism of HDC using a homology model of HDC that was built based on two incomplete crystallographic structures of aromatic-L-amino-acid decarboxylase (AADC) (PDB codes: 1JS3 and 1JS6) [270, 271]. In this study, the authors proposed that this step should be the rate-limiting one since it was characterized by a high activation energy of 20.4 kcal/mol. However, this study was not completed, since neither the product of the reaction of the first step nor the second step of the reaction were described. In addition, in 2012, the X-ray structure of HDC was determined for the first time (PDB code 4E1O) [261]. This new structure of the enzyme revealed that the homology model presents several pitfalls in the region of the active site. For example, in the X-ray structure, the carboxylate group of the substrate is pointing towards the solvent, whereas in the homology model it is pointing in the opposite direction. The homology model also lacks a network of hydrogen bonds that surround the PLP cofactor and was found to be important to stabilize it during the catalytic process. The computed energies for the decarboxylation process are also higher than what should be expected to take into account the reaction rate that was determined experimentally.

When put together, all of these facts suggest that the mechanism of HDC may be different from the one previously proposed by Moya-Garcia *et al.*. Thus, in this work, the catalytic mechanism of HDC was studied using a ONIOM QM/MM computational approach. This studied showed, for the first time, the complete mechanism for the

decarboxylation of a PLP-dependent enzyme. In particular, the mechanism for the second step, where the quinonoid intermediate is protonated [189].

1.1.2. Methodology

a. Model Preparation

At the time of this study, only one crystallographic structure of HDC was available, and it is deposited under the PDB code 4E1O [261]. This structure has three equivalent dimeric structures: AB, CD, and EF. Among them, dimer AB was selected to pursue the study because it presents a less distorted structure of the external aldimine and it does not have any artificially introduced tag of amino acid residues.

As mentioned before, the HDC consist of a homodimer, so from the two active sites, only one was selected to be include in the high-level (HL) layer of the QM/MM calculations.

The structure was obtained through the co-crystallization of HDC with L-histidine methyl ester inhibitor (HME). Therefore, the methyl group bonded to the carboxylate group of the HME was removed to get the natural substrate (L-histidine). Additionally, Ser^{180} and Ser^{418}, from each subunit, were replaced by Cys since they have been mutated to help the crystallization process. These two residues are far from the active site and do not affect the catalytic efficiency [261, 272].

The same approach was followed to build the quinonoid intermediate, completely deleting the carboxylate group from the external aldimine available in the original PDB.

b. MD Simulations

After the system's preparation, the resulting biomolecular complex were solvated using a TIP3P type of water molecules [273, 274], distancing at least 12.0 Å to the face of the box.

The protonation states of the amino acids, at pH 7.0, were assessed using PROPKA 3.1 [275, 276], and the subsequent protonation and preparation were made using the SANDER software from AMBER 12 package [277].

Antechamber software from AMBER12 package were used to assign parameters to the external aldimine and quinonoid intermediate. The atomic charges for these molecules were obtained from RESP of the atomic charges calculated using HF/6-31G(d).

A set for four minimizations, an equilibration, and a 60 ns production were followed. The MD simulation was conducted in an isothermal-isobaric ensemble, using the ff99SB [149, 278, 279] force field, at the temperature of 310.15 K, pressure of 1.0 bar, integration step of 1 fs, and a 10.0 Å cutoff for non-bonded interactions.

c. QM/MM Calculations

From the MD simulations, five clusters were made for each simulation, in order to select the most promisors to conduct the QM/MM calculations.

The QM/MM model was built to include the full dimer, the external aldimines in each active site, and all the water molecules located at a maximum of 2 Å from the protein. All atoms positioned more than 30 Å away from the active site were fixed during the geometry optimizations.

Given the system's size (more than 17,000 atoms) and the fact that the geometry optimizations are very time-consuming, the ONIOM method was employed to perform the calculations (Figure 15). In the HL layer, it was included the external aldimine or the quinonoid intermediate and all the relevant amino acid residues for the reaction. The composition of the HL layer changes according to the studied model, so they are going to be described in the results section. The atoms belonging to the low-level (LL) layer were considered using AMBER force field, and hydrogen atoms were used as link-atoms for the cleaved bonds that span over the two layers. The mechanical embedding scheme was applied due to the size of the system and the limited computational resources.

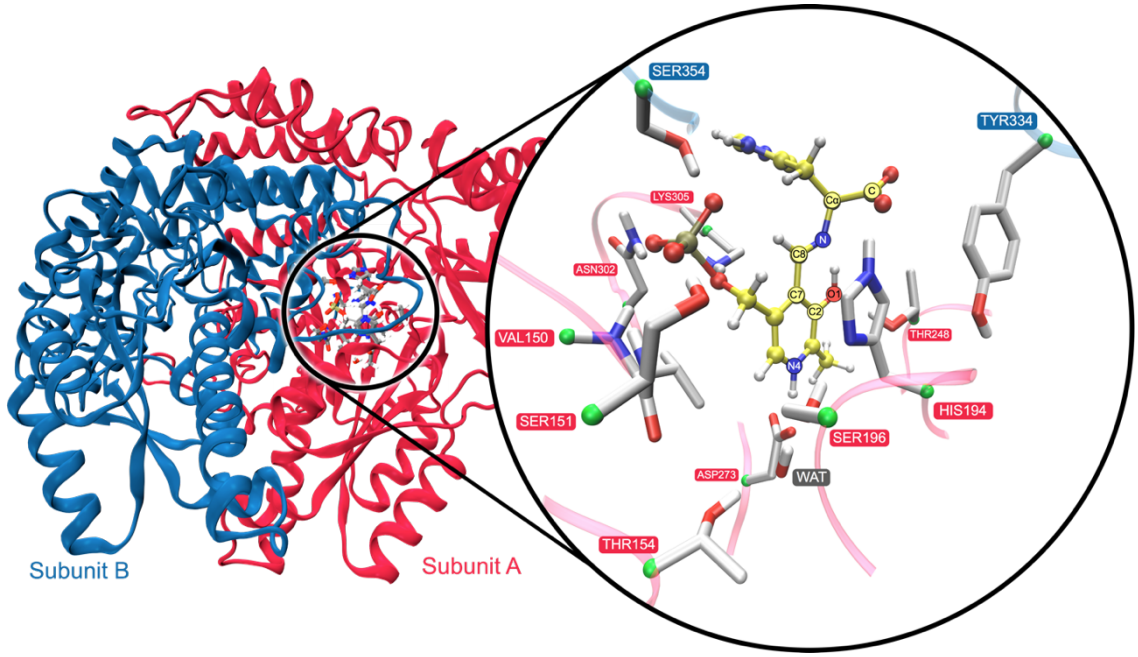


Figure 15. (Left) New cartoon representation of the X-ray structure of HDC from *Homo sapiens* (PDB code 4E1O), with the external aldimine. Subunits A and B are colored in red and blue, respectively. (Right) Representation of the atoms that were placed in the HL layer during the geometry optimizations. The carbon atoms from amino acid residues and the water molecule are colored in gray whereas the ones from the external aldimine are colored in yellow. The representation was simplified by hiding the hydrogen atoms bonded to nonpolar atoms of the amino acid residue. The green spheres represent the hydrogen atoms that were used as link atoms.

All the calculations were performed using Gaussian09 [280] using the following scheme for geometry optimizations B3LYP/6-31G(d):AMBER [182, 245-247, 281-285], and M06-2X(GD3)/6-311++G(3df,2pd):AMBER [281, 284, 286-290] for the single-point energy calculations.

All the minima and TS were confirmed by vibrational frequency calculations, and the thermal correction were computed at 310.15 K and 1.0 bar.

The atomic charges analyses of the HL layer were calculated considering Merz-Singh-Kollman type of charges [291].

Visual Molecular Dynamics (VMD) [292] and GaussView 5 [293] were used as visualization tools.

1.1.3. Results and Discussion

The catalytic mechanism of HDC was computationally studied using two QM/MM models based on the available crystallographic structure. In first model, the HME inhibitor was modified to represent the histidine and study the initial step. The second model was built removing the carboxylate group of the substrate to get a structure of the quinonoid intermediate.

a. MD Simulations

Two independent MD simulations were conducted for both models, each one with a duration of 60 ns.

Both simulations presented an equilibrated region for the last 20 ns of simulation, with a RMSd value of 1.86±0.06 Å and 2.11±0.11 Å for the external aldimine and quinonoid intermediate simulations, respectively (Figure 16).

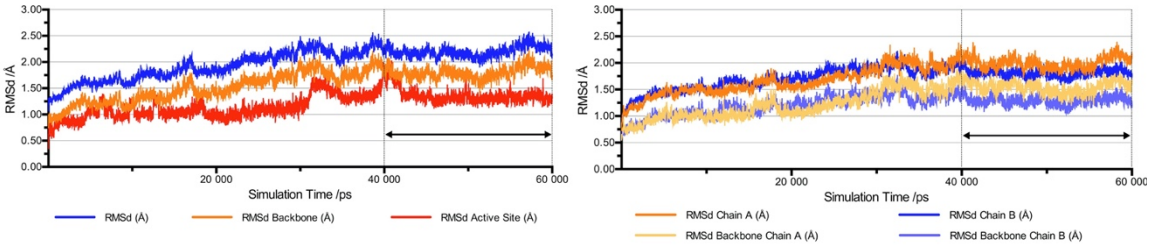


Figure 16. RMSd values for the MD simulation of HDC with (Left) the external aldimine and (Right) the quinonoid intermediate. The blue, orange, and red lines correspond to the RMSd values of all atoms in the system; only the α-carbons; or the active site amino acid residues plus the external aldimine/quinonoid intermediate, respectively. The simulation was considered equilibrated during the last 20 ns.

A cluster analysis was then conducted in the equilibrated region of each MD simulation. The most representative structures of each cluster were visually analyzed and considered for the subsequent QM/MM calculations. Some interactions have been shown conserved among all the clusters, in particular the Ser^{151}, Val^{150}, Asn^{302}, Ser^{345}, and Lys^{305} that are responsible for stabilizing the phosphate group of the PLP cofactor, through a set of hydrogen bonds. Moreover, the pyridine ring of PLP establishes a strong hydrogen bond with the Asp^{273}, via its NH group. His^{194} residue is also responsible for this stabilizing by a π-stacking interaction.

A high flexibility was observed for the position of amino acid residues belonging to the loop that closes the active site. Among those residues, Tyr^{334} residue is the one presenting a higher mobility.

b. *The Catalytic Mechanism*

The catalytic mechanism of HDC was initially studied resorting to QM/MM methodologies using one structure that was retrieved from a representative conformation of the most populated cluster obtained from the MD simulation.

The chosen structure has the substrate (Figure 17 Center) in the same conformation of the HME inhibitor (Figure 17 Left) as observed in the crystallographic structure.

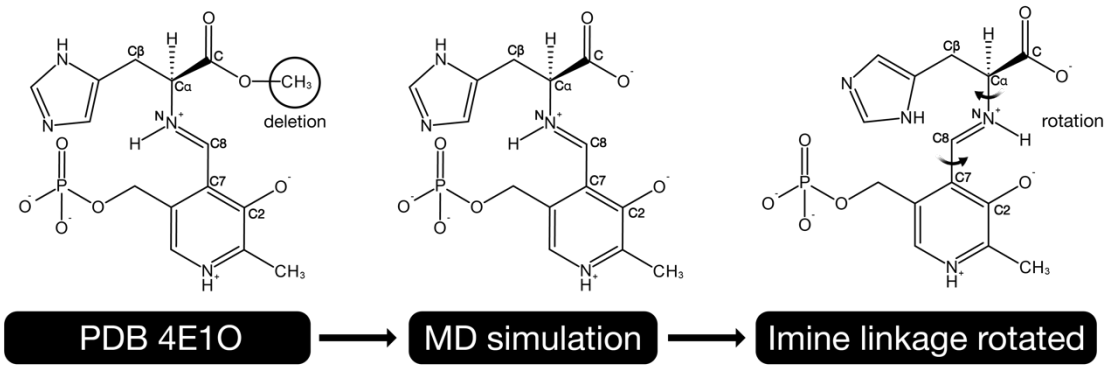


Figure 17. Structures of the external aldimine (Left) available in the PDB file 4E1O where the PLP cofactor is bonded to the HME inhibitor; (Center) where the PLP cofactor is bonded to the natural substrate; (Right) where the imine linkage was re-orientated to fit the common configuration of other PLP-dependent enzymes.

Such conformation is not common among the PLP-dependent enzymes. Usually, the NH group of the imine linkage is pointing towards the ketone group and not to the phosphate group as it is found in this X-ray structure. This suggests that such configuration could be a consequence of the inhibitory process. However, since the inhibitor is very similar to the natural substrate and only differs on a methyl group, this new configuration could be an intrinsic characteristic of the enzyme and important for the catalytic process. Taking this into account, we started our study investigating the decarboxylation process with the external aldimine adopting the uncommon configuration observed in the co-crystalized X-ray structure 4E1O (Figure 17 Center).

The QM/MM calculations revealed that the decarboxylation cannot occur, under physiological conditions, when the external aldimine adopts this uncommon conformation. The computed activation barrier and reaction energy are (28.9 kcal/mol and 28.5 kcal/mol, respectively), making the reaction too slow to be biological viable. Moreover, the calculated barrier is much higher than the one expected taking into account the experimental turnover number [261] (~17.9 kcal/mol)[10]. Therefore, more

[10] Estimated based on the Transition-State Theory

than 30 different conformation rearrangements of the surrounding amino acid residues were tested. Among all these models, four main models were tested, A (A, A', and A''), B, C, and D (Figure 18).

Regarding the model A, the displacement between the Tyr^{334} residue and the carboxylate group of the substrate was tested. This distance was increased over the A, A', and A'' models based on the evidence that the Tyr^{334} residue belongs to a highly flexible loop of the enzyme.

In model A, the hydroxyl group of Tyr^{334} interacts, by a hydrogen bond, with the carboxylate group of the external aldimine (1.61 Å), whereas in models A' and A'' this distance increases to 4.09 Å and 8.46 Å, respectively. (Figure 18)

The QM/MM structure of each of these models is identical. The HL layer includes the external aldimine, and the side chains of Ser^{354}, Tyr^{334}, Ser^{151}, Asp^{273}, Ser^{196}, His^{194}, and Thr^{248}. In none of the cases (models A, A', and A'') was possible to characterize the TS for the decarboxylation reaction and the energy profile was very unfavorable. The energy barriers have reached over 30 kcal/mol (Figure 18). However, these models were useful to conclude that the reaction is favored when Tyr^{334} is further away from the carboxylic group of the external aldimine. It can be rationalized that Tyr^{334} residue hinders the carbon dioxide dissociation.

The model B was built based on the previous model A''. Since the Tyr^{334} is considerable away from the external aldimine, it was excluded from the HL layer. In counterpart, the Thr^{154} residue and a nearby water molecule were included in the HL layer (Figure 18). This new conformation of the active site allows the formation of a hydrogen bonds network that includes the Asp^{273}, a water molecule, Ser^{196}, and His^{194}, improving the interaction of the active site residues with the external aldimine (Figure 18). The model B was able to provide a TS structure (101.02i cm^{-1}) and related minima. However, the computed activation and reaction energies were too high considering the expected values for a quite fast enzyme (E_a = 28.9 kcal/mol and E_R = 28.4 kcal/mol).

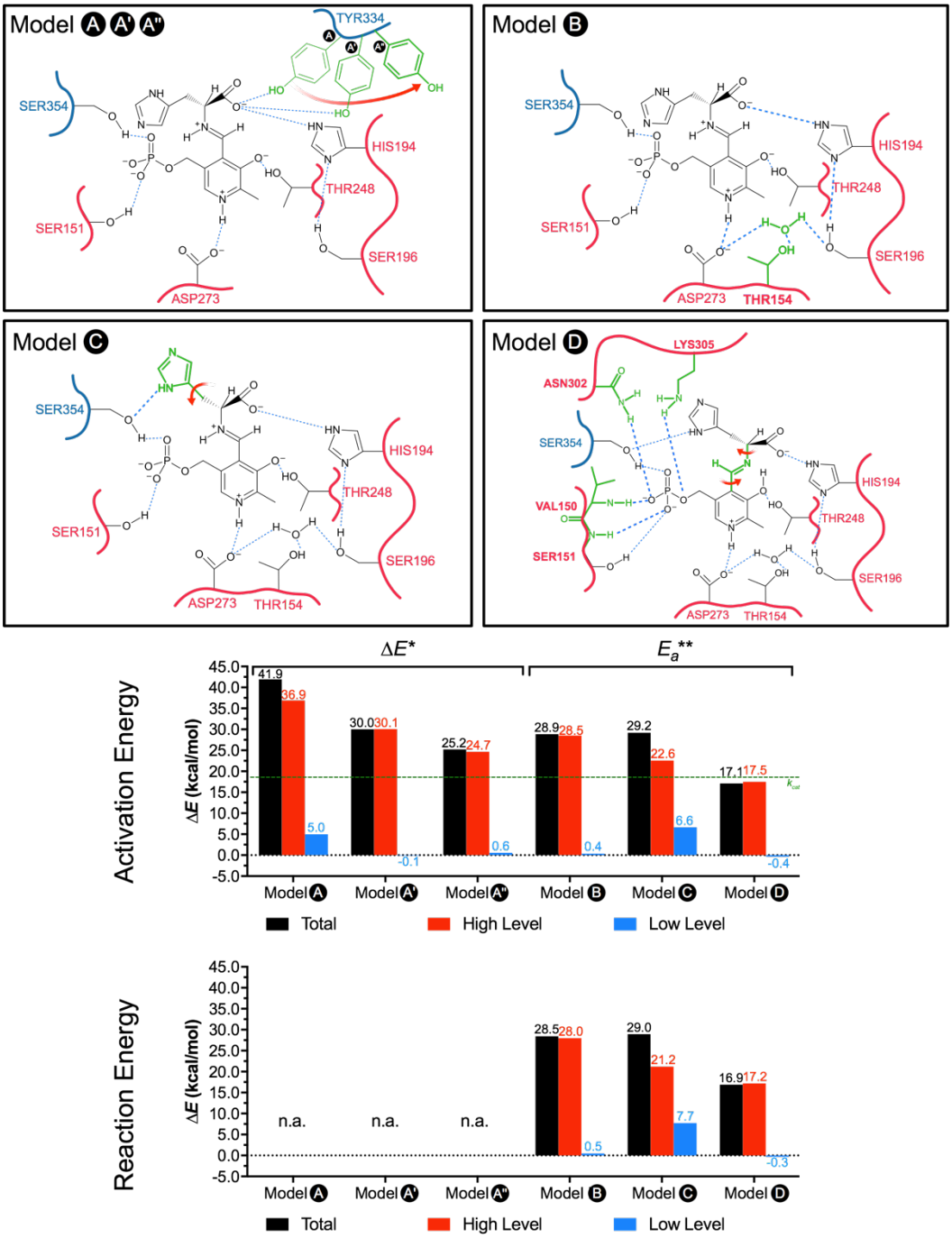


Figure 18. Main models that were built to study the first step of the catalytic mechanism of HDC. The structures colored by green correspond to those were modified comparing to the previous model. The graphs show the activation and reaction energies for all the main models presented above. Dashed lines colored by blue represent important non-bond interactions at the active site. * Energy calculated from the difference between the higher energy from the scan and the reactant. ** Energy calculated from the difference between the optimized transition state and the reactant. No ZPE or thermal corrections were included in these calculations.

Further improvements were made (model C), and the imidazole ring of the substrate was rotated by 180° in relation to the model B.

The model C favors the formation of an additional hydrogen bond between the external aldimine and Ser^{354} residue (Figure 18). This modification agrees with site-directed

mutagenic experiment on Ser^{354}, where the mutation of this residue by a Gly decreases the affinity of HDC to L-histidine [261]. Although it was possible to isolate the TS structure (90.64i cm^{-1}), the activation and reaction energies were, again, too high: E_a = 29.2 kcal/mol, E_R = 28.9 kcal/mol.

All of these models were unsuccessful to describe the decarboxylation process of HDC. So far, all the models share the same orientation of the NH group of the imine linkage, so, in model D, this group was rotated to fit what it is commonly observed in other PLP-dependent enzymes. The rotation was accomplished by rotating the two adjacent dihedral angles (C2-C7-C8-N and C8-N-C2-C) by 180° in opposite directions (**Figure 17** Right).

This new conformation of the external aldimine fits very well inside the active site and does not imply additional rearrangements of the neighbor active site residues. Moreover, it allows the keto-enol tautomerism between the NH group of the imine linkage and the ketone group of the PLP cofactor as it has been proposed by spectroscopic studies [266].

The same computational approach was employed using the model D as starting point, and, for the first time, the computed energies closely agree with the experimental predictions.

A relaxed coordinate scan was performed to incrementally increase the Cα-C bond of the substrate. Since model D favors the perpendicular positioning of the Cα-C bond in relation to the PLP and the PLP's conjugated system withdraws the electron density of this bond, the reaction occurs seamless with an activation Gibbs free energy of 17.6 kcal/mol. The calculated activation energy is very close to the one predicted based on the experimental turnover number [261] (17.9 kcal/mol). The reaction Gibbs free energy calculated for this first step was 13.7 kcal/mol (**Figure 19**).

In the reactant of step 1, the negatively charged Asp^{273} residue stabilizes the protonated and positively charged nitrogen of the pyridine ring of the PLP cofactor (1.52 Å). The imidazole group of the substrate is stabilized by an important hydrogen bond that is endorsed by Ser^{354} (1.73 Å). The phosphate group of the PLP cofactor has a charge of -2 and it is extensively stabilized by a network of hydrogen bonds endorsed by Ser^{151} (OH: 1.64 Å; NH: 2.64 Å), Val^{150} (NH: 2.47 Å), Asn^{302} (NH_2: 2.18 Å), Ser^{354} (OH: 1.51Å), and Lys^{305} (NH_2: 2.45 Å).

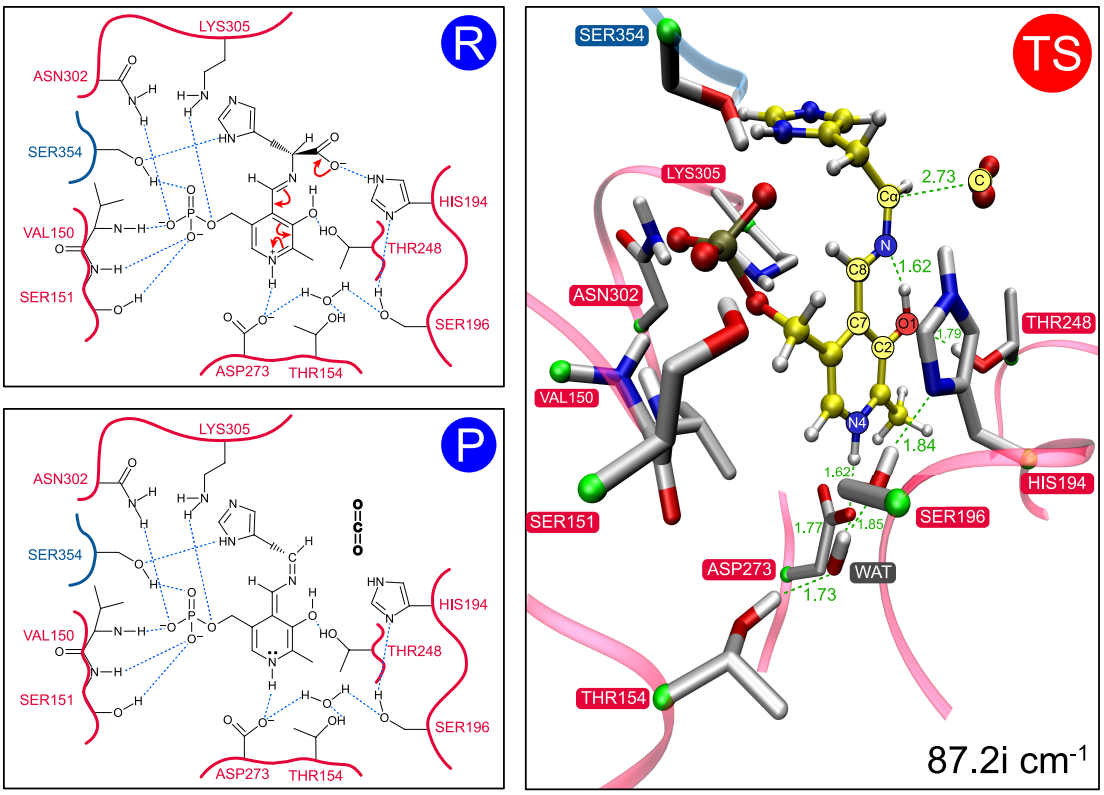


Figure 19. Structures of reactant, TS, and product of the first step of the reaction catalyzed by HDC. (Left) Wedge-Dash representation of the structures of reactant and product. (Right) *New cartoon* and licorice representations of the TS structure of the active site.

In the course of the decarboxylation process, the Cα-C bond becomes weaker, promoting the decarboxylation and the release of the carbon dioxide. The TS of this reaction is characterized by only one imaginary frequency at 87.2 cm^{-1}, in which the interatomic distance between the Cα and C atoms is 2.73 Å (Figure 19). At the end of the decarboxylation process, one molecule of carbon dioxide is formed, and it is released from the active site, and the quinonoid intermediate is formed.

After the decarboxylation process takes place, it is observed an electronic withdrawing effect that is mediated by the conjugated system of the PLP cofactor. The negative charge of the carboxylate group is accommodated on the PLP cofactor (-1.18 a.u. in the reactant versus -1.72 a.u. in the product) and in particular at the nitrogen N4 belonging to the pyridine ring (-0.35 a.u. in the reactants versus -0.53 a.u. in the product). This effect is confirmed by a decrease in the bond length of N4-H (1.10 Å in the reactants versus 1.06 Å in the product) and the consequent increase of the intermolecular interaction between Asp^{273} and the N4-H of the PLP cofactor (1.52 Å in the reactants versus 1.64 Å in the product). In order to compensate the decreasing stabilization of the negative charge of Asp^{173}, the interaction of this residue with the neighbor water molecule is strengthened (1.81 Å in the reactants versus 1.77 Å in the product). At the same time,

the hydrogen bond established between Ser^{196} and His^{194} becomes weaker (1.77 Å in the reactants versus 1.87 Å in the product) since Ser^{196} is pushed by this water.

The calculations also reveal alterations in the conjugated π-system along the PLP cofactor, during the charge transfer effect. Two double bonds are formed between the Cα-N and the C6-C7 atoms because their length decreases from 1.45 Å and 1.48 Å (reactant) to 1.31 Å and 1.40 Å (product), respectively. At the same time, the double bond C8-N becomes single (1.29 Å in the reactants versus 1.37 Å in the product). These results indicate that the negative charge is delocalized along these three bonds, but it is not accommodated here since the charge variation at this position of the PLP is almost insignificant (+0.06 a.u.). The carbon Cα acquires a sp^2 configuration, becoming co-planar with its adjacent atoms (carbon Cβ, hydrogen Hα, and nitrogen N). As expected, no significant changes are observed with respect to the amino acid residues that are stabilizing the phosphate group of PLP cofactor.

Once the formation of carbon dioxide is complete, it leaves the active site, and the PLP cofactor remains covalently bonded to the quinonoid intermediate through an imine linkage.

The second step of this mechanism involves the protonation of carbon Cα from which results the external aldimine, in which the PLP cofactor is bonded to histamine, the final product. It is known that the proton that is required for such reaction comes from the solvent, but as the active site is not freely solvent accessible, there must be an amino acid that catalyzes this process. There was no previous evidence about which residues could protonate the quinonoid intermediate. Consequently, two hypotheses were tested as emerged from the MD simulations.

The MD simulations performed with HDC and the quinonoid intermediate show that the loop located in the top region of the active site (containing Tyr^{334} residue) precludes any direct interaction of the quinonoid intermediate with the solvent. The cluster analysis revealed that His^{194} and Tyr^{334} are close to carbon Cα. Among these two residues, Tyr^{334} is more likely to be involved in the proton transfer because it is located in a flexible loop that is exposed to the solvent. However, the participation of His^{194} in this process cannot be discarded because it establishes an important hydrogen bond with Tyr^{334} and can, therefore, mediate this process.

Based on the QM/MM structure of the product of the first step, two hypotheses were studied concerning the second step (**Figure 20**).

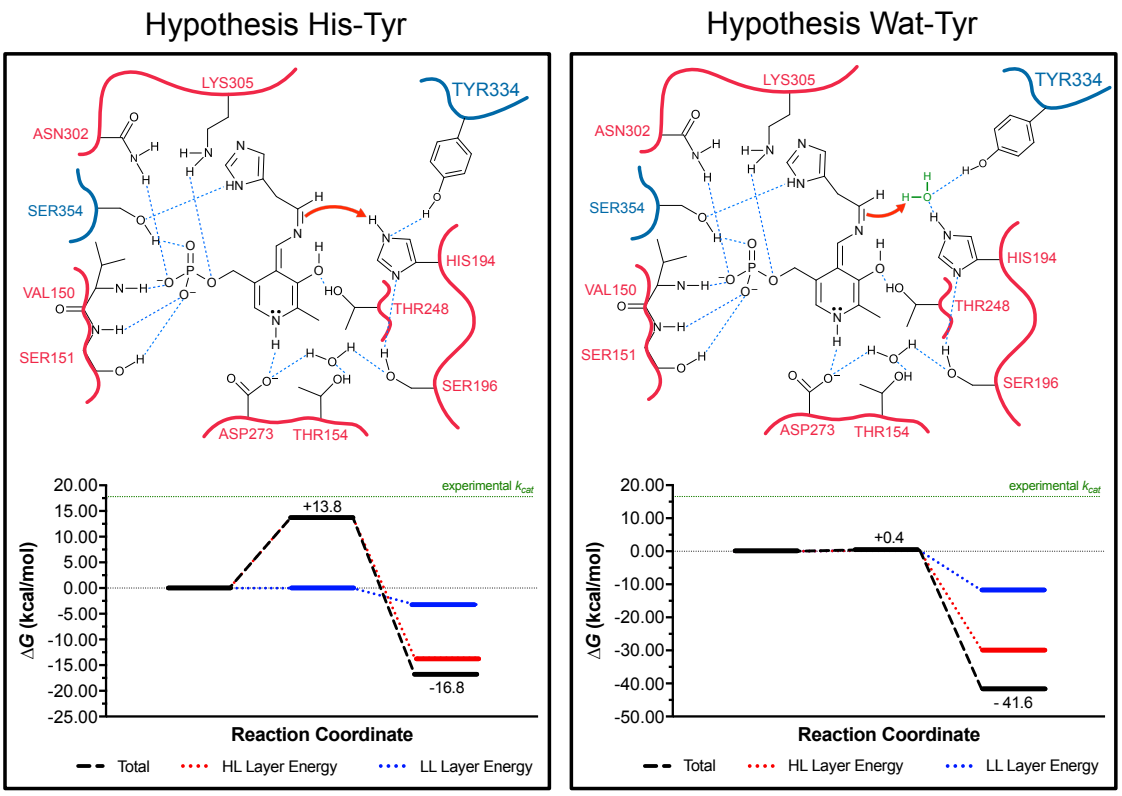


Figure 20. Two hypotheses were studied to address the catalytic mechanism for the second step. The structures colored in green correspond to those were modified comparing to the previous hypothesis. The energy profiles are shown for each hypothesis. Dashed lines colored by blue represent important non-bonded interactions at the active site. Energies were calculated from the difference between the optimized TS and the reactant.

In one model, the His^{194} residue mediates the proton transfer from Tyr^{334} to carbon Cα of the quinonoid intermediate (Figure 20 Left). In the second one, Tyr^{334} is directly involved in the proton transfer through a water molecule that shuttles the proton to the carbon Cα of the quinonoid intermediate (Figure 20 Right).

The system used to study the second step is identical to the one used to study the decarboxylation reaction, however, the sidechain of the amino acid residue Tyr^{334} was included in the HL layer, which now accounts for 140 atoms. In the case of the hypothesis Wat-Tyr, an additional water molecule was included in the HL layer, accounting additional three atoms.

The protonation of carbon Cα of the quinonoid intermediate by Tyr^{334} and mediated by His^{194} requires an activation energy of 13.8 kcal/mol and it is exothermic in 16.8 kcal/mol. The TS of this step was characterized by a unique imaginary frequency at 1517.8 cm^{-1}. However, this activation energy is prohibitive when we considered the whole mechanism, since the first step of the catalytic mechanisms is very endothermic and this reaction requires 13.8 kcal/mol. It means that the global activation energy of the full process would

be 27.4 kcal/mol, which is much greater than the one predicted by the k_{cat} (17.9 kcal/mol) and would not be possible under physiological conditions.

On the other hand, the proton transfer from Tyr^{334} to carbon Cα of the quinonoid intermediate was almost spontaneous. This reaction was only possible to characterize including a water molecule between the Tyr^{334} and the quinonoid intermediate. A water molecule can occupy this position because, upon the release of carbon dioxide to the solvent, a free position is available in the active site that can promptly be occupied by a water molecule. The MD simulations confirm that Tyr^{334} is located in a loop with high flexibility and in direct contact to the solvent, and therefore, it can enhance this event.

Once a water molecule enters the active site, Tyr^{334} can mediate the proton transfer to the quinonoid intermediate using this water molecule as a proton shuttle.

In the reactant state, we observed the same conformation of the product from the first step. However, in this case, a water molecule occupies the region of the released carbon dioxide molecule, and one of its protons is ready to undergo a nucleophilic attack from the Cα atom. The oxygen atom belonging to this water molecule is surrounded by two hydrogen bonds established with His^{194} and Tyr^{334} residues. When the reaction occurs, and the proton passes to the Cα carbon of the quinonoid intermediate, the generated hydroxyl ion is simultaneously protonated by the hydroxyl group of the Tyr^{334} residue (**Figure 21**).

The energy profile of this reaction was calculated, and all the minima were characterized including the TS which has solely one imaginary frequency at 429.6 cm^{-1}. The reaction is almost spontaneous with an activation Gibbs free energy of 1.9 kcal/mol. In contrast with the first step, the protonation of the quinonoid intermediate is exergonic by 33.2 kcal/mol.

Although the direct participation of His^{194} in the protonation step can be excluded, it continues to have an important role in the process. The reaction mediated by Tyr^{334} and the water molecule is only possible due to the hydrogen bond that it establishes with His^{194} residue, promoting, therefore, the correct position and alignment of the water molecule in relation to the Cα carbon and the Tyr^{334} residue. These results are in agreement with the experimental data that shows that the mutation of His^{194} by a Gly decreases the catalytic efficiency by almost 12 times, but it does not prevent the reaction as it does when Tyr^{334} residue is mutated [268].

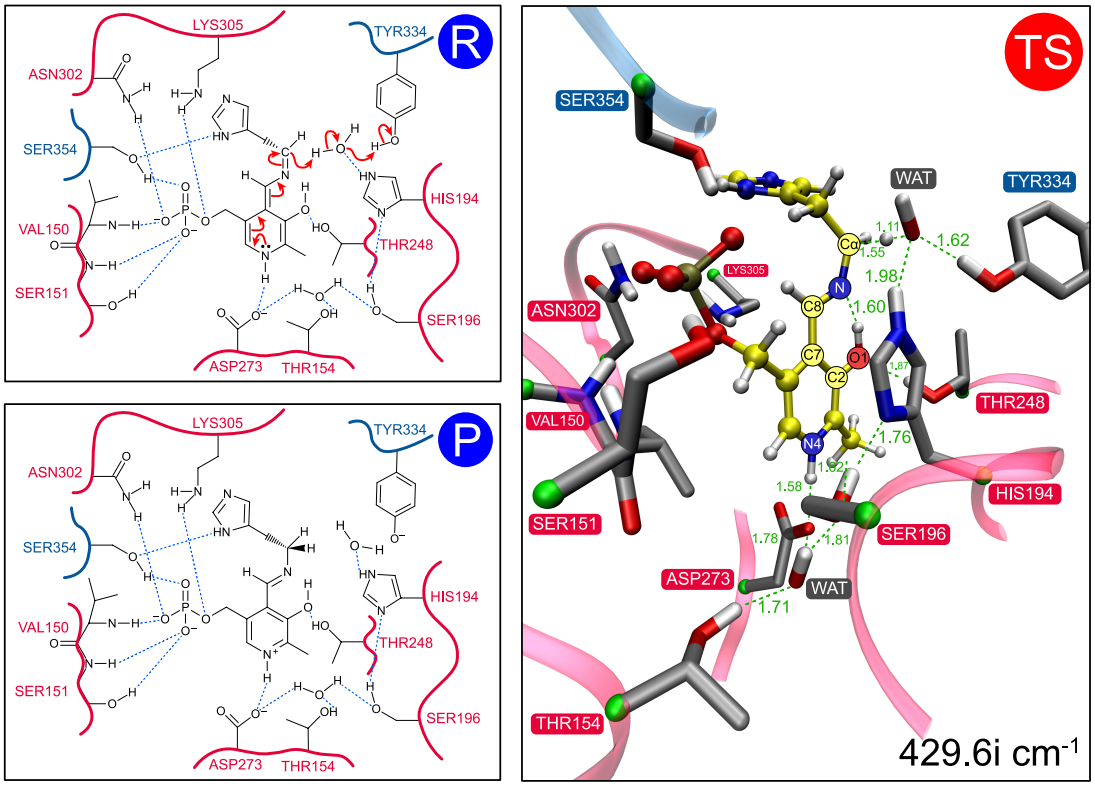


Figure 21. Structures of reactant, TS, and product of the second step of the reaction catalyzed by HDC. (Left) Wedge-Dash representation of the structures of reactant and product. (Right) New cartoon and licorice representations of the TS structure of the active site.

Once more, the PLP cofactor has a preponderant role in this second step, but in the different way. When the first step occurs, the PLP cofactor becomes negatively charged favoring, by this way, the decarboxylation process. In the second step, the reverse process occurs and allows the charge to become lodged at the Cα carbon, turning it into a better nucleophile atom to accept one proton. This process is followed by an electronic delocalization that inverts the conjugated π-system of C7, C8, N, and Cα atoms. The double bonds between C7-C8 and N-Cα atoms become single again, since the interatomic distances increase from 1.42 Å and 1.33 Å (reactant) to 1.48 Å and 1.47 Å (product), respectively. On the other hand, the single bond between C8-N atoms becomes double (1.35 Å in the reactants versus 1.28 Å in the product). Simultaneously, the quinonoid intermediate becomes less negative (-1.29 a.u. in the reactant versus -1.00 a.u. in the product) and the hydrogen bond between the protonated nitrogen N4 and the carboxylate group belonging to Asp^{273} residue becomes stronger (1.62 Å in the reactant versus 1.50 Å in the product). As a result, the negative charge of Asp^{273} residue is stabilized, and the intermolecular interaction with the neighbor water molecule turns weaker (1.77 Å in the reactant versus 1.81 Å in the product). At the end, the hydroxyl group of Tyr^{334} residue is deprotonated and negatively charged. However, due to its

exposure to the solvent, the Tyr^{334} residue can easily be re-protonated by a water molecule or any other acid molecule in the cell.

1.1.4. Conclusion

This work described, for the first time, the catalytic mechanism of a PLP-dependent decarboxylase, in particular the HDC, with an atomistic level of detail.

The reaction was found to take place in two sequential steps. The first step is the rate-limiting one and involves the decarboxylation process. It requires a free activation energy of 17.6 kcal/mol, and it is endergonic by 13.7 kcal/mol. The second step involves the protonation of the quinonoid intermediate by Tyr^{334} with the direct participation of a water molecule and with the help of His^{194}. This step is almost spontaneous, requiring a free activation energy of 1.9 kcal/mol, and it is also very exergonic (-33.2 kcal/mol) (Figure 22).

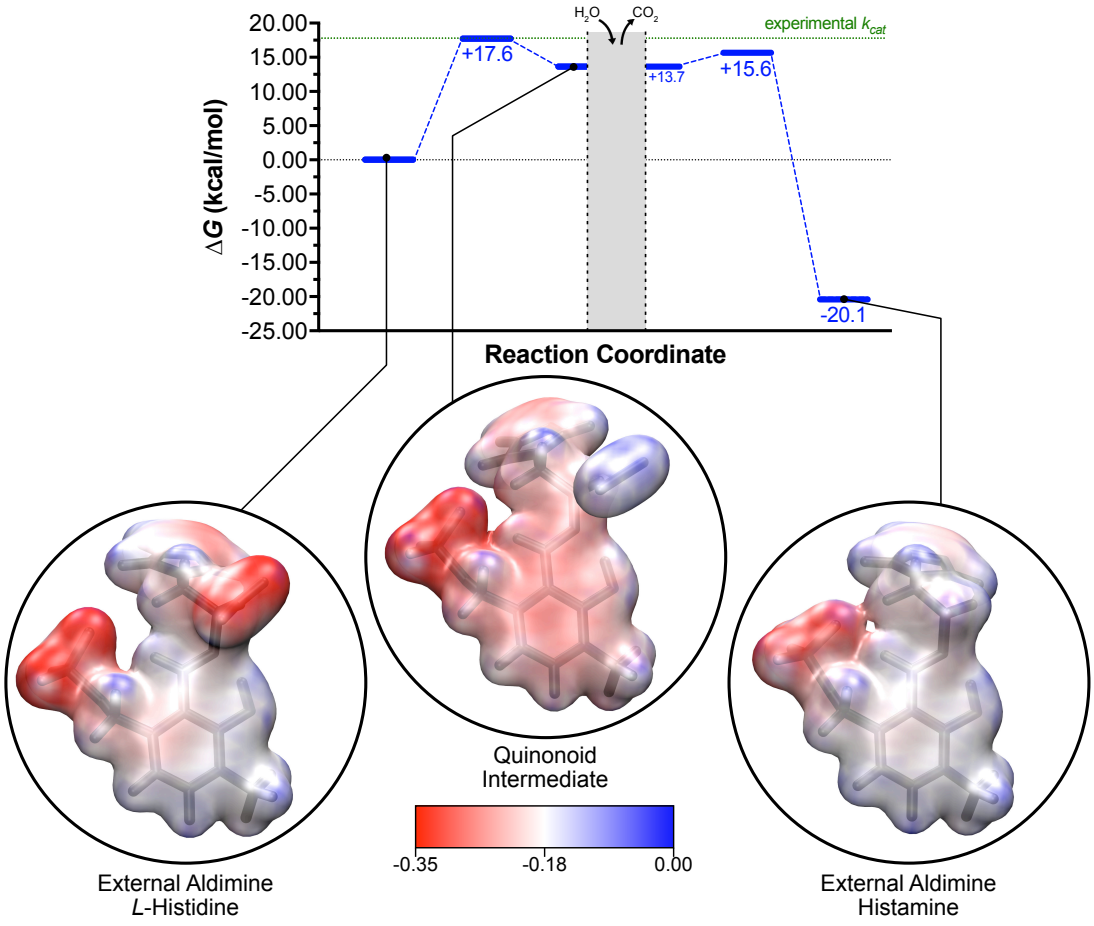


Figure 22. (Top) The energy profile for the two specific steps catalyzed by HDC. (Bottom) Surface mapping of the Merz-Singh-Kollman charges of the external aldimines with the reactant and product and quinonoid intermediate.

The global energy profile of the full mechanism of HDC agrees very well with the experimental k_{cat} (1.73 s^{-1}) that is equivalent to an activation barrier of 17.9 kcal/mol. Moreover, it also corroborates the experimental observation that shows an irreversible

conversion of histidine into histamine. According to the computed energies, the reverse reaction would require more than 30 kcal/mol to occur, which make the reverse reaction extremely slow to occur under biological conditions.

All the calculations performed to study the catalytic mechanism of HDC were based on the only available X-ray structure of this enzyme (PDB code 4E1O) in which the enzyme was co-crystalized with an inhibitor (HME). The inhibitor was very similar to the natural substrate, and therefore, it was expected that the natural substrate would interact with the active site and with the PLP cofactor in a similar conformation. However, found out that this was not the case. The calculations indicate that the orientation of the imine linkage, as it is found in the co-crystallized X-ray structure, prevents the formation of an important keto-enol tautomerism between the NH group of the substrate and the ketone group of the PLP cofactor. This interaction was found to be essential for the decarboxylation process. These two groups must be pointing to the same side of the imine linkage and interacting by a hydrogen bond, similar to what is found in other PLP-dependent enzymes [266, 294]. This means that the configuration of the external aldimine present in the co-crystallized X-ray structure is a consequence of the inhibitory process and should not be present under normal catalytic conditions.

The calculations disclosed an important interaction between the protonated nitrogen belonging to the imidazole group of the substrate and the hydroxyl group of Ser^{354}. The position that is occupied by this amino acid residue is known to be crucial for the control of substrate specificity of the PLP-dependent enzymes that catalyze decarboxylation reactions [261]. What differentiates the substrates among these PLP-dependent enzymes is the side chain of their amino acid substrates, and therefore, in the case of HDC, the interaction of the side chain of L-histidine with Ser^{354}. This interaction promotes the correct positioning of the substrate inside the active site, favoring the reaction.

This work also showed, for the first time, the important role of the Tyr^{334} residue. This residue is located in a flexible loop that is highly exposed to the solvent and was found to be important for the protonation of the quinonoid intermediate, from which results an external aldimine in which the PLP cofactor is bonded to the product of the reaction (histamine).

An important network of hydrogen bonds, involving His^{194}, Ser^{196}, a water molecule, and Asp^{273}, was also disclosed. This set of intermolecular interactions plays a key role to stabilize the negative charge of Asp^{273} when the PLP accommodates the negative charge of the quinonoid intermediate.

The role of PLP was also understood for these enzymes that catalyze decarboxylation of amino acids. During the first step, PLP cofactor triggers the decarboxylation process through the withdrawal of the electron density from the Cα-C bond. Once the quinonoid intermediate is formed, the negative charge is accommodated through the PLP cofactor making it ready to the protonation step that takes place in the second step of the catalytic process (Figure 22 Bottom). The PLP cofactor has, therefore, a double role that is fundamental for catalysis. It is able to spread the electron density in itself, in order to promote the bond C-Cα cleavage, and, since PLP does not kidnap the negative charge, it is able to make the electron density available for the protonation of the carbon Cα. Otherwise, the second step would never occur.

This mechanistic proposal also goes in line with what was proposed by Dunathan, where the bond that is formed or broken during the catalytic process should be placed perpendicularly to the conjugated system of the PLP cofactor [295, 296]. In the case of HDC, the bond that is cleaved (Cα-C) is closer to a perpendicular position in relation to the PLP cofactor since the dihedral angle C2-C7-Cα-C is -77.9° for the reactant and -72.2° when the TS is formed (Figure 23).

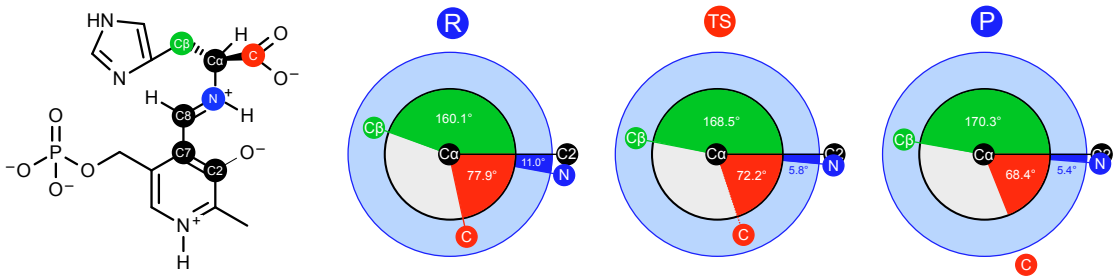


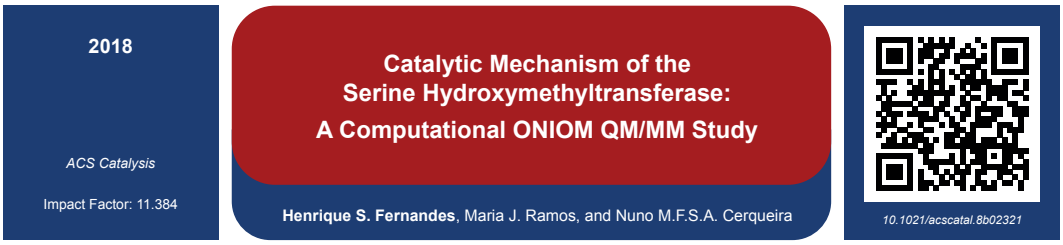
Figure 23. Analysis of the relevant dihedral angles of the reactant, TS, and product for the first step. Dihedral angles: C2-C7-Cα-C; C2-C7-Cα-Cβ; C2-C7-C8-N.

The conclusions that were obtained with this work can also give a hint about the catalytic mechanism of other enzymes belonging to family II of PLP-dependent decarboxylases. This family of enzymes includes HDC, aromatic-L-amino-acid decarboxylase (AADC) (EC 4.1.1.28), and glutamate decarboxylase (GAD) (EC 4.1.1.15). The only difference that is observed in the active site among these enzymes is the residue that in HDC is occupied by Ser^{354}. This residue makes an important interaction with the side chain of the substrate and, therefore, it is important to modulate the specificity of the enzyme. In spite of this difference, no changes are expected regarding the catalytic mechanism since this residue does not play a catalytic role during catalysis.

1.2. α-Elimination

The catalytic mechanism for the α-elimination catalyzed by PLP-dependent enzymes were studied using the serine hydroxymethyltransferase (SHMT) enzyme as example [188]. SHMT was chosen because of its importance as a drug target against malaria.

All the calculations were performed by Henrique S. Fernandes, as well as, the writing of the article, which was revised considering the contributions of all the authors. The work was published in the ACS Catalysis journal, and the following content is almost a transcription of the published version.



SHMT is an important drug target to fight malaria, which is one of the most devastating infectious diseases, with 216 million cases cited and accounting for ~450 000 deaths in 2016. In this paper, computational studies were performed to unveil the catalytic mechanism of SHMT using QM/MM methodologies. This enzyme is responsible for the extraordinary cyclization of a tetrahydrofolate (THF) into 5,10-methylene-THF. This process is catalyzed by a PLP cofactor that binds L-serine and, from this, one molecule of glycine is produced. The results show that the catalytic process occurs in eight sequential steps that involve an α-elimination, the cyclization of the 5-hydroxymethyl-THF intermediate into 5,10-methylene-THF, and the protonation of the quinonoid intermediate. According to the calculated energetic profile, the rate-limiting step of the full mechanism is the elimination of the hydroxymethyl group, from which results a formaldehyde intermediate that then becomes covalently bonded to the THF cofactor. The calculated barrier (DLPNO–CCSD(T)/CBS:ff99SB) for the rate-limiting step (18.0 kcal/mol) agrees very well with the experimental kinetic results (15.7–16.2 kcal/mol). The results also highlight the key role played by Glu^{57} during the full catalytic process and, particularly in the first step of the mechanism, where the anionic Glu^{57} is required, contrasting with some proposals available in the literature for this step. It was also concluded that the cyclization of THF must occur in the enzyme, rather than in solution, as it has been proposed also in the past. Together, all of these results provide knowledge and insight on the catalytic mechanism of SHMT, which can now be used to develop inhibitors targeting SHMT and, therefore, antimalaria drugs.

This work was also presented in the following conferences:

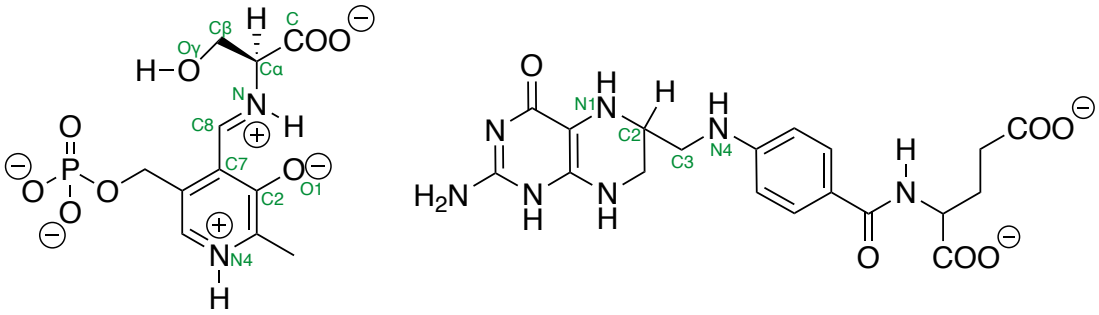
- (Poster) *III Symposium on Medicinal Chemistry of University of Minho*, **2017**, Braga, Portugal (Appendix: page 381)
- (Oral Communication) *7th EuCheMS Chemistry Congress*, **2018**, Liverpool, United Kingdom
- (Poster) *UCIBIO Annual Meeting*, **2018**, Lisboa, Portugal (Appendix: page 382)
- (Poster – Award for the best poster) *6º Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, **2018**, Porto, Portugal (Appendix: page 383)

1.2.1. Introduction

Malaria is one of the most devastating infectious diseases that is caused by a parasitic infection transmitted by mosquitoes. According to the last World Health Organization (WHO) report, 228 million new cases of malaria and 405 thousand deaths were caused by this disease in 2018. Despite the resources and investments allocated to find a universal cure for this disease (2.7 billion dollars in 2018), the control and eventual elimination of malaria is seriously questioned due to the resistance of parasites to antimalarial drugs. This effect has become even more dramatic in five countries of the Greater Mekong Subregion and in the Cambodia-Thailand border, where *Plasmodium falciparum* (the species that causes the deadliest form of malaria) has become resistant to the major antimalarial drugs used clinically [297, 298].

All of these facts together demand an urgent need for new drug targets, which can be used to design new and more efficient therapeutics against the malaria pathogen. Particular attention is being given to those that provide a selective interference with the parasite metabolism without harming the human host [297, 299, 300]. In this regard, some of the most promising targets are enzymes involved in the *de novo* purine biosynthesis in the malaria parasites, such as SHMT [301-304].

The SHMT (EC 2.1.2.1) is a PLP-dependent enzyme that catalyzes the reversible transfer of a carbon-unit from the L-serine to a second cofactor, the THF (Figure 24). At the end of the reaction, a glycine molecule is released as well as a 5,10-methylene-THF molecule, which is a precursor in the synthesis of purines and thymidylate [36]. Since the purines biosynthesis is vital for the pathogen survival (synthesis of nucleic acids), the inhibition of SHMT has been pointed out as an important drug target for malaria [301]. Furthermore, the comparison between the tridimensional structure of the human and *P. falciparum* forms of the SHMT revealed key differences that can help to rationalize the specific synthesis of new inhibitors for the parasitic form of the SHMT [305].



Pyridoxal-5'-Phosphate
PLP

Tetrahydrofolate
THF

Figure 24. Structure of the external aldimine (PLP + L-serine) (Left) and the THF cofactor (Right).

Despite the importance that SHMT has gathered in the last two years, this enzyme continues to be poorly understood. This happens due to the complex chemistry that it catalyzes, which requires the participation of two cofactors (PLP and THF) and the multi sequential steps needed for the cyclization of the THF into the final product, 5,10-methylene-THF. This knowledge is very important since it will provide clues about the enzymatic activity and how it can be inhibited. This information can then be used to develop new inhibitors targeting this enzyme with potential application as new anti-malaria drugs.

Similarly to other PLP-dependent enzymes, the catalytic activity of the SHMT is composed of three stages. The first two stages are common to all PLP-dependent enzymes and were already studied [27, 28], involving the activation of the enzyme and the formation of the external aldimine (transimination). In this work, the third stage was studied, which corresponds to the set of reactions that are specific for this enzyme. In particular, the catalytic mechanism for the α-elimination of L-serine that is subsequently inserted into the THF cofactor (**Figure 25**).

The experimental kinetic data available for the *E. coli* form of the SHMT indicates an activation barrier limit of 15.7 – 16.2 kcal/mol based on the available turnover number of 686.6 ± 21.7 min^{-1} (pH 7.2, 20 °C) [306] and 790 min^{-1} (pH 7.2, 30 °C) [307].

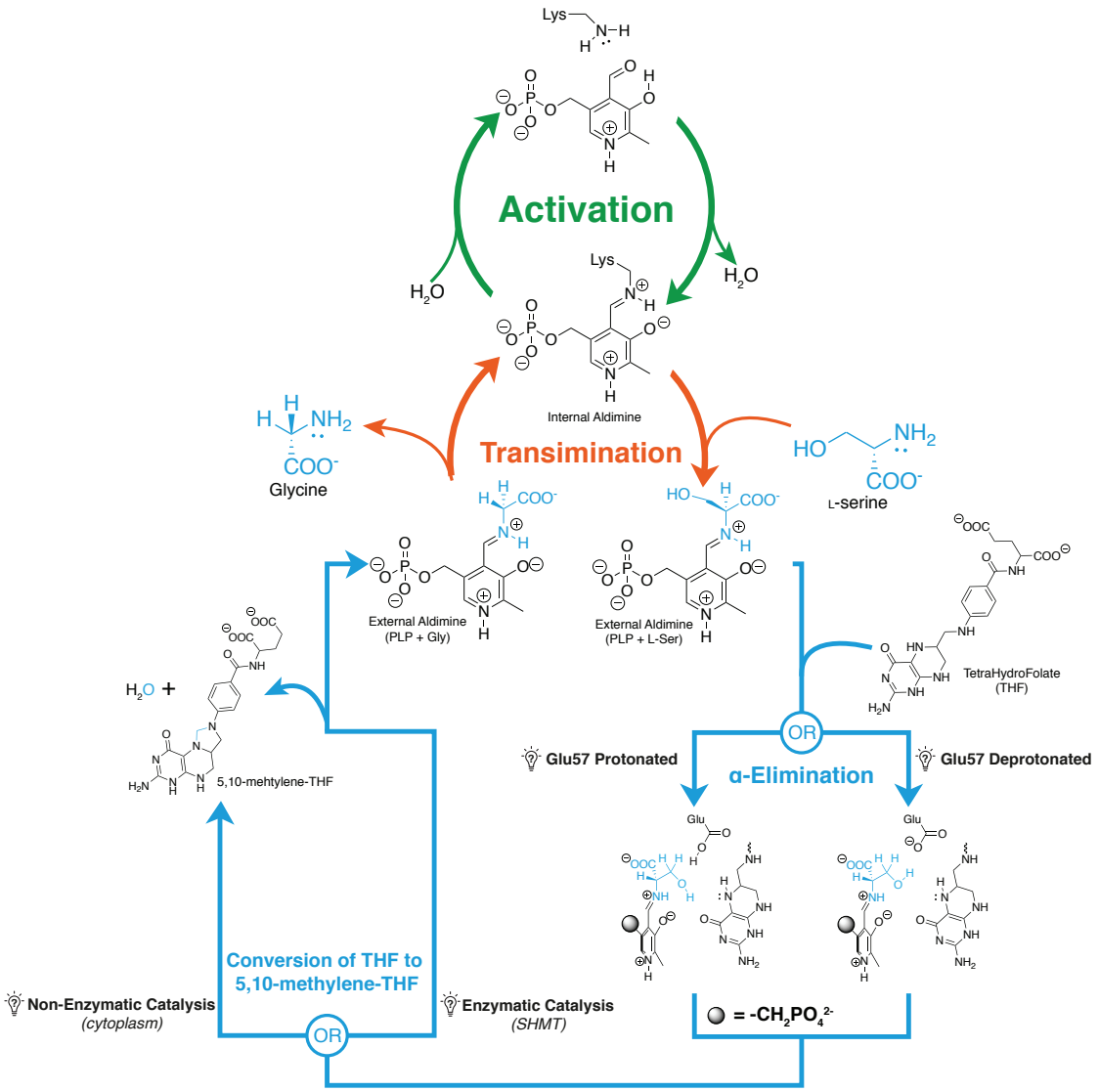


Figure 25. Scheme representing the mechanism behind the internal and external aldimine formation for PLP-dependent enzymes and the specific α-elimination reaction.

Until this work has been conducted, there was no accepted mechanism for the α-elimination catalyzed by SHMT, and only two hypotheses were available in the literature. They diverge in the way through which the hydroxymethyl group is transferred from serine (substrate) to the THF cofactor. One of the hypotheses suggests that the reaction takes place in one step, where nitrogen N1 of THF makes a nucleophilic attack on the Cβ of the serine promoting the α-elimination reaction. In the other hypothesis, the reaction requires two steps. The first step involves the formation of a formaldehyde molecule (through the dissociation of the hydroxymethyl group from the external aldimine), which then undergoes a nucleophilic attack from nitrogen N1 of the THF cofactor [308]. These two hypotheses also diverge regarding the protonation state of the Glu^{57} residue, which interacts very closely with the hydroxyl group of the external

aldimine. In the first proposal, Glu^{57} is proposed to be protonated, whereas, in the second proposal, the mechanism requires an unprotonated Glu^{57} [308-310].

The role played by Glu^{57} in the reaction is known to be fundamental for the catalysis since several mutagenic studies have shown that its mutation by Gln, Leu, or Lys residues reduces the activity of the enzyme drastically (more than 300 times), but the role played by this residue in the mechanism continues to be poorly understood [309, 311].

Some authors also suggest that the reaction of the formaldehyde molecule with THF can occur in the solvent and not in the enzyme [308, 312]. However, so far, no experimental result is currently available to provide insights into this hypothesis.

In this work, the catalytic mechanism of the SHMT was studied employing a ONIOM QM/MM methodology. Based on the mechanistic proposals available in the literature, and described above, two models of the enzyme were set up to judge the role of the protonation state of the Glu^{57} residue in the catalytic mechanism and also to understand if the hydroxymethyl group is transferred directly to the THF cofactor or it involves the formation of a formaldehyde intermediate in an intermediate step. A third model was also built to study the possibility in which the cyclization of the THF intermediate occurs in the solvent and does not require the presence of the enzyme.

1.2.2. Methodology

a. Model Preparation

This study was conducted using the crystallographic structure of SHMT from *Escherichia coli* deposited in the PDB with the reference 1DFO [310]. Although the aim to study the SHMT was studying the SHMT from *Plasmodium falciparum*, this PDB structure for the *E. coli* form was chosen because it contains the external aldimine with the product of the reaction (PLP + glycine) and an analog of first THF intermediate. This sort of data is not available in the PDB file of SHMT from *P. falciparum* (PDB code: 4O6Z) [313]. Although the identity between both proteins is not ideal (43%, Figure 26 A) (BLASTP 2.8.0+) [314, 315], the structure of both proteins (Figure 26 B) is very similar, and the region of the active site can be almost overlapped (Figure 26 C).

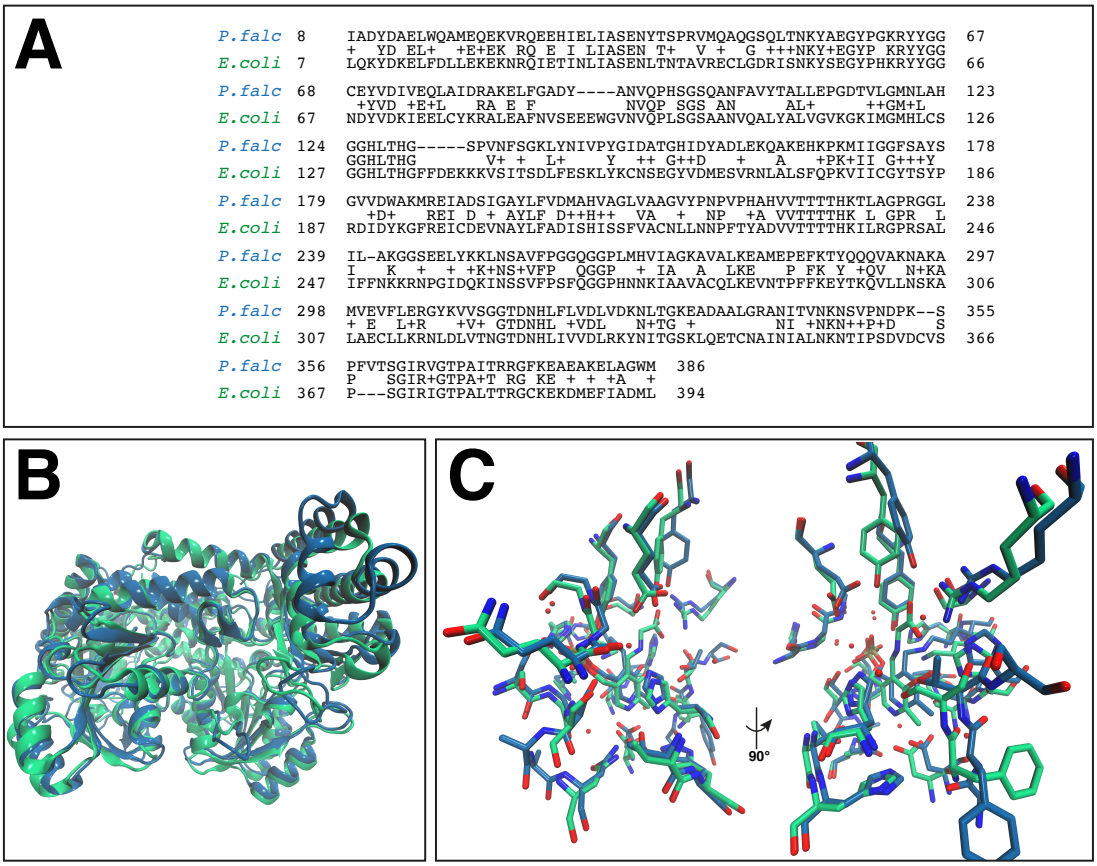


Figure 26. (A) Alignment of the protein sequences of the *E.coli* and *P. falciparum* forms of the SHMT (BLASTP 2.8.0+). Overlapping of the structures of (B) all dimer and (C) the active site residues and PLP cofactor from both forms of the SHMT enzyme: *E.coli* (green) and *P. falciparum* (blue).

The model system that was used in the computational studies includes one of the homodimers of SHMT that is present in the crystallographic structure with the PDB code 1DFO. The external aldimine (PLP + glycine) available on the PDB was used to model the quinonoid intermediate of the reaction, deleting only one proton from the Cα carbon

of the substrate. The 5-formyl-THF inhibitor was used as a template to build the 5-hydroxymethyl-THF intermediate that results from the transfer of the hydroxymethyl group during the first step of the reaction.

From this model, two additional models were built to accommodate different protonation states of the Glu^{57} residue. These models were entitled “Model 1” and “Model 2”.

An additional model was built to study if the formation of the 5,10-methylene-THF could occur in a non-enzymatic process. This model (named “Model 3”) only contains the 5-hydroxymethyl-THF molecule that results from the reaction of the THF cofactor with the formaldehyde molecule (that takes place inside the enzyme). To model the solvent effect an implicit solvent model was employed (water, ε = 78.3553) using the IEFPCM formalism integrated with Gaussian09 [280].

b. MM Minimization

The protein, the quinonoid intermediate, and the THF cofactor were solvated by using TIP3P [273, 274] type water molecules. The protein was placed inside a water box so that each atom of the protein distanced at least 12 Å from each face. The SANDER software from the AMBER12 package was used to add the hydrogen atoms to the model [277]. The protonation states for all amino acids were predicted by PROPKA 3.1 [275, 276] at pH 7.0, excepting for the Glu^{57} residue whose protonation state was tested in models 1 and 2.

The 5-hydroxymethyl-THF and external aldimine molecules were parametrized performing a HF/6-31G(d) optimizations to obtain the atomic charges and the antechamber software from AMBER12 package [147, 148, 277] to assign the atom types.

The geometry of the full system was minimized through four sequential steps using the AMBER12 package and the GAFF and ff99SB force fields [147-149, 277-279]. Firstly, minimization of all water molecules was conducted, and afterwards all of the hydrogen atoms. Subsequently, all atoms from the system, except backbone, were minimized, and finally, a minimization of all system with no constraint was conducted.

c. QM/MM Calculations

The ONIOM QM/MM model was built loading the minimized structure from the previous procedure, using the molUP plugin [316] for VMD software [292]. The QM/MM model included the full dimer, the two quinonoid intermediates, the two 5-hydroxymethyl-THF molecules, and a 5.0 Å coat of water molecules.

The atoms included in the HL layer were adjusted from step by step since different types of reactions take place during the full catalytic process. In the α-elimination reactions, the HL includes the atoms from the external aldimine, THF cofactor, and some active site residues that in total account to 106 atoms (Arg^{363}, Asp^{200}, His^{129}, Asn^{102}, Glu^{57}, Ser^{35}, and a water molecule). In the study of the conversion of the THF intermediate to 5,10-methylene-THF only the THF intermediate and the active site residue Glu^{57} were included in the HL layer, accounting for 85 atoms. In the last stage of the mechanism, the similar set of atoms used in the α-elimination was used (121 atoms in total) (Figure 27). The remaining atoms of the system were included in the LL layer. Hydrogen atoms were used as link atoms to complete the valence of the bonds crossing between the two layers of the ONIOM QM/MM scheme.

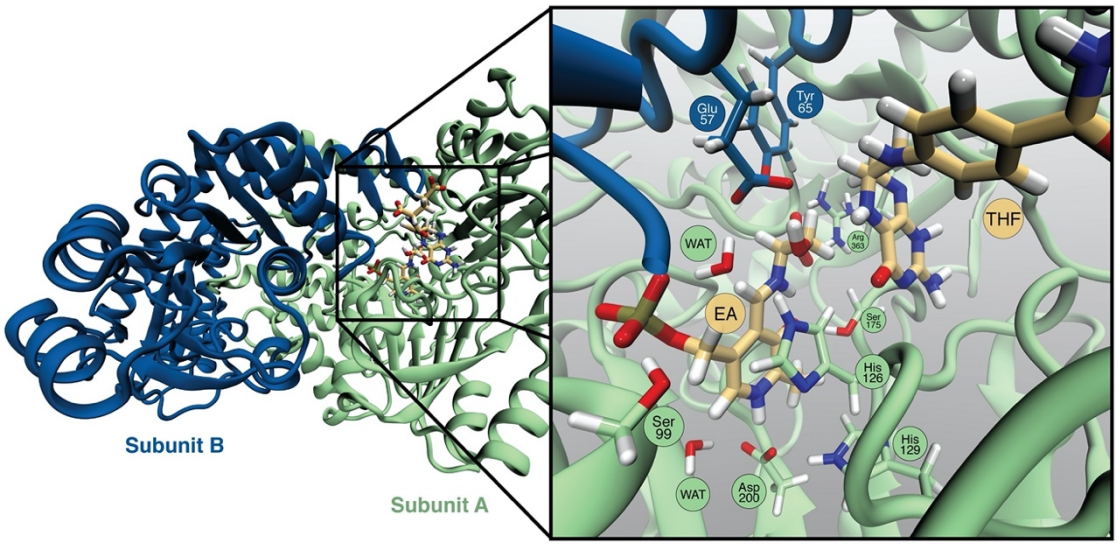


Figure 27. (Left) Structure of the SHMT dimer (subunit A: green; subunit B: blue) and licorice representation of the HL layer. (Right) The substrate, cofactors and the active site residues are represented in licorice and colored according to the subunit that they belong to (subunit A: green; subunit B: blue). Each molecule is identified by a circular tag.

This work was conducted using the same methodology described for the study of the catalytic mechanism of HDC (Chapter C – 1.1.2. page 95). However, single-point (SP) energy calculations for all TSs, reactants, and products were calculated using DLPNO-CCSD(T)/CBS, for the HL layer. The HL layer was isolated to perform these calculations using the ORCA software (version 4.0.1.2) [250]. The SPs were carried out using the cc-pVDZ and cc-pVTZ basis sets [185], and the cc-pVDZ/C and cc-pVTZ/C correlation fitting basis sets [317, 318], respectively. The combination of the energies obtained with cc-pVDZ | cc-pVDZ/C and cc-pVTZ | cc-pVTZ/C basis sets for the HL layer were used to extrapolate to the complete basis set (CBS) limit according to the Truhlar's extrapolation

approach [319], excepting for the α and β exponents. The values 4.42 and 2.46 were considered as α and β exponents in the extrapolation according to the ORCA documentation.

Concerning the interaction between the HL and LL layers, the mechanical embedding method was used to perform all the calculations because the electrostatic embedding scheme is computational very expensive considering the size of the system and the number of reactions studied in this work.

1.2.3. Results and Discussion

The computational calculations regarding the catalytic mechanism of SHMT were divided into three main stages: i) the first stage involved the α-elimination process; ii) the second stage, the conversion of the THF intermediate to 5,10-methylene-THF; and iii) the third stage, the protonation of the quinonoid intermediate (QI) and subsequent enzymatic turnover. Each of these stages is described and discussed in detail in the following subsections of this thesis.

a. First Stage: α-Elimination

The α-elimination of the hydroxymethyl group of the serine is the main point of discussion considering the catalytic mechanism of the SHMT. In the literature, there are two proposals for this reaction. In the first one, it is proposed that Glu^{57} is protonated and involves the nucleophilic attack of nitrogen N1 from THF to carbon Cβ of the substrate and from which results the elimination of the hydroxymethyl group [308-310]. (Figure 28 Top)

The other hypothesis proposes that the Glu^{57} residue is not protonated and, at the beginning of the reaction, it abstracts a proton from the hydroxyl group of the substrate. From this reaction results the formation of formaldehyde that afterwards reacts with the nitrogen N1 from THF and becomes covalently bonded to it (Figure 28 Bottom).

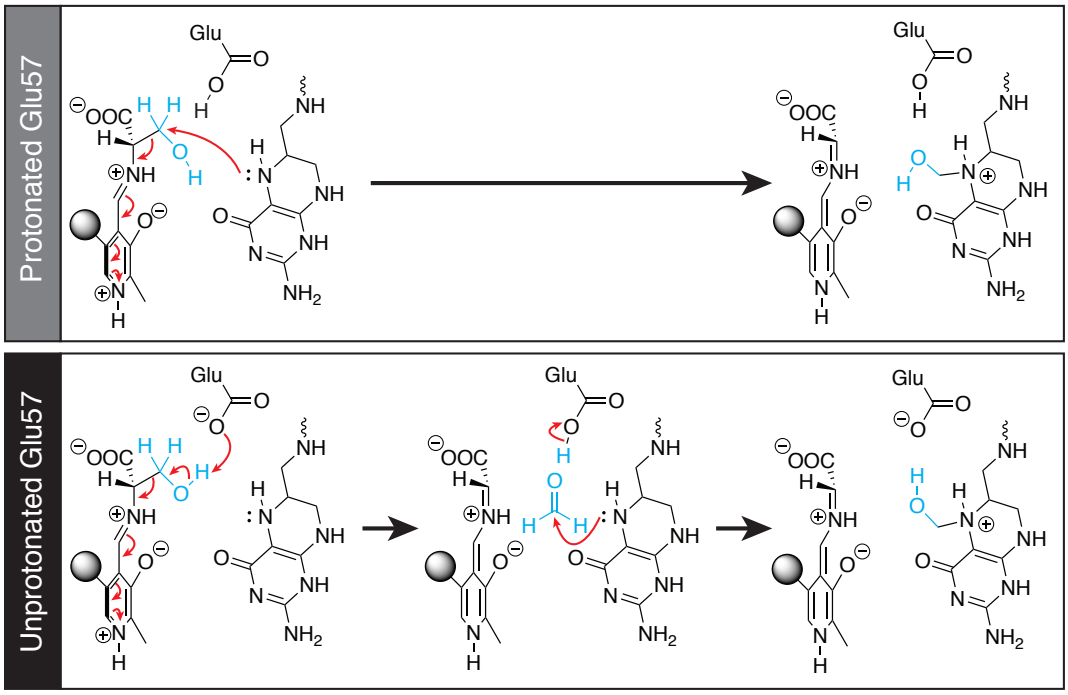


Figure 28. Schematic representation of the two mechanistic proposals for the α-elimination of serine that is bonded to the PLP cofactor and forms the external aldimine, considering different protonation states of the Glu^{57} (Top: protonated and Bottom: unprotonated). The spheres represent the phosphate group of the PLP cofactor.

Unprotonated Glu^{57}

When Glu^{57} is not protonated, it interacts with the hydroxyl group of the external aldimine by a hydrogen bond, and this interaction endorses a good alignment of the latter group with the THF cofactor. This interaction will be very important during the two steps that are required for the α-elimination process.

The first step is characterized by a proton transfer between the hydroxyl group of the external aldimine and the negatively charged Glu^{57}. The computational calculations also show that this reaction triggers, simultaneously, the cleavage of the Cα-Cβ bond (the α-elimination reaction), and the formation of a double bond between Oγ Cβ atoms (1.38 Å in the reactant versus 1.23 Å in the product). At the end of this reaction, one molecule of formaldehyde is formed, Glu^{57} becomes protonated, and the PLP cofactor adopts a quinonoid configuration that is negatively charged (Charge of PLP's ring: +0.18 a.u. in the reactant versus -0.04 a.u. in the product) (**Figure 29** Step 1). This reaction has an activation Gibbs free energy of 18.0 kcal/mol, and it is endergonic in 8.1 kcal/mol.

The TS structure of this reaction is characterized by one imaginary frequency at 229.1i cm^{-1}, which is associated with the vibration of the atoms involved in the reaction (**Figure 29** – Step 1 Center). In the TS structure, the bond that is cleaved is perpendicular to the PLP cofactor ring (Cβ-Cα-C7-C2 angle: 98.6°), a condition that is important in the chemistry of the PLP cofactor to have the reaction kinetically favored [295, 296, 320].

The second step involves the nucleophilic attack of N1 atom of THF to Cβ carbon of the recently formed formaldehyde, and the simultaneous proton transfer from Glu^{57} to Oγ atom of the formaldehyde. At the end of this reaction, a hydroxymethyl group becomes covalently bonded to the THF cofactor, which turn out to be positively charged (+1.74 a.u. in the reactant versus -1.17 a.u. in the product). In this step, the quinonoid intermediate that resulted from the first step does not participate in the reaction (**Figure 29** Step 2).

The TS of this reaction is characterized by one imaginary frequency at 206.8i cm^{-1}, in which Cβ carbon of formaldehyde adopts a sp^3 hybridization and the Cβ-Oγ bond becomes elongated, preparing it for the nucleophilic attack endorsed by the THF and the proton transfer from Glu^{57}, respectively.

This reaction is very fast account only 2.0 kcal/mol for the activation Gibbs free energy, and it is thermoneutral.

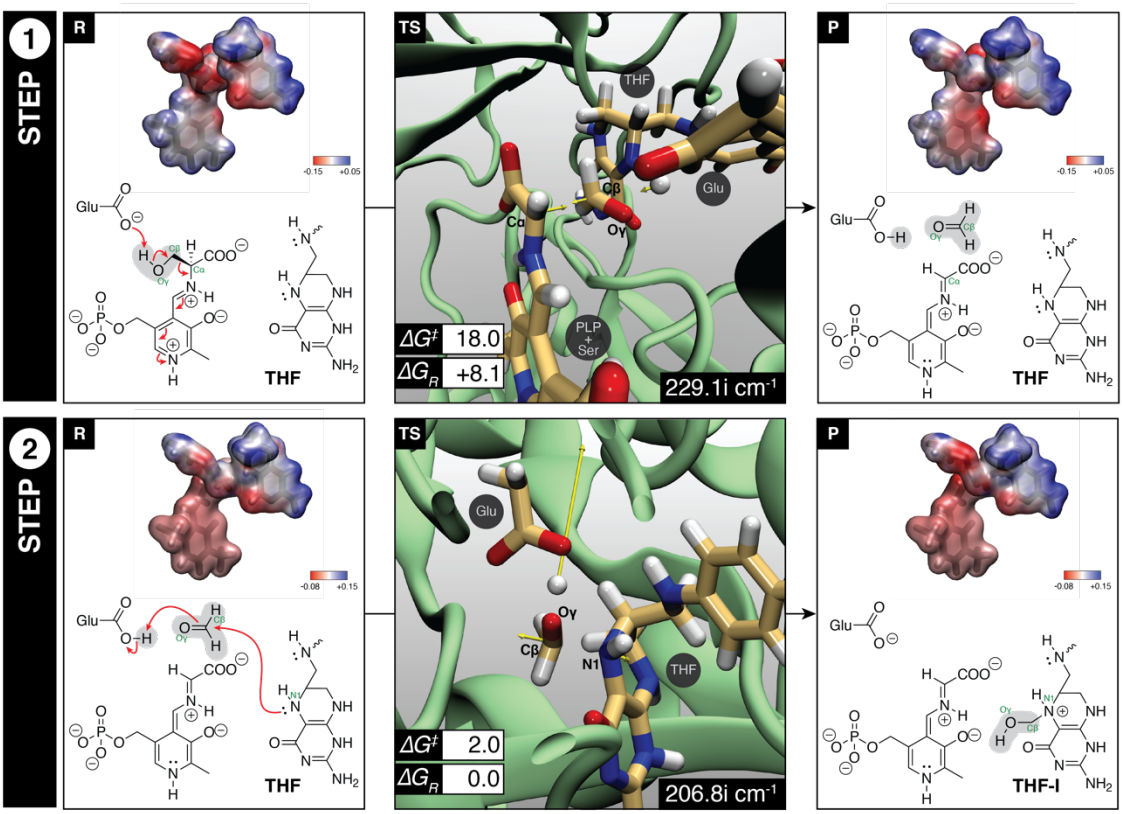


Figure 29. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the α-elimination reaction catalyzed by SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction.

Protonated Glu57

When Glu[57] is protonated, the α-elimination takes place in a single step. However, the computational results show that the nucleophilic attack of N1 nitrogen from THF to Cβ carbon of the external aldimine is impossible under biological conditions since the computed activation barrier is too high ($\Delta E^{\ddagger}$ = 53.9 kcal/mol) to occur at reasonable conditions of temperature and pressure. In addition, the reaction is also very endothermic (ΔE_R = 46.7 kcal/mol) which results from the lack of stabilization allowed by Glu[57], contrasting to what was observed with the previous model. These results indicate therefore that Glu[57] cannot be protonated during the α-elimination reaction catalyzed by the SHMT (Figure 30).

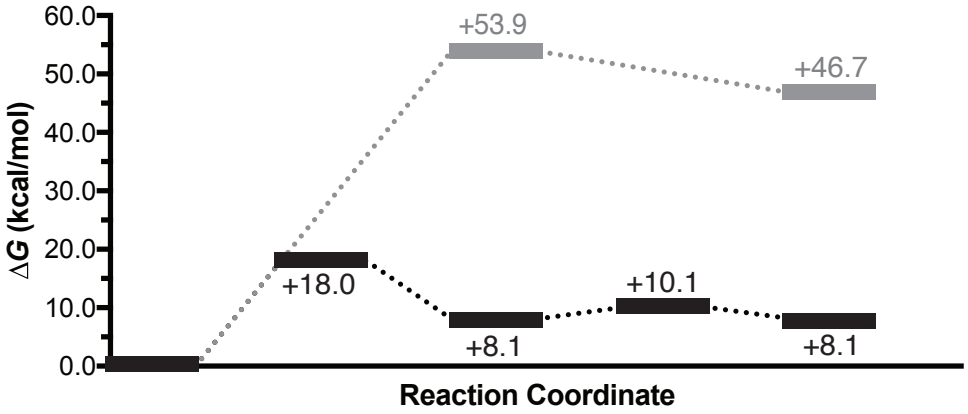


Figure 30. Energy profile of the two evaluated mechanisms for the α-elimination process catalyzed by SHMT with Glu[57] protonated (Δ*E*) (gray) or unprotonated (Δ*G*) (black).

b. Second Stage: Conversion of THF intermediate into 5,10-methylene-THF

After the α-elimination took place, the THF is covalently bonded to a hydroxymethyl group. The next step involves the cyclization of this reaction intermediate (THF-I) to yield one of the final products of the reaction, the 5,10-methylene-THF.

In the literature, there are two hypotheses for this reaction. Some authors propose that this reaction does not require that participation of the enzyme, and therefore, it could take place in the cytoplasm of the cell (water environment). Other authors suggest that the reaction needs to continue inside of the enzyme since the cyclization process is a rather complex task and the enzyme will be very important to catalyze this process [308, 312]. Therefore, both hypotheses were tested in this project and they will be discussed in the following sections.

Enzymatic Catalysis

Under enzymatic catalysis, the formation of 5,10-methylene-THF requires five steps that include: i) three proton transfers, ii) the formation and release of one water molecule, and iii) the cyclization process. In most of these reactions, Glu[57] plays an active role, and it is directly involved in the catalysis. The quinonoid intermediate is neither involved in any of these reactions nor does it affect the cyclization process. However, it was kept in all of the models used to study these reactions, because it stays in the quinonoid intermediate stage waiting for the final reaction where it is protonated, and the final product (glycine) is released.

Proton transfer reactions

The first two steps consist of three proton transfers. The first step involves the intramolecular proton transfer from N1 to N4 nitrogen atoms, resulting in the migration of the positive charge that was concentrated at N1 nitrogen, in the reactant, to N4 nitrogen. This reaction has a very small activation Gibbs free energy ($\Delta G^{\ddagger}$ = 0.7 kcal/mol), and it is exergonic in 8.5 kcal/mol.

The low activation energy results from the configuration of the TS structure (1230.6i cm^{-1}) in which both nitrogen atoms are very close to each other favoring the proton transfer (N1-N4: 2.69 Å). This proximity is endorsed by Glu^{57} that establishes two hydrogen bonds with the THF intermediate: one with the hydroxyl group (1.47 Å) and another with the proton that is bonded to the N4 nitrogen (2.57 Å). The stabilization of the reaction product relies on the better delocalization of the positive charge (mesomeric effect) that is provided by the benzyl group that is attached to N4 nitrogen (**Figure 31** Step 3).

The second step involves two coupled proton transfers: one between N4 and Oγ atoms and another between Oγ oxygen and Glu^{57}. The full process is almost spontaneous ($\Delta G^{\ddagger}$ = 1.3 kcal/mol), and the reaction is very exergonic in 8.3 kcal/mol (**Figure 31** Step 4). The low activation energy of this reaction is only possible due to the proximity of the atoms involved in the chemical reaction at the TS structure (639.6i cm^{-1}). The higher stabilization of the reaction product (THF-III) is due to the absence of point charges in the active site, which contrasts with the reactants where N4 nitrogen of the THF intermediate (THF-II) was positively charged, and Glu^{57} was negatively charged. This effect can be seen in the electrostatic maps shown in **Figure 31**. The key role played by Glu^{57} in these two steps reinforces, once again, its importance in the catalytic process as well as the importance of its protonation state at the beginning of the catalytic process.

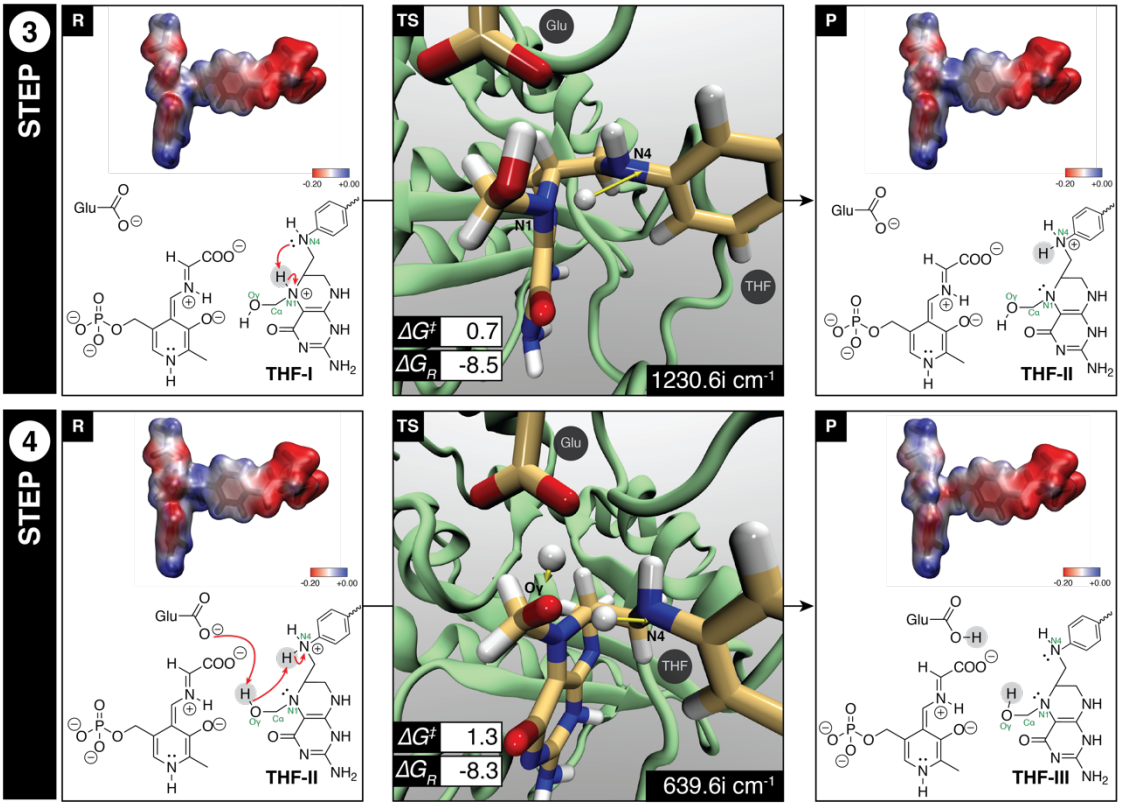


Figure 31. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the three proton transfers catalyzed by SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction.

Dehydration reaction

The next step of the cyclization process involves the formation of a carbocation through the elimination of a water molecule.

Our initial attempts to study this reaction involved the direct protonation of the hydroxyl group by Glu[57], leading to the formation of a water molecule and subsequent release (Figure 31 Step 4). Although this step is rather straightforward, the computational results indicate that this reaction is only possible when the hydroxyl group does not interact with the N4 nitrogen. Only in this conformation, it was possible to characterize a TS structure for the dehydration process. This new conformation is possible through the rotation of the N1-Cβ bond from the product of the previous step by -125.0º and has an energy cost of 9.6 kcal/mol (Figure 32).

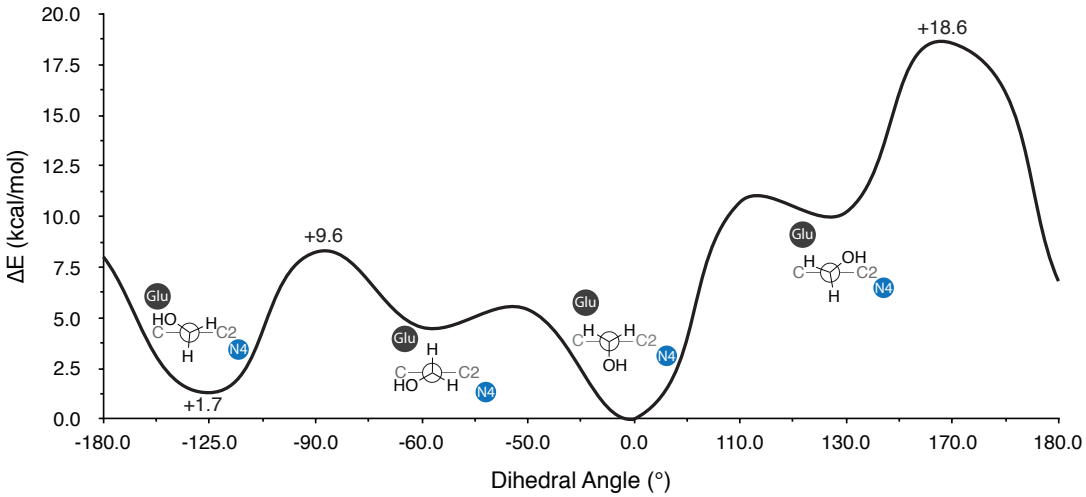


Figure 32. Energy profile for the dihedral rotation across the N1-Cβ bond. Newman representation was used to show the relative configuration of the Cβ and N1 atoms. The energies were computed using the ONIOM(DLPNO-CCSD(T)/CBS:ff99SB) scheme.

In the new conformation of the THF intermediate (THF-III), the elimination of the water molecule requires an activation Gibbs free energy of 5.4 kcal/mol, and the reaction is slightly endergonic in 3.0 kcal/mol.

This fastest reaction results from the good alignment between Glu^{57} and the hydroxyl group of the THF intermediate in the TS structure (113.6i cm^{-1}) (Figure 33). Although, in this new conformation, Glu^{57} does not interact directly with N4 nitrogen, it establishes a hydrogen bond with a water molecule (1.83 Å) that interacts with that atom (N4) through an additional hydrogen bond (2.00 Å).

At the end of this step, the release of the water molecule restores the negative charge of Glu^{57} and generates a positive charge at the N1 nitrogen that becomes bonded to carbon Cβ by a double bond (N1-Cβ: 1.43 Å in the reactant versus 1.32 Å in the product). The formation of the new point charges in the THF intermediate (THF-IV) is very important for the polarization of the Cβ-N1 bond. This bond polarization will be crucial to increase the electrophilic character of the Cβ atom, required for the cyclization of the THF intermediate (THF-IV) in the consecutive step.

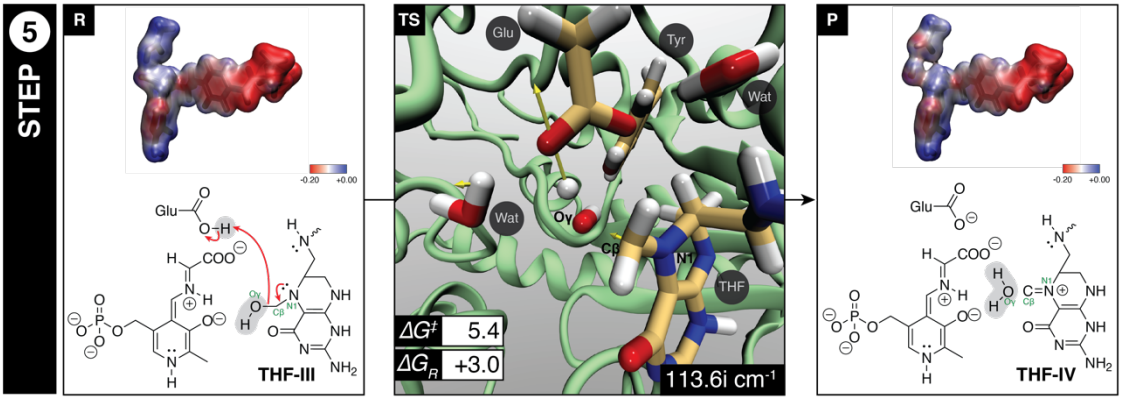


Figure 33. Structures of reactant (R), transition state (TS), and product (P) of the dehydration of the THF intermediate catalyzed by the SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped at the surface of the molecules involved in the reaction.

Cyclization

The cyclization of the THF intermediate (THF-IV) to yield the 5,10-methylene-THF molecule requires two sequential steps: i) an intramolecular nucleophilic attack and ii) a deprotonation.

The first step involves the nucleophilic attack of N4 nitrogen to Cβ carbon of the THF intermediate (THF-IV), from which results a 5-membered ring. This reaction is favored by the correct orientation of the Cβ atom in relation to the N4 nitrogen. The favored orientation is facilitated by a hydrogen bond that Glu^{57} residue makes with one water molecule present in the active site and the latter one with the N4 nitrogen. For this reason, the TS structure (205.0i cm^{-1}) is very close to the reactant, from which results a spontaneous reaction (Figure 34 Step 6). The positive charge that was lodged at the N1 nitrogen of THF-IV intermediate migrates to the N4 atom, where it is better delocalized due to the presence of the bonded benzyl group. The reaction Gibbs free energy of this reaction reflects this stabilization being exergonic in 16.9 kcal/mol.

The final step of the cyclization process involves the proton transfer from the N4 nitrogen of the THF-V to the Glu^{57} residue. This reaction proceeds with the participation of a water molecule that catalyzes the proton transfer between both groups. The TS of this reaction is characterized by one imaginary frequency at 670.7i cm^{-1}. At the end of the reaction, the Glu^{57} becomes protonated and one of the final products of the enzymatic catalysis, 5,10-methylene-THF, is obtained.

This step is almost spontaneous presenting an activation Gibbs free energy of 1.1 kcal/mol. In addition, this final step is slightly endergonic in 0.5 kcal/mol (Figure 34 Step 7).

Overall, it can be concluded that the cyclization process requires an activation Gibbs free energy of 1.1 kcal/mol and it is exergonic in 16.4 kcal/mol.

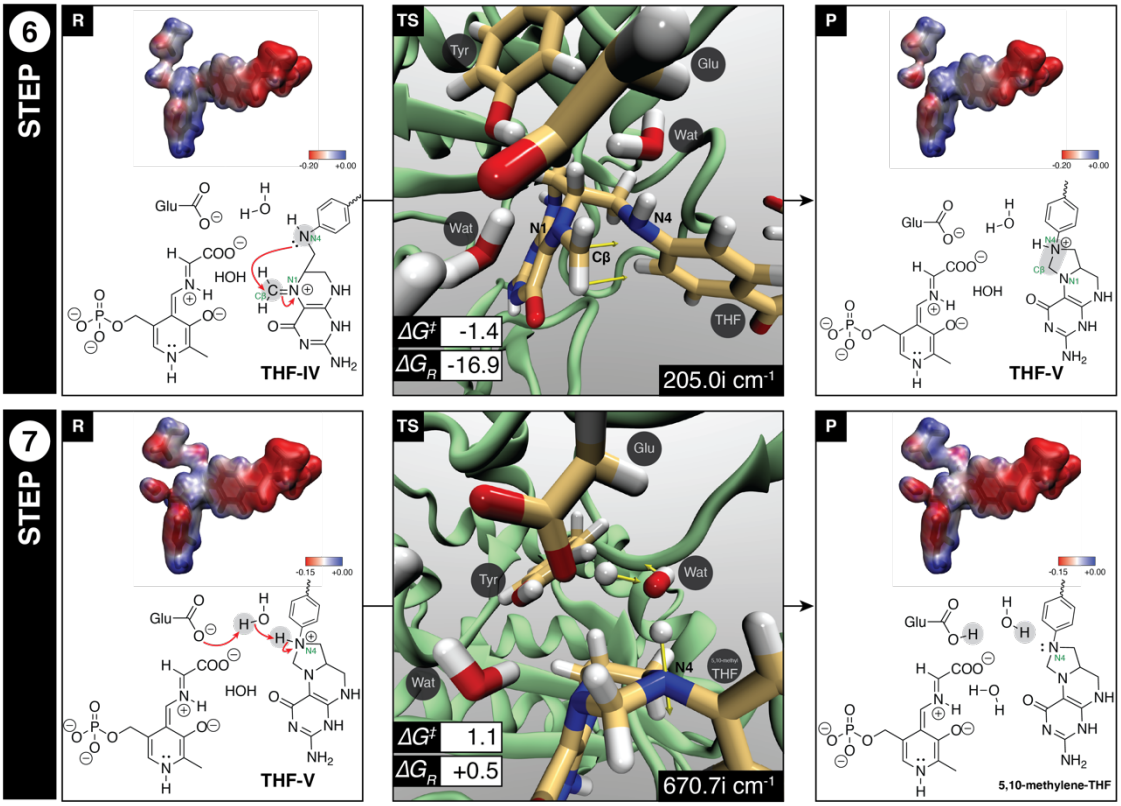


Figure 34. Structures of reactant (R), transition state (TS), and product (P) of the steps involved in cyclization of the THF intermediate catalyzed by the SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows corresponds to the major vibrational frequency associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction.

Non-enzymatic catalysis

In the previous section, it was described the synthesis of the 5,10-methylene-THF molecule under enzymatic catalysis. Now, it will be described the energy profile of the same reaction when occurred in the cytoplasm, which was considered as an aqueous environment.

Under enzymatic catalysis, the full process requires five sequential steps, in which the Glu[57] plays an active role, involving the conformational rearrangement (the rotation of the N1-Cβ bond) of the THF intermediate (THF-I(S)). In the cytoplasm, the cyclization

process requires only three steps and does not need any conformational rearrangement of THF.

The first step of the reaction, in the cytoplasm, is very similar to the one that takes place during the enzyme catalysis (**Figure 35** Step 1 and **Figure 31** Step 3). It involves the proton transfer from the N1 to the N4 atom of the THF intermediate (THF-I(S) versus THF-I).

In the cytoplasm, the reaction is marginally slower requiring as activation Gibbs free energy of 1.4 kcal/mol (versus 0.7 kcal/mol in the enzymatic catalysis). The most significant difference is observed in the reaction free energy. In the cytoplasm, this step is endergonic in 0.7 kcal/mol whereas under enzymatic catalysis it is exergonic in 8.5 kcal/mol. This difference comes from the stabilization of the product of the reaction that is provided by the enzyme, and mainly by Glu[57], which in the cytoplasm is absent.

The next step involves the dehydration process. This is where the differences between both mechanisms are observed. In the cytoplasm, the abstraction of the proton bonded to the N4 nitrogen, by the Oγ atom, occurs in a single step without requiring any conformational rearrangement of the reaction intermediate. At the end of this step, one water molecule is formed, and the same reactional product is obtained. Despite the simplicity of the reaction, the energetic cost for this reaction in the cytoplasm is much higher than under enzymatic catalysis. In the cytoplasm, this step requires an activation Gibbs free energy of 10.2 kcal/mol (versus 5.4 kcal/mol), which means that the dehydration process is approximately 3,000 times slower than when it happens with enzymatic catalysis. As expected, the reaction Gibbs free energy is almost the same for the dehydration whether the reaction occurs under enzymatic catalysis (ΔG_R = +3.0 kcal/mol) or in the cytoplasm (ΔG_R = +1.9 kcal/mol) (**Figure 35** Step 2 and **Figure 33** Step 5).

The last step required for the cyclization of the THF intermediate (THF-III(S) versus THF-IV) involves the nucleophilic attack of the N4 nitrogen to the Cβ carbon. This reaction is spontaneous in both the cytoplasm and the enzyme ($\Delta G^{\ddagger}$ = -3.5 kcal/mol versus $\Delta G^{\ddagger}$ = -1.4 kcal/mol), which means that both processes are energetically similar (**Figure 35** Step 3 versus **Figure 34** Step 6). However, it requires only one step instead of two. This energy difference is explained by the higher freedom of the structure in the cytoplasm that allows an easier cyclization of the 5-membered ring of THF without the geometric constraints imposed by the active site of the enzyme. However, once more, the enzyme is able to provide further stabilization of the reaction product (ΔG_R = -16.9 kcal/mol versus ΔG_R = -6.8 kcal/mol), contributing to a more favorable reaction.

All the TS structures of these reactions were checked using the vibrational frequency analysis and are characterized by a single imaginary frequency for each TS. Those frequencies correspond to the vibration of the atoms directly involved in each step and can be found in Figure 35.

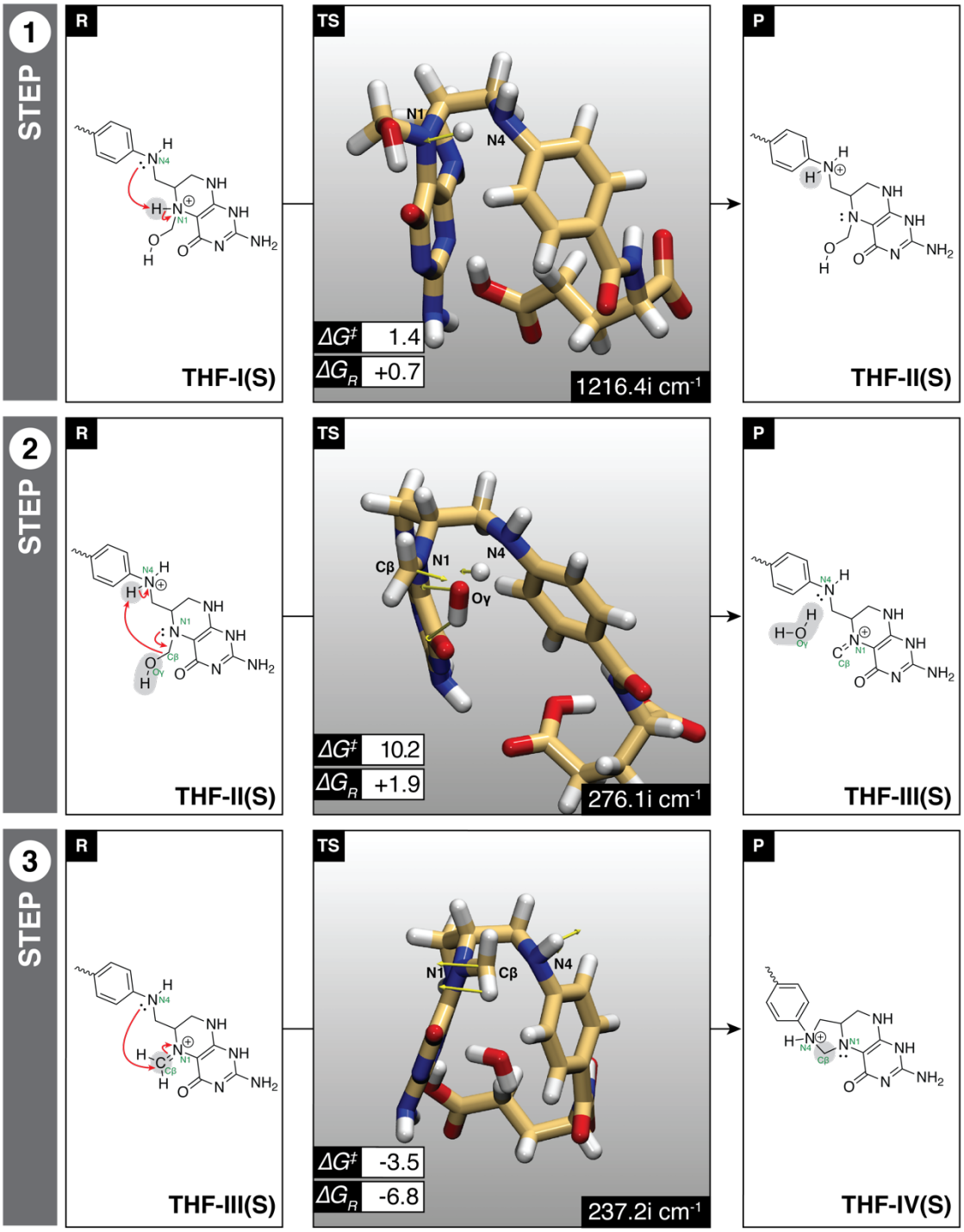


Figure 35. Structures of reactant (R), transition state (TS), and product (P) of the steps involved in cyclization of the THF intermediate in the cytoplasm. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS.

Looking at the overall results for the cyclization of the THF intermediate in the cytoplasm, it can be concluded that although the reaction in water requires fewer steps (three instead of four steps), the reaction is, from the kinetic and thermodynamic point of view, favored with enzymatic catalysis. Under enzymatic catalysis, the rate-limiting step of the cyclization process is the dihedral rotation across the N1-Cβ bond and requires 9.6 kcal/mol, whereas, in the cytoplasm, the rate-limiting step is the elimination of the water molecule with an activation barrier of 10.2 kcal/mol (Figure 36). When the reaction takes place in water, the process is almost thermoneutral (-4.2 kcal/mol), whereas with enzymatic catalysis of the same process originates a product that is highly stabilized in 28.5 kcal/mol (Figure 36)

In sum, these results show that the production of the 5,10-methylene-THF can occur in the cytoplasm and without enzymatic catalysis, but it would be much slower. However, taking into account that this process requires the migration of the THF intermediate to the cytoplasm, which implies an additional energetic cost (dissociation and solvation effects), the full process will be even slower. For this reason, the cyclization of THF intermediate (THF-I) has to occur through an enzymatic process catalyzed by the SHMT.

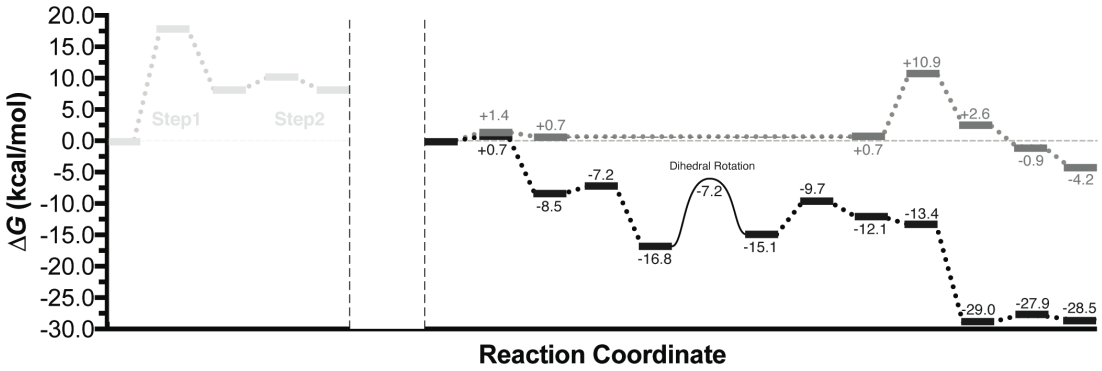


Figure 36. Energy profile of the two proposals for the conversion of the THF intermediate into 5,10-methylene-THF: (black) the process is catalyzed by the SHMT enzyme or (gray) the reactions occur in the cytoplasm. The values in brackets correspond to the energy of each single step.

c. Third Stage: Protonation of the Quinonoid Intermediate

Once the 5,10-methylene-THF is formed, the active site has Glu^{57} protonated and a quinonoid intermediate (Figure 34 Step 7) that needs to be converted into an external aldimine in order to enhance the enzymatic turnover and the release of one molecule of glycine.

It is worth mentioning that nothing is known if the 5,10-methylene-THF is released from the SHMT active site to the cytoplasm before or after the protonation of the quinonoid intermediate. For this reason, we chose to maintain the 5,10-methylene-THF inside the active site during this step. We observed that it is neither involved in any chemical step nor is it involved in the catalytic process.

This step involves a proton transfer from the Glu^{57} residue to the Cα carbon of the quinonoid intermediate, through a water molecule that mediates the reaction. This water molecule is the one that was generated at step 5 (Figure 33), which plays, in this final step, an important role for the continuity of the catalytic process. This reaction has an activation Gibbs free energy of 9.0 kcal/mol, and it is exergonic in 6.1 kcal/mol. The TS was validated through the vibrational frequency analysis showing a single imaginary frequency assigned to the atoms that are intrinsically related to the reaction (775.0i cm^{-1}). This reaction is favored because the quinonoid intermediate is negatively charged and therefore it meets good conditions to act as a nucleophile and become protonated (Figure 37).

At the end of this reaction, the Glu^{57} becomes negatively charged, restoring its initial protonation state, and an external aldimine with a glycine bonded to the PLP cofactor is obtained. After the transimination reaction, a process that is common to all PLP-dependent enzymes, the glycine molecule is released from the active site. At the same time, the PLP cofactor becomes bonded to the Lys^{229}, forming an internal aldimine, which makes the enzyme ready for a new catalytic cycle, once a new molecule of substrate becomes available.

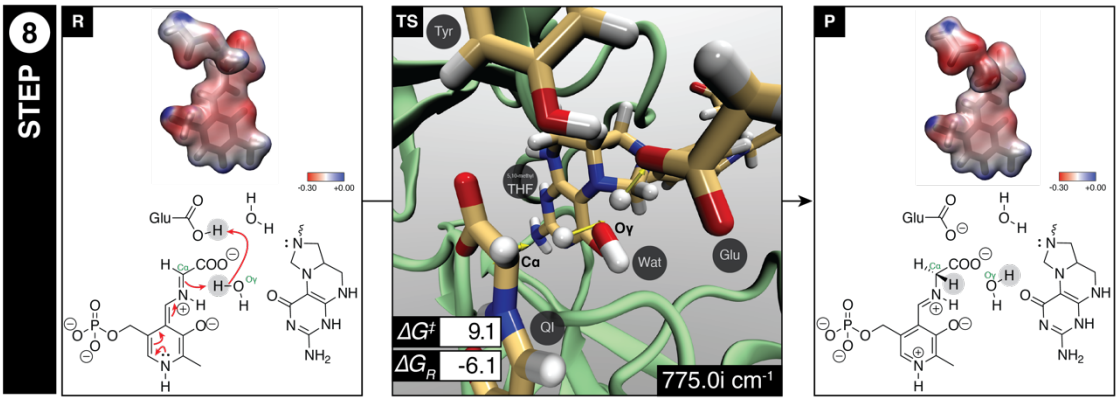


Figure 37. Structures of reactant (R), transition state (TS), and product (P) of the protonation of the quinonoid intermediate (QI) catalyzed by the SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The yellow arrows correspond to the major vibrational frequencies associated with the imaginary frequency of the TS. The electrostatic potential was calculated and mapped on the surface of the molecules involved in the reaction.

1.2.4. Conclusion

In this project, the catalytic mechanism of the SHMT was assessed using ONIOM QM/MM methodologies. The SHMT catalyzes the reversible conversion of L-serine and THF into glycine and 5,10-methylene-THF, with the contribution of the PLP cofactor. The reaction requires eight sequential steps that can be divided into three main stages: the α-elimination, the cyclization of the 5-hydroxymethyl-THF intermediate into 5,10-methylene-THF, and the protonation of the quinonoid intermediate.

The full mechanism is represented in Figure 38.

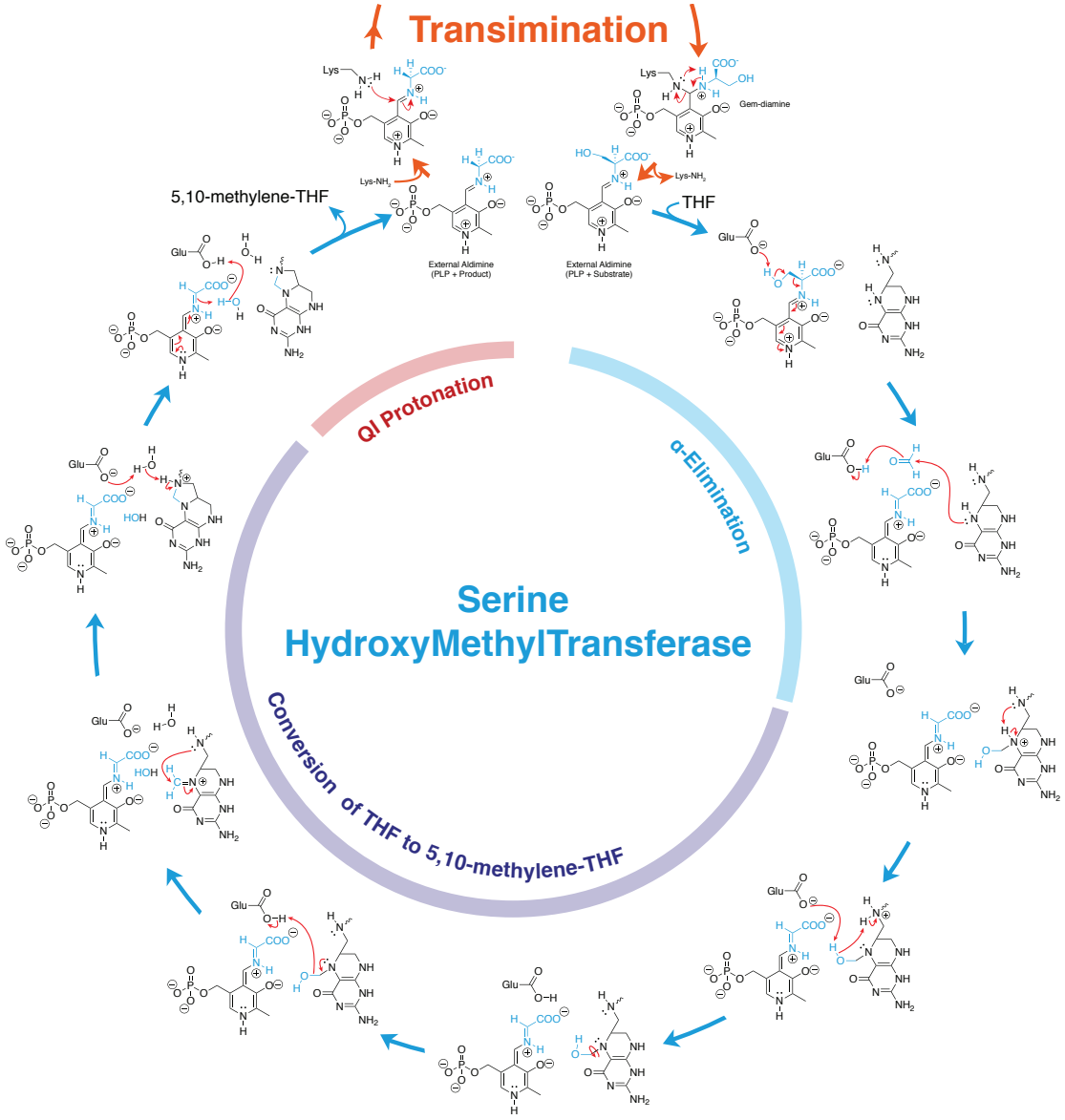


Figure 38. Overview of the catalytic mechanism of the PLP-dependent enzyme SHMT.

In the first step, the hydroxymethyl group of the substrate is eliminated in the form of a formaldehyde molecule that undergoes, in the second step, a nucleophilic attack by the

N1 atom of the THF cofactor. The resulting THF-I intermediate undergoes an intramolecular reaction, that involves a proton transfer from the N1 to the N4 nitrogen (step 3) and finally to Glu^{57} (step 4). Afterwards, a conformational rearrangement of the THF-III intermediate takes place, required for the elimination of a water molecule that enhances the electrophilic nature of the Cβ carbon, which in turn is required for the step 5. Subsequently, the Cβ atom becomes suitable to undergo a nucleophilic attack by the N4 nitrogen, producing a 5-membered ring THF intermediate (step 6). Immediately after, the Glu^{57} residue abstracts the remaining proton from the N4 nitrogen and produces the 5,10-methylene-THF (product) (step 7). Finally, the water molecule that was released in step 5 assists, in the final step, the proton transfer from the Glu^{57} to the Cα carbon of the quinonoid intermediate. Consequently, the external aldimine composed of a glycine molecule bonded to the PLP cofactor is formed (step 8). At the end of this reaction, the release of 5,10-methylene-THF and glycine (after the transimination reaction) from the active site takes place, and the enzyme is ready for a new enzymatic cycle.

According to the calculated energy profile (**Figure 39**), the rate-limiting step of the entire mechanism is where the elimination of the hydroxymethyl group occurs, forming a formaldehyde intermediate that becomes immediately bonded to the THF cofactor. The intramolecular passage of the proton between N1 and N4 atoms of the THF-I intermediate also contributes to assign this step as the rate-limiting one. These two reactions require an activation energy boost of 18.0 kcal/mol. These results closely fit the experimental kinetic data that predicted an activation barrier of 15.7-16.2 kcal/mol [306, 307, 321].

It is worth mentioning that SHMT is also capable of catalyzing the reverse reaction, in other words, the conversion of glycine and 5,10-methylene-THF into L-serine and THF, respectively. Analyzing the energy profile in the reverse order (**Figure 39**), all the steps are feasible under biological conditions, agreeing with the experimental evidence of reversible catalysis. However, there is no kinetic result about the kinetics of the reverse reaction catalyzed by the SHMT from *E.coli*. There is only a kinetic study using *Hydrogenobacter thermophilus* from which it was concluded that the reaction is almost 4 times slower in the reverse direction [322]. Based on this evidence, it might be expected a slower conversion of glycine into L-serine. This also agrees very well with the mechanism proposed in this work, since the rate-limiting step, for the reverse reaction, would be 29.7 kcal/mol, which compared with the rate-limiting step of the direct catalytic would lead to a slower reaction.

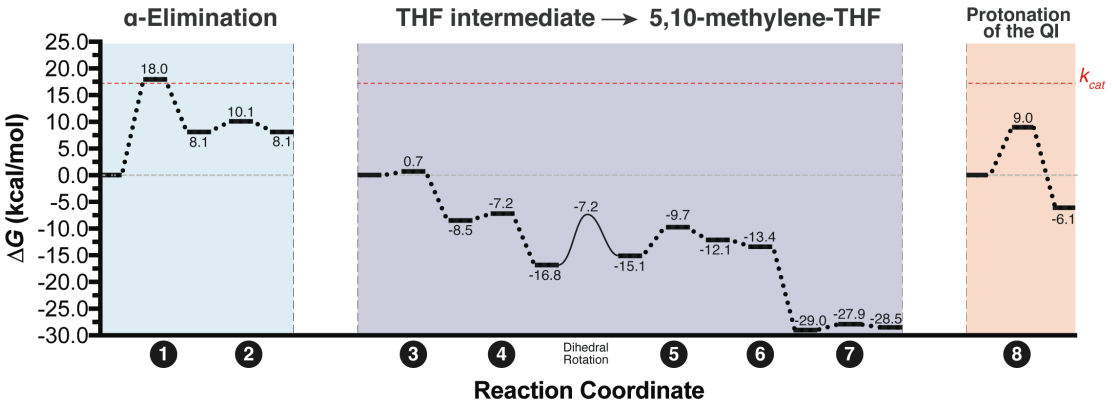


Figure 39. Complete energy profile (DLPNO-CCSD(T)/CBS::ff99SB) for the catalysis of L-serine and THF into glycine and 5,10-methylene-THF by the SHMT.

The results obtained with this work also provide new insights into the catalytic process and the production of the 5,10-methylene-THF molecule.

Regarding the two proposals that are available in the literature for the α-elimination reaction, the results presented in this paper indicate that the reaction proceeds with the formation of the formaldehyde molecule as an intermediate, and only then, it becomes covalently bonded to the N1 nitrogen of the THF cofactor. In this process, the Glu^{57} cannot be protonated since it plays a key role in holding the proton of the hydroxymethyl group. Although structural studies showed a close proximity between the Glu^{57} and the hydroxyl group of the substrate, from which was suggested that the Glu^{57} was protonated [309, 310], the calculations showed that this configuration of the active site does not allow the α-elimination. Such reaction was only possible to model when the anionic form of the Glu^{57} was considered. In the anionic state, the hydroxyl group from the substrate establishes a hydrogen bond with the Glu^{57}, agreeing with the experimental evidence that shows a close proximity between the Glu^{57} residue and the Oγ atom of the substrate.

The Glu^{57} was found to play a key role in several reactions of the mechanism, namely steps 1, 2, 4, 5, 7, and 8. This goes in line with the mutagenic studies that show that any mutations to Glu^{57} caused a drastic inhibition of the enzyme. [309, 311] This study proved, for the first time, the role played by the Glu^{57} residue. Furthermore, Glu^{57} is a key residue to handle most of the proton transfers that are required for the reaction to occur, holding a negative charge during some steps (step 3 and 6) to provide conditions for intermediate reactions to happen. Then, the Glu^{57} becomes protonated during the remaining steps of the mechanism. In sum, this glutamate residue works as an proton pump that handles the charge transfers across the mechanism and establishes a link between the external aldimine and the THF cofactor. (Figure 40)

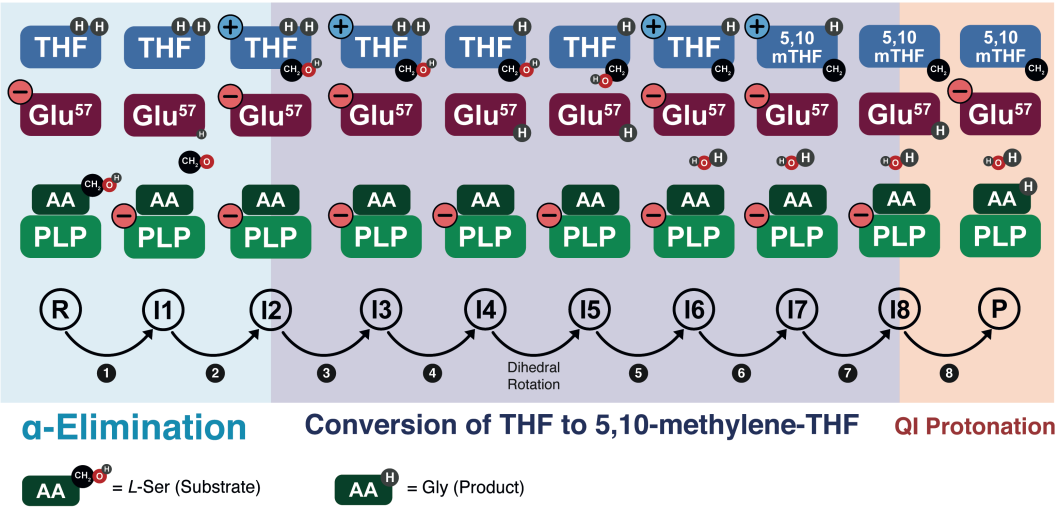


Figure 40. Schematic representation of the behavior of the main groups and charges across the entire catalytic mechanism.

In the literature, there are also two different proposals suggesting that the conversion of the THF intermediate to 5,10-methylene-THF can alternatively occur outside the enzyme through a non-enzymatic mechanism (in the solvent) [308, 312]. In this work, both proposals were studied, and the results showed that the reaction is thermodynamically and kinetically favored under enzymatic catalysis. The results presented in the manuscript are also in line with the observed experimental evidence [308], confirming that the conversion of the THF into 5,10-methylene-THF is very fast and almost spontaneous. Overall, this stage of the mechanism (step 3 to 7) is very exergonic and fast, agreeing with the experimental observation of an almost spontaneous process. Additionally, the steps involved in the cyclization of the THF are important to prepare the enzyme for the final step where a proton is transferred to the quinonoid intermediate to produce the glycine. This evidence agrees with the fact that the THF cyclization must occur inside the active site of the SHMT.

Moreover, our results agree with the proposal of Dunathan [295, 296], which postulates that the bond that is formed or cleaved during the reaction involving the PLP cofactor must be placed perpendicularly to the conjugated system of the PLP cofactor. The hydroxymethyl group is released almost perpendicularly to the PLP cofactor (Cβ-Cα-C7-C2 angle: 98.6 °), and the protonation of the quinonoid intermediate occurs in the same way (H-Cα-C7-C2 angle: 115.7 °).

Together with the previous stages that are common to all PLP-dependent enzymes, this work provides, for the first time and with an atomistic level of detail, a portrait of the entire mechanism catalyzed by the SHMT. This knowledge can be used to enhance the rational molecular design of new drugs against malaria.

This page was intentionally left blank.

1.3. α-Addition

In the previous work, the catalytic mechanism of SHMT was studied and completely described by the application of computational methods. In this work, the mechanism for a different reaction catalyzed by SHMT was studied following a similar approach.

Since it was already described the ability of SHMT to catalyze the stereoselective synthesis of a different α,α-dialkyl-α-amino acids, we studied the mechanism behind this reaction. Moreover, the mechanism of two additional mutants (Glu57Leu and Glu57Gln) were also investigated to produce a bioengineered enzyme for the catalysis of this kind of reaction.

This work is still ongoing. Therefore, this section will address the main results already available. Nonetheless, this project was presented in the following conferences:

- (Poster) *III Symposium on Medicinal Chemistry of University of Minho*, **2017**, Braga, Portugal (Appendix: page 381)
- (Oral Communication) *7th EuCheMS Chemistry Congress*, **2018**, Liverpool, United Kingdom
- (Poster) *19th FEBS Young Scientists' Forum*, **2019**, Krakow's Jagiellonian University Campus, Krakow, Poland (Appendix: page 385)
- (Poster) *The 44th FEBS Congress*, **2019**, ICE Krakow Congress Centre, Krakow, Poland (Appendix: page 385)
- (Poster) *7º Encontro De Jovens Investigadores Em Biologia Computacional Estrutural*, **2019**, Lisbon, Portugal (Appendix: page 386)

1.3.1. Introduction

The SHMT (EC 2.1.2.1) is a PLP-dependent enzyme that, under biological conditions, also requires the THF cofactor to catalyze the reversible conversion of L-serine into glycine and THF into 5,10-methylTHF. We mentioned "under biological conditions" because the SHMT can catalyze a retroaldol cleavage of 3-hydroxy-amino acids (e.g., L-*allo*-threonine and β-phenyl-serine) in a THF-independent mechanism [308]. This capacity of SHMT to catalyze the reversible retroaldol cleavage of 3-hydroxy-amino acids is being explored by scientists and engineers as a powerful biocatalyst for the stereoselective synthesis of a huge variety of α,α-dialkyl-α-amino acids [143].

α,α-disubstituted α-amino acids occupy a central role in biotechnological and biomedical chemical processes. Moreover, they are substructures of several biologically active molecules, such as drugs for the immune system (myriocin), infectious diseases (sphingofungines E and F), the neuronal system (lactacystin), and cardiovascular system (AADC inhibitors) (Figure 41) [143].

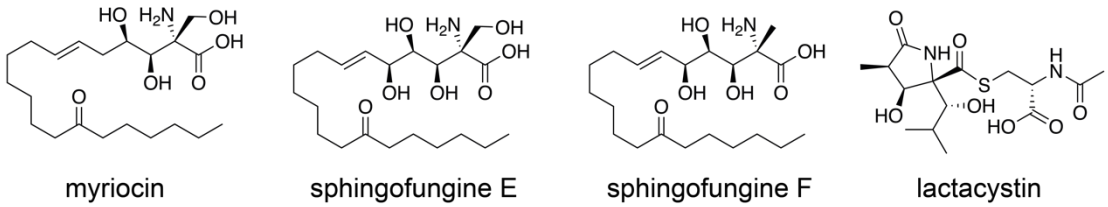


Figure 41. Structures of myriocin, sphingofungines E and F, and lactacystin.

However, the synthesis of chirally pure α,α-disubstituted α-amino acids is still a challenging procedure due to the difficulty of synthetizing stereogenic tetrasubstituted carbon centers with stereoselectivity control. Therefore, some enzymes, in particular SHMT, have been pointed out as powerful catalysts to produce these compounds with controlled and good stereoselectivity under mild aqueous conditions [323-326].

SHMT was already been shown to catalyze the aldol addition of glycine to aldehydes with interesting rates [327]. However, the enzyme is not efficient enough to catalyze this reaction. In order to improve the rates of the reaction and the stereoselectivity of the reaction, the enzyme has been subject to enzymatic engineering. A good starting point is to choose the orthologue enzyme with the best performance for the catalysis of these retroaldol reactions. Some kinetic data revealed that the best yields and reaction rates are achieved using the SHMT from *Streptococcus thermophiles* [309, 328]. Nevertheless, the details about the full mechanism are still unknown, even though M. Ishii and co-workers have already provided insights about the catalytic mechanism for

the aldol reaction catalyzed by SHMT [328]. However, the system was studied using a small QM model where only figures the EA (PLP + glycine) and part of the sidechain of an Arg residue that establishes a saline bridge with the carboxylate group of the substrate. The lack of all the surrounding residues does not provide insights into how the enzyme could be ingeniously modified to enhance its performance in the industrial context.

For this reason, the catalytic mechanism of the aldol addition of glycine to acetaldehyde was studied using an ONIOM QM/MM approach to get a full picture of all players involved in the reaction. Particular attention will be given to the role of catalytic Glu^{57} residue that is essential for the biological reaction [309, 311] but seems to impair the retroaldol reaction of L-*allo*-threonine that is the one with industrial relevance [309, 329].

Considering the biological reaction where L-serine is converted into glycine in the presence of THF, the mutation of the Glu^{57} residue by a Gln slows down the reaction 500 times, and the mutation by Leu completely impairs the reaction [309]. Interestingly, these two mutations improve the rates of the THF-independent retroaldol reaction. The retroaldol cleavage of the L-*allo*-threonine is 10% faster when the Glu^{57} residue is replaced by a leucine, and the reaction is even faster, about 430% when the Glu^{57} is mutated by a glutamine residue (**Table 1**).

Table 1. Experimental kinetic data available for the reactions catalyzed by SHMT.

	k_{cat} (s^{-1})		
	Rabbit cytosolic (30 °C) [309]		*S. thermophiles* (37°C) [329]
	L-Ser + THF	*L-allo*-Threonine	*L-allo*-Threonine
Wild-Type	10	2.17	4.32
Glu57Leu	not detectable	2.40	n.a.
Glu57Gln	0.020	9.33	n.a.

In this work, the catalytic mechanism of the aldol addition of glycine to acetaldehyde (the reverse reaction of the retroaldol cleavage) was studied using the crystallographic structure of the SHMT from *S. thermophiles* (PDB code: 4WXG [143]). This structure was used as a template for the wild-type form of the enzyme. Based on the kinetic data, the rate-limiting step of the entire mechanism should not be higher than 17.3 kcal/mol. The impact of the mutations (Glu57Leu and Glu57Gln) was also tested, studying the respective mechanisms of each mutated form of the enzyme.

This study was focused on the retroaldol mechanism since the previous activation mechanisms where the PLP becomes bonded to the apoenzyme [27], and the consequent substrate binding to the PLP cofactor [28] have been already studied.

Moreover, these two initial activation mechanisms are common to all PLP-dependent enzymes and occur faster than the specific reactions that occur afterwards. (Figure 42)

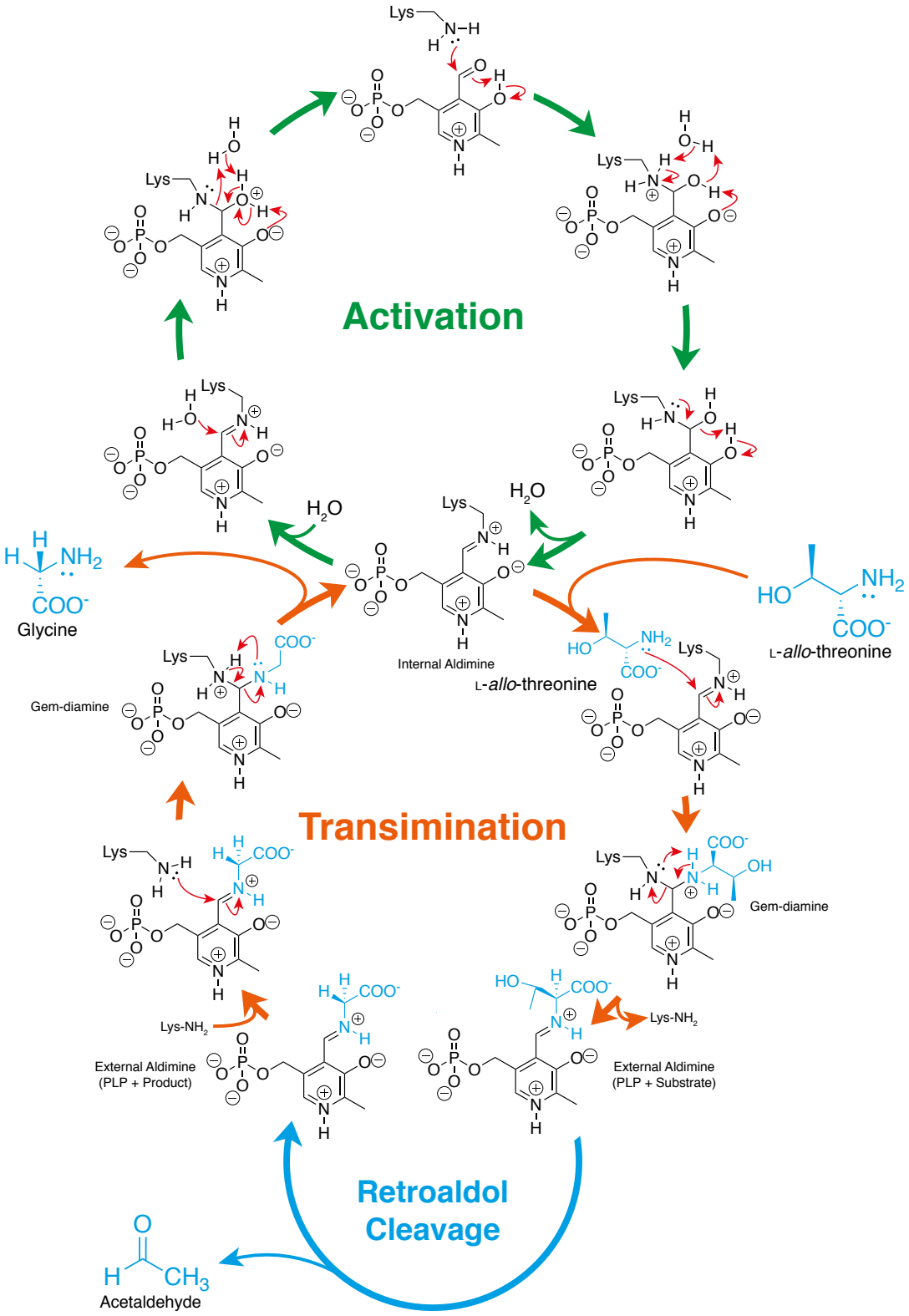


Figure 42. Schematic representation of the mechanism for the IA and EA formation for PLP-dependent enzymes and the specific retroaldol cleavage reaction.

1.3.2. Methodology

a. Model

The crystallographic structure (PDB code: 4WXG) [143] used in this study includes the homodimer structure of SHMT from *Streptococcus thermophilus*. This structure was obtained through the co-crystallization of SHMT with L-threonine. Due to the lack of information about hydrogens position on crystallographic structures, the stereochemistry of the β-carbon was taken into account during the protonation procedure to fit the L-*allo*-threonine conformation (*2S*,*3S*) instead of L-threonine (*2S*,*3R*). The structure lacks the initial six amino acid residues of each subunit, but those residues were not included in the model because they belong to a highly solvent-exposed loop and are far from the active site. Therefore, the model contains two equal subunits with 410 amino acid residues each, as well as two EAs (PLP + L-*allo*-threonine) placed inside each active site.

Two additional models were prepared to study the influence of the Glu^{57} mutation, using the wild-type structure and mutating the Glu^{57} by a Gln and a Leu residue. In the end, three initial models were built, and all of them undergo the same procedure as following described.

b. MD Simulations

The protein and the EA were solvated by using TIP3P [274] type water molecules. The water box was set up, allowing a minimum distance of 12 Å from protein to each face of the box. Hydrogen atoms were added to the model using SANDER from the AMBER12 package. [277] In this work, we considered the protonation states for all amino acids at pH 7.0 according to PROPKA 3.1 [275, 276].

The parameterization of the external aldimine was made through an optimization using the Hartree–Fock (HF/6-31G(d)) method to depict the atomic charges and the antechamber software from AMBER12 package [277] to assign the atom types.

Antechamber software from AMBER12 package was used to assign parameters to the EA. The atomic charges for these molecules were obtained from RESP of the atomic charges calculated using HF/6-31G(d).

A set for four minimizations, an equilibration (20 ps), and a 10 ns production were employed to each system. The production stage of the MD simulation was conducted in an isothermal-isobaric ensemble, using the ff99SB [149, 278, 279] force field, at the temperature of 310.15 K, pressure of 1.0 bar, integration step of 1 fs, and a 10.0 Å cutoff for non-bonded interactions.

From the equilibrated region of each MD simulation, the frames were grouped in five clusters according to their RMSd. Then, a representative structure from the most populated cluster was selected to proceed with the QM/MM calculations.

c. QM/MM Calculations

The QM/MM system was built loading the structure isolated in the previous procedure, using the molUP plugin [316] for VMD software [292]. The ONIOM QM/MM model included the full dimer, both EAs placed in each active site, and all water molecules that are not further than 5 Å of protein.

The HL layer of the wild type (WT), Glu51Gln, and Glu51Leu models include 93, 92, and 97 atoms, respectively. The HL layer includes the atoms of the EA, a water molecule, the sidechain of Glu/Gln/Leu^{57B}, Arg^{363A}, and Asp^{201A} residues, and the backbone of Ser^{97A}, Gly^{98A}, and Ser^{99A} residues (Figure 43). Each model's size increases by three atoms due to the inclusion of an additional water molecule in the last step.

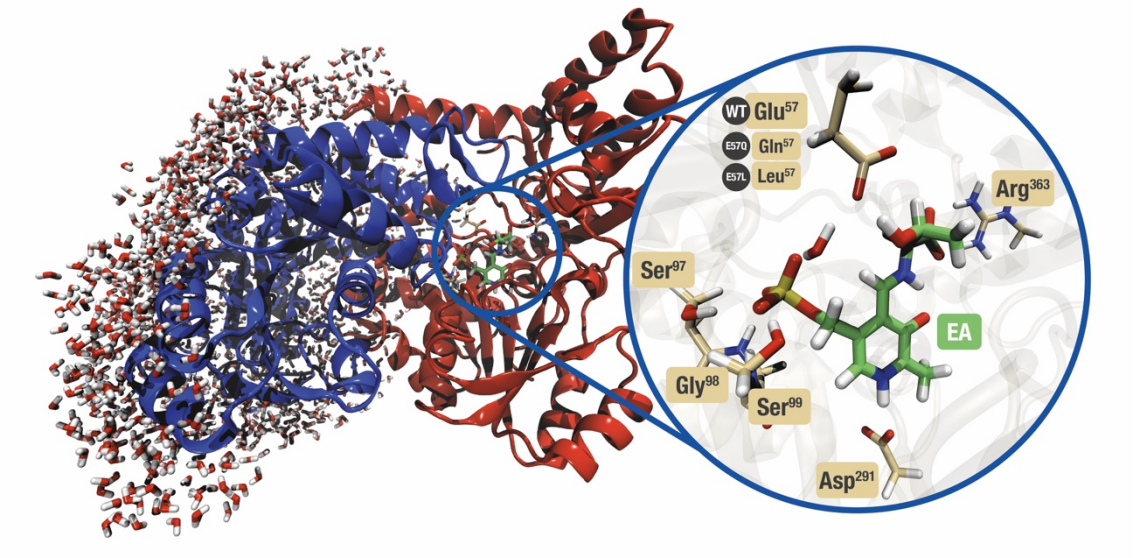


Figure 43. (Left) Secondary structure representation of the SHMT dimer (subunit A: red; subunit B: blue) and licorice representation of the HL layer. (Right) Licorice representation of the HL layer where the carbon atoms of the EA are colored green.

The remaining atoms of the system were included in the LL layer. Hydrogen atoms were used as link atoms to complete the valence of the bonds spanning between the two layers of the ONIOM QM/MM Scheme.

The interaction between HL and LL layers was considered according to the electrostatic embedding approach.

The geometry optimization of the HL layer was performed using the B3LYP [330] functional (DFT). The 6-31G(d) [182, 282-285] basis set was employed for the geometry optimizations as available in Gaussian 09 [280]. All TS and minima were validated by vibrational frequency calculations. The thermal corrections for estimation of the activation and reactions Gibbs free energies were calculated at 298.15 K and 1 atm.

The SP energy calculations are going to be performed to refine the computed energies, using the DLPNO-CCSD(T)/CBS scheme for the HL layer as applied in the study of the natural mechanism of SHMT (Chapter C – 1.2.2. page 144).

1.3.3. Results and Discussion

Since this is an ongoing research project, the following section provides a description and conclusions based on the currently available results. This work aims to clarify the catalytic mechanism behind the retroaldol cleavage of L-*allo*-threonine catalyzed by the wild-type SHMT (WT) and two mutated forms of this enzyme (E57L) and (E57Q). Currently, the study of the mechanism for the E57Q model is completed. With respect to the WT form, calculations devoted to the last step of the mechanism are being conducted. Finally, considering the E57L mutated form, the first step is still being studied under the same methodologic approach.

This section is divided into the two retroaldol cleavage mechanisms studied for the WT, E57Q forms of SHMT, whose results are already partially available.

a. Wildtype (WT)

The structure obtained from the MD simulation was optimized, and the resulting structure was used as the starting point for the relaxed scan of the α-cleavage. Accordingly, the Cα-Cβ bond (Figure 44 – Step 1) was incrementally increased.

The calculations unraveled the nucleophilic attack of the Glu^{57} residue to the Hδ proton of the substrate (O1-Hδ distance: R 1.49 Å; TS 1.04 Å; P 1.01 Å). This proton transfer triggers the formation of a double bond between Cβ and Oγ atoms (Cβ-Oγ distance: R 1.39 Å; TS 1.27 Å; P 1.23 Å) (Figure 44 – Step 1) that cleaves the Cα-Cβ bond (Cα-Cβ distance: R 1.59 Å; TS 2.17 Å; 2.91 Å). This reaction is possible due to the intrinsic features of PLP to stabilize the negative charge formed during this first step, involing the formation of the quinonoid intermediate. Consequently, these electrons are stabilized across the conjugated system of the PLP cofactor (Figure 44 – Step 1). All the minima were verified through vibrational frequencies calculations, and only the TS presented one imaginary frequency at 225.4i cm^{-1}, corresponding to the vibration of the atoms involved in the reaction (Figure 44 – Step 1: green arrows). The reaction requires an activation Gibbs free energy of 13.0 kcal/mol, and it is endergonic by 9.5 kcal/mol. At the end of step 1, the ethanal molecule is released, and the quinonoid intermediate is formed (Figure 44 – Step 1: P)

Before the second step takes place, a water molecule occupies the same region that was previously occupied by the ethanal molecule. This is essential for the protonation step to occur since the Glu^{57} residue cannot directly protonate the Cα atom. The direct protonation was tested but resulted in significant energy barrier (over 25 kcal/mol), from which no product was obtained. Since in this reaction, SHMT does not employ the THF

cofactor, the active site is plenty of water molecules that easily diffuse to the near region of the quinonoid intermediate.

The second step involves the protonation of the Cα atom before releasing glycine as the final product of the catalytic process (Figure 44 – Step 2). The pair of electrons that was previously stabilized by the PLP cofactor is delocalized, in step 2, to the Cα atom. Then, the Cα atom makes a nucleophilic attack on the HW proton of the water (Figure 44 – Step 2). Simultaneously, the Hδ is transferred from Glu^{57} residue to the water molecule, restoring the initial protonation state of the Glu^{57} residue. At this moment, the Glu^{57} residue is ready to catalyze a new retroaldol cleavage, after the release of glycine as the latest final product of the catalytic process. (Figure 44 – Step 2)

The TS for this second step is still being calculated. Therefore, both activation and reaction energies for this step are only predictions based on the relaxed scan calculation. According to those calculations, the activation energy is 9.3 kcal/mol, and the reaction is exergonic by 2.7 kcal/mol.

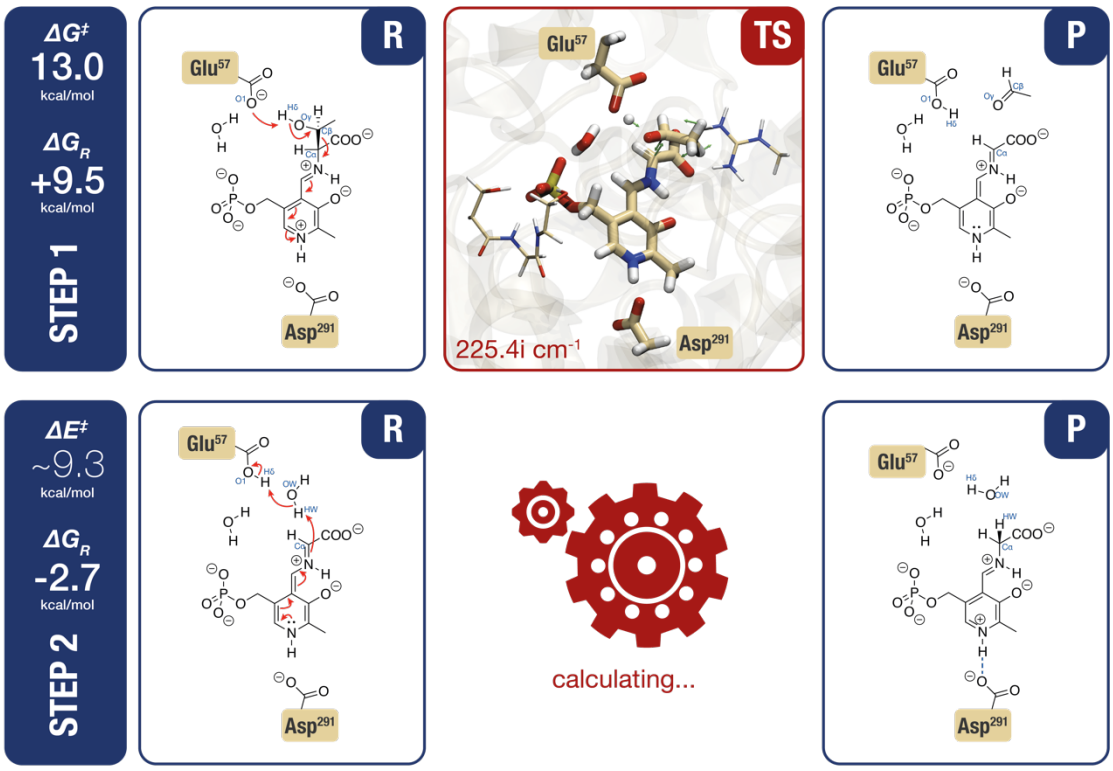


Figure 44. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the retroaldol cleavage catalyzed by the wildtype form of SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism.

b. Glu57Gln (E57Q)

The catalytic mechanism of the Glu57Gln mutated form of SHMT was conducted using the same approach used for the WT SHMT. In this case, the structure obtained from MD simulation shows a hydrogen bond between the hydroxyl group of the substrate and a conserved water molecule near the phosphate group of the PLP cofactor (**Figure 45** – Step 1: R). This fact immediately discloses a difference between both mechanisms, since Gln^{57} residue cannot accept the Hδ proton as Glu^{57} does in the WT. Alternatively, the proton transfer occurs between the hydroxyl group of the substrate and the phosphate group of the PLP factor, mediated by a water molecule.

As performed in the WT study, the Cα-Cβ bond was incrementally increased in a relaxed scan calculation (Cα-Cβ distance: R 1.59 Å; TS 2.16 Å; P 2.98 Å). During the α-elimination, the H proton is transferred to the water molecule (OW) (OW-Hδ distance: R 1.66 Å; TS 1.02 Å; P 0.99 Å). Simultaneously, the water molecule loses the HW proton to the phosphate group of PLP (OP1) (OP1-WH distance: R 1.57 Å; TS 1.04 Å; P 1.01 Å). This reaction is slower than the one catalyzed by the WT form since the activation Gibbs free energy is 17.1 kcal/mol (vs. 13.0 kcal/mol). Moreover, this first step is also slightly more endergonic with a reaction Gibbs free energy of 12.0 kcal/mol (vs. 9.5 kcal/mol). Both minima and TS were validated by vibrational frequency calculations, and only TS showed one imaginary frequency at 230.3i cm^{-1}. This frequency corresponds to the vibration of the atoms involved in the first step of the Glu57Gln mechanism as represented by the green vectors in **Figure 45** – Step 1: TS.

Similar to the WT mechanism, this mutated SHMT form requires the approximation of one water molecule to allow the protonation of the quinonoid intermediate (**Figure 45** – Step 1: P). Accordingly, a water molecular can easily diffuse and occupy the position left by released ethanal molecule at the end of step 1.

The pair of electrons lodged by the PLP cofactor after step 1 is now available to abstract a proton (HW1) from the water molecule (Cα-HW1 distance: R 2.24 Å; TS 1.32 Å; P 1.09 Å) (**Figure 45** – Step 2: R). This reaction automatically triggers a simultaneous cascade of proton transfers. The Hδ proton is transferred to the water molecule that protonates the Cα atom (Hδ-OW1 distance: R 1.78 Å; TS 1.44 Å; P 0.98 Å) and the HW proton is recovered from the phosphate group of PLP (OP1-HW distance: R 1.61 Å; TS 1.42 Å; P 1.00 Å). This reaction restores the initial protonation state of the PLP phosphate group and produces the external aldimine with the final product (glycine). (**Figure 45** – Step 2: P)

The second step of this mechanism has an activation Gibbs free energy of 9.2 kcal/mol, which is very similar to the one predicted for the WT SHMT (9.3 kcal/mol). However, the reaction is considerably more exergonic than the WT mechanism, showing a reaction Gibbs free energy of -10.8 kcal/mol (vs. -2.7 kcal/mol). This evidence provides important insight into how more stable is the negative charge of the phosphate group rather than in the Glu^{57} residue. Reactant, TS, and product of step 2 were validated by vibrational frequencies calculations, where only TS presented one imaginary frequency at 1033.5i cm^{-1}. This vibrational mode is associated with the vibration of the atoms involved in the proton transfer from the phosphate group to the Cα atom of the quinonoid intermediate (Figure 45 – Step 2: green arrows).

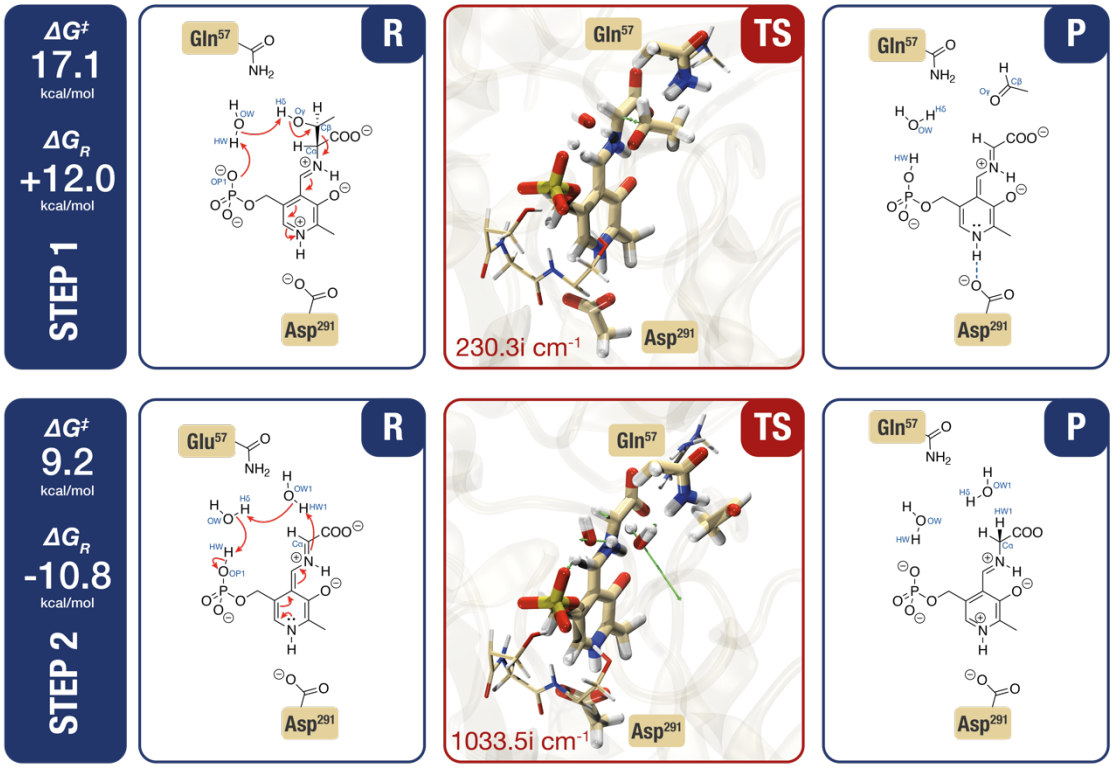


Figure 45. Structures of reactant (R), transition state (TS), and product (P) of the two sequential steps associated with the retroaldol cleavage catalyzed by the Glu57Gln form of SHMT. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism.

c. *Energy Profile*

The energy profiles for the retroaldol cleavage catalyzed by the WT and E57Q forms of SHMT are shown in Figure 46 (Left). From the kinetic point of view, the results show that the WT form is much more efficient in catalyzing the elimination of ethanal from L-*allo*-threonine than the mutated form, where Glu^{57} was replaced by a Gln. This difference is more evident in the first step, where the Glu57Gln mutation causes an increase of 4.1 kcal/mol in activation energy. The resulting product is also less stabilized in the mutated form when compared with the WT SHMT. The nucleophilic character of Glu^{57} residue is greater than the phosphate group of PLP. Therefore, the reaction occurs faster when the Hδ proton is abstracted by the Glu^{57} residue.

However, the Glu57Gln mutation turns the overall process almost thermoneutral (+1.2 kcal/mol), whereas the WT SHMT originates an endothermic process (+6.8 kcal/mol). This difference can be explained by the fact that the negative charge of the PLP phosphate group is more stabilized by the surrounding residues than Glu^{57}.

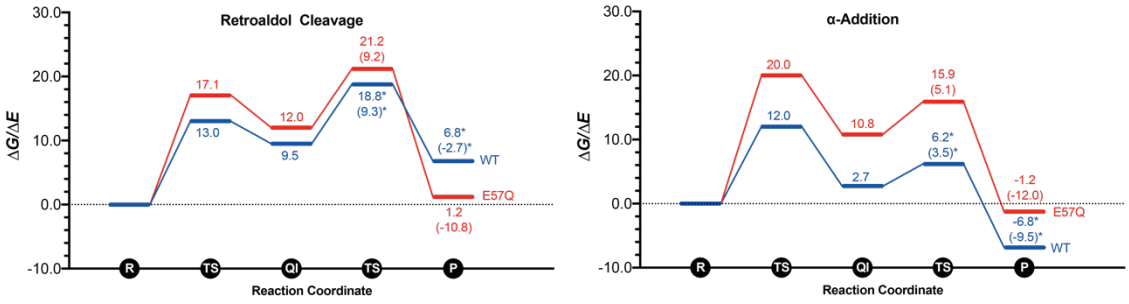


Figure 46. Energy profile of the (Left) retroaldol cleavage and (Right) α-addition catalyzed by the WT (blue) and E57Q (red) mutated form of SHMT. The energy profile of the α-addition is the profile of the reverse reaction presented in the retroaldol cleavage profile. * Energies estimated by relaxed scan calculations.

In this work, the retroaldol cleavage mechanism was studied, but the interesting industrial reaction is the exact reverse process, where glycine is deprotonated and an aldehyde, as ethanal, is added to the Cα atoms in a stereospecific manner. In order to facilitate the analysis of reverse reactions, the mechanism of the α-addition was plotted in Figure 46 (Right).

According to the results obtained in this work, the Glu57Gln mutation does not favor the reaction from either the thermodynamic or kinetic points of view. Consequently, the WT form of SHMT seems to provide better catalytic efficiency in the addition of ethanal to glycine than this particular mutated form.

1.3.4. Conclusion

This still ongoing study of the catalytic mechanism of retroaldol cleavage of L-*allo*-threonine on the WT, E57Q, and E57L forms of SHMT has already provided some insights about the catalytic mechanisms and the impact of the mutations.

So far, the WT and E57Q forms have different chemical species involved in the nucleophilic attack during step 1. The WT enzyme catalyzes the Hδ proton abstraction through the Glu^{57} residue. On the other hand, when Glu^{57} residue is mutated by a Gln, the reaction proceeds through a water molecule, and the final proton acceptor is the PLP phosphate group. However, this different mechanism is not advantageous since the reaction becomes slower than the one catalyzed by the WT form.

Although no final data about the Glu57Leu mutation is available, our initial results suggest a mechanism similar to the one obtained for the Glu57Gln. Therefore, it is not expected a better performance, in terms of kinetics, of Glu57Leu form of SHMT in the α-addition of aldehydes to glycine. Since this residue appears in the interface of the two subunits, its mutation could have impact on the stability of the enzyme. Therefore, further investigation will be conducted to assess that effect. Nonetheless, other residues can be mutated to measure its impact in the catalysis of this reaction, looking for a better engineered version of SHMT.

This page was intentionally left blank.

1.4. γ-Elimination

The catalytic mechanism for the γ-elimination catalyzed by PLP-dependent enzymes is being studied using the cystathionine-γ-lyase (CGL) enzyme as example.

All the calculations are being performed by Henrique S. Fernandes. Since this is an ongoing project, the following section of this thesis provides an introduction about the CGL, the computational methodology employed in the study, and the most recent results.

1.4.1. Introduction

Cystathionine-γ-lyase (CGL; EC 4.4.1.1) is PLP-dependent enzyme that catalyzes the cystathionine conversion into cysteine, α-ketobutyrate, and ammonia (Figure 47) [331, 332]. This is the main reaction catalyzed by CGL, and it is essential for the maintenance of the cysteine cell pool (Figure 7). Additionally, CGL can also catalyze three additional elimination reactions: the conversion of homoserine into water, ammonia, and α-ketobutyrate; the conversion of cystine into thiocysteine, pyruvate, and ammonia [333]; and the conversion of cysteine into pyruvate, ammonia, and hydrogen sulfide [331] (Figure 47).

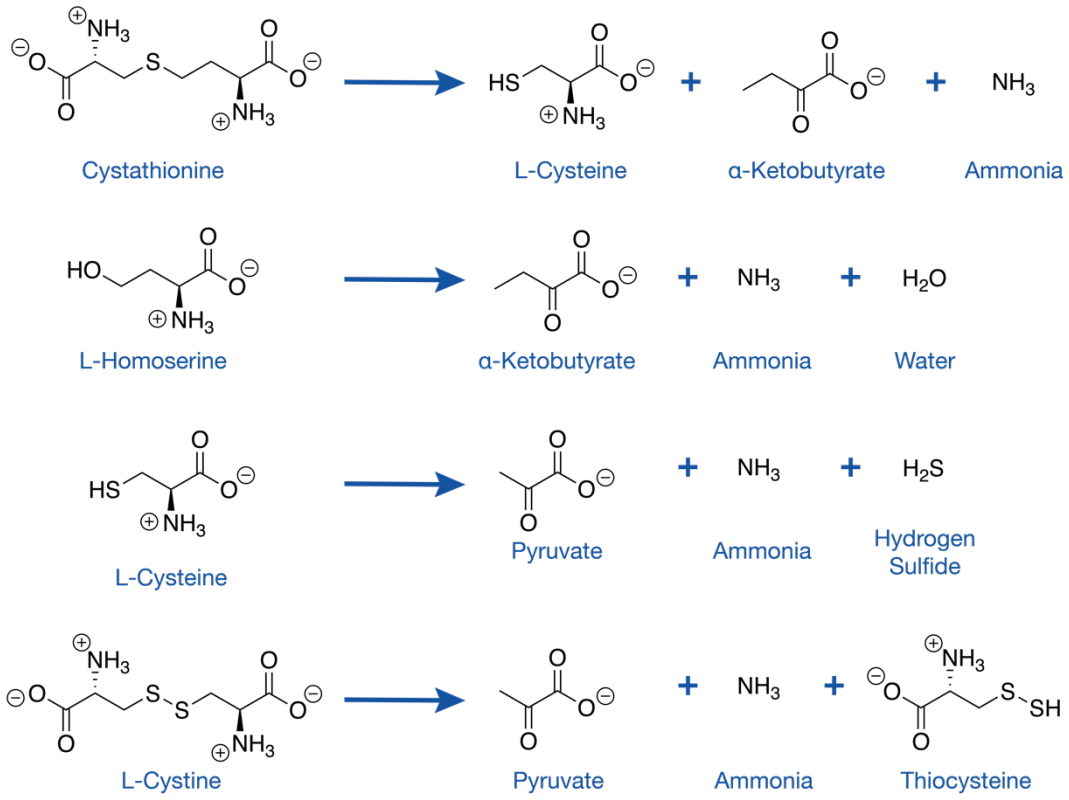


Figure 47. Reactions catalyzed by CGL.

Hydrogen sulfide is a very important biological signaling molecule involved in vascular and nervous functions [331, 334, 335]. In humans, the main source of hydrogen sulfide is the desulfhydration of cysteine catalyzed by CGL and cystathionine β-synthase (CBS), another PLP-dependent enzyme. Moreover, there is evidence that CBS is responsible for producing hydrogen sulfide in the central nervous system, while CGL is the enzyme that catalyzes its production in peripheral tissues [336-340]. The primary importance of CGL in the homeostasis of cysteine and the production of hydrogen sulfide turns this

enzyme a key participant in several biological functions, namely the cell growth regulation, the blood pressure control [341, 342], the macrophages activity [343], gluconeogenesis [344], and the kidney protection in situations of ischemia/reperfusion [345].

This central role played by CGL has been studied to address new therapeutic strategies against several diseases. The increased activity of CGL showed good results in preventing nonalcoholic fatty acid disease [346], atherosclerosis [347-349], and helping in the delayed gastric emptying in diabetic patients [350]. However, other pathologies have also been assigned to the activity of CGL, mainly due to its hydrogen sulfide production activity. Lowering the activity of CGL, using specific inhibitors, for example, was proposed to lower the incidence of allergic asthma at a young age [351] and attenuates the inflammatory response in a murine model of burn [352]. Furthermore, CGL activity is being linked to the pathological process of acute pancreatitis-associated lung injury [353], the nephrotoxic effects of gentamicin [354], and periodontal damage [355].

Despite all the important roles assigned to the CGL, the catalytic mechanism of CGL is still unknown. Deeper knowledge about how CGL catalyzes these reactions would provide valuable insights in terms of therapeutical approaches targeting this enzyme. Therefore, in this work, the catalytic mechanism of the cystathionine elimination into cysteine, α-ketobutyrate, and ammonia was studied employing computational methodologies.

The human form of CGL is a homotetramer with 405 amino acid residues per chain. Four identical active sites are available in a tetramer of CGL, and each one is composed of residues from two adjacent chains [356]. In each active site, one PLP molecule is anchored by Lys^{212} through a covalent bond, forming an internal aldimine (IA). Then, a substrate molecule enters the active site and reacts with the imine group of the IA, which results the substrate covalently bonded to the PLP cofactor. This structure is commonly known as external aldimine (EA) (**Figure 48**). So far, no further information is known about the catalytic mechanisms, except for some proposals where unknown basic and acid residues at the active site are involved [29, 332, 357-359]. The computational study of the mechanism of the methionine-γ-lyase (MGL) [29] can also provide insights into the mechanism of CGL since it belongs to the same family of enzymes.

C. Steegborn *et al*. determined the γ-lyase activity towards cystathionine of human CGL, reporting a k_{cat} of 1.9 s^{-1} at 30 °C [360]. According to the transition-state theory, this

rate-limiting step of the mechanism should account for 17.4 kcal/mol as activation Gibbs free energy.

Three directed-mutagenesis studies [358, 359, 361] have disclosed the importance of some residues in the catalytic mechanism of human CGL. W. Zhu *et al.* tested the impact of the Thr^{67} and Gln^{240} mutations in the γ-lyase activity towards cystathionine. These two residues are not in the active site and do not participate in the reaction. However, their mutations (Thr67Ile and Gln240Glu) cause an almost total impairment of the enzymatic activity, which is more prominent in the case of Gln240Glu mutation (**Figure 48**). The authors observed that these two residues are important for the PLP binding since the mutant forms present a considerable low amount of the PLP cofactor bonded in the active site. [358] The other two publications [359, 361] report the impact of the mutation of several active site residues in the γ-lyase activity towards cysteine, measuring the production of hydrogen sulfide (**Figure 47**). An almost total enzymatic impairment was accomplished when the Lys^{212}, Arg^{375}, Tyr^{114}, Thr^{189}, Asp^{187}, Arg^{62}, or Tyr^{60} residues were substituted by Ala (**Figure 48**). Lys^{212} is absolutely essential for the anchoring of PLP cofactor and formation of the IA. Moreover, it has been already proposed that Lys^{212} plays a critical role in some steps of the catalytic mechanism. Arg^{375} is important to appropriately place the substrate molecule within the active site since it establishes a strong salt bridge with the carboxylate group of the substrate. Tyr^{114} seems to provide important π-stacking interaction with the pyridine ring of PLP. In other PLP-dependent enzymes [188, 189], this role is usually assumed by a His residue, but this new data suggest that any single-ring aromatic residue can provide such stabilization. Even more so, the substitution of Tyr^{114} by a Phe in CGL causes an almost four times increase in enzymatic activity [359]. Asp^{187} and Thr^{189} seem to be important in the stabilization of the positive charge of the pyridine ring of PLP. Finally, Tyr^{60} and Arg^{62} show a crucial role in the stabilization of the phosphate group of PLP, contributing to anchor the PLP cofactor in the active site. Thr^{211} and Ser^{209} also contribute to the stabilization of phosphate group since their mutation by Ala lowers the enzymatic activity in ~40%. Interestingly, Glu^{339} mutation by more non-polar residues leads to a significant increase in hydrogen sulfide production. However, the impact of these mutations is not known for the γ-elimination activity of cystathionine.

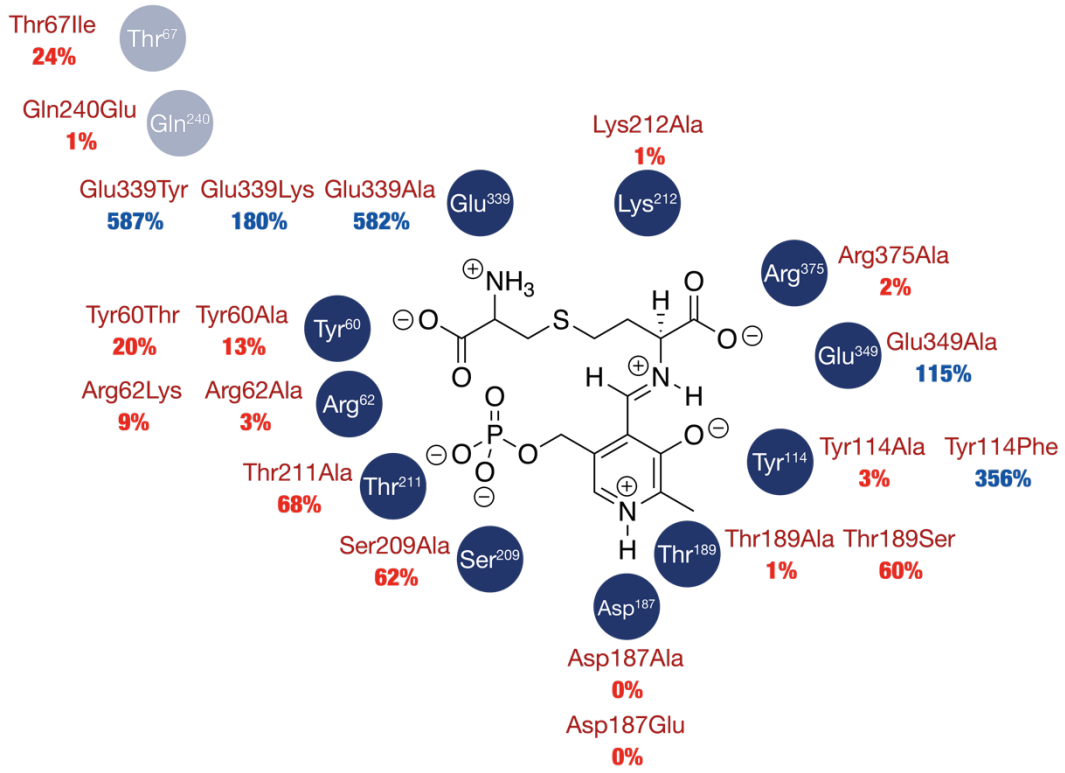


Figure 48. Map of amino acid residue of CGL for which there are mutagenesis studies. The residues are labelled in blue circles (dark: residues at the active site; light: residues far from the active site). Each reported mutation is shown with the respective ratio of enzymatic activity relative to the wild type form. [358, 359, 361]

Based on the available experimental data and published mechanism proposal, the catalytic mechanism of CGL was studied considering multiple possibilities in key steps of the mechanism. The mechanisms of activation of the enzyme and the formation of the EA [27, 28] were already studied and shown to be common to all PLP-dependent enzymes. Therefore, this work is exclusively devoted to the specific reactional steps catalyzed by CGL.

1.4.2. Methodology

a. Model

Among the eight crystallographic structures of human CGL available in the PDB, the structure with the PDB code 3COG [357] was chosen. This X-ray structure presents the best resolution (2.0 Å) and was obtained through the co-crystallization of CGL with PLP bonded to the Lys^{212} (IA) and 2-aminopent-4-enoic acid [357]. The structure is not completed, and some fragments are missing in some chains: 50-63 in chain A, 352-369 in chain B, and 56-57 in chain C. Since the structure of CGL is a homotetramer, it was possible to fulfill the missing residues superimposing the structures of all four chains of CGL. The EA (PLP + cystathionine) was placed in the active site using the atomic coordinates of the co-crystallized PLP cofactor as template for the alignment.

b. MD Simulation

The CGL structure and the EAs were solvated using TIP3P-FB [362] type water molecules. The protein was placed inside a water box so that each atom of the protein distanced at least 15 Å from each face. The system was neutralized by the addition of 30 sodium ions. The SANDER software from the AMBER18 package was used to add the hydrogen atoms to the model [363]. The protonation states for all amino acids were predicted by PROPKA 3.1 [275, 276] at pH 7.0.

The molecular mechanics parameters for the EA were obtained through a HF/6-31G(d) optimization to obtain the atomic charges (RESP) and the antechamber software from AMBER18 package [147, 148, 363] to assign the atom types.

The resulting solvated model underwent four sequential energy minimizations using the AMBER18 package and the GAFF and FB15 force fields [148, 364]: (1) minimization of all water molecules; (2) minimization of all hydrogen atoms; (3) minimization of all atoms, except backbone; (4) minimization of all system with no constraint.

The resulting minimized structure was used as the starting point for an equilibration of two steps: 50 ps where the temperature was increased from 0.00 K to 310.15 K in NVT conditions, and 50 ps at 310.15 K in NVT conditions. A 2.0 kcal mol^{-1} $Å^{-2}$ restraint force was applied to all backbone atoms during the equilibration.

Finally, a 200 ns MD simulation was performed using the NPT ensemble at 310.15 K and 1 bar. Both equilibration and production were conducted using SHAKE algorithm [365], an integration step of 2 fs, and a 10 Å cut-off for non-bonded interactions.

c. QM/MM

A set of key interatomic distances were evaluated along the entire MD simulation, looking for the best initial structure for the first step of the mechanism. Based on the mechanistic proposals, the first involves the abstraction of Hα by the Nζ of the Lys^{212}. In addition, based on the mechanism of enzymes of the same family, the proximity of O11, Nζ, and Sδ is important for the mechanism [29]. Therefore, the structure that minimizes the key distances 1 (2.49 Å), 2 (2.88 Å), 3 (2.98 Å), 4 (2.97 Å), and 5 (4.11 Å) (Figure 49) was chosen for building the ONIOM QM/MM [254, 366-368] model.

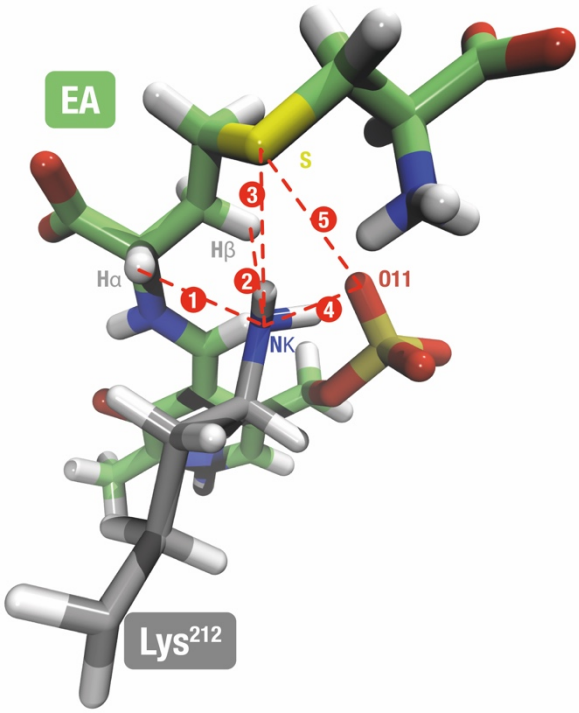


Figure 49. Structure of the EA and Lys^{212} at the active site of CGL that minimizes the distances 1, 2, 3, 4, and 5 among all frames of the MD simulation.

The QM/MM model was prepared using the molUP plugin [316, 369] for VMD software [292]. Due to the size of the system, only chains A and B, the respective EAs, and a 5 Å coat of water molecules were included in the model. The layer treated with QM includes all the atoms of the EA and the atoms from the sidechains of the Arg^{375}, Asp^{179}, Lys^{204}, and Arg^{62} residues (106 atoms). The remaining system was considered using MM (22,095 atoms) (Figure 50). Hydrogen atoms were used as link atoms to complete the valence of bonds that span across the two layers of the QM/MM scheme. The interaction between both layers was considered according to the electrostatic embedding scheme as implemented in Gaussian 09 [280].

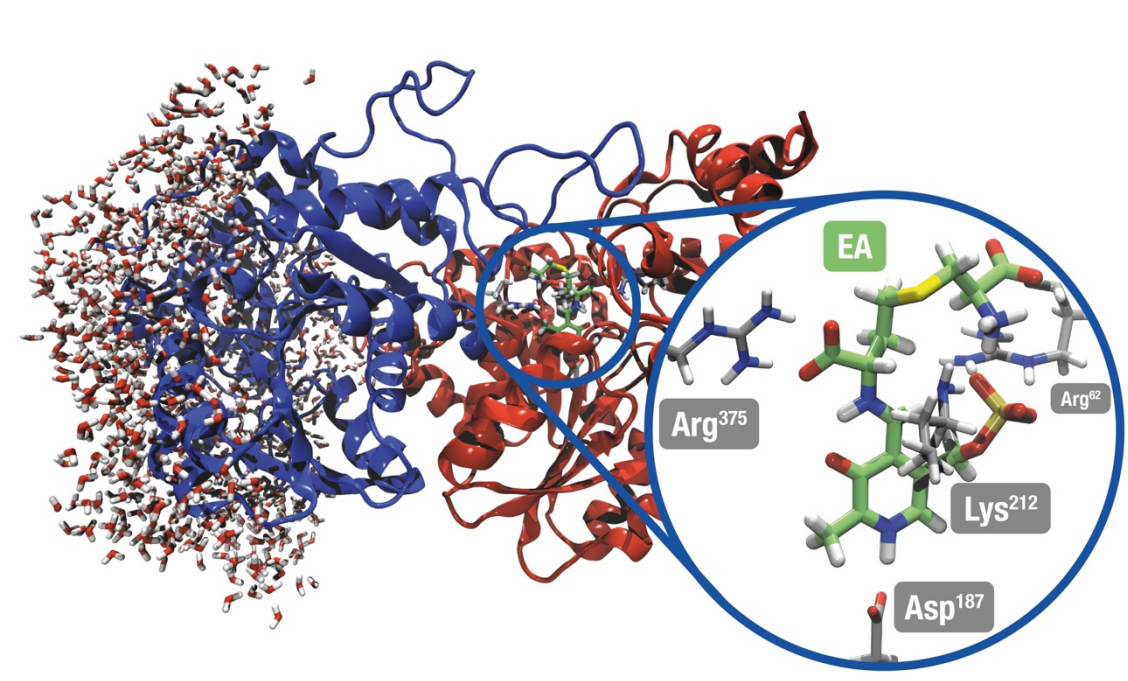


Figure 50. (Left) Structure of chain A (red) and B (blue) of CGL, representation of the QM layer, and part of the 5 Å coat of water molecules. (Right) Structure of the EA and sidechains of the active site residues included in the QM layer.

Each step of the mechanism was studied through linear coordinate scans to isolate the transition state (TS) of each reactional step. All TS were optimized and validated by vibrational frequency calculations, showing only one imaginary frequency associated with the vibration of the atoms involved in each step. The initial guesses for reactants and products were obtained by intrinsic reaction coordinate (IRC) [370] starting from the optimized TS. Each structure (reactant and product) obtained by IRC calculation was fully optimized and validated by imaginary frequency calculation, showing no imaginary frequencies.

The electronic energies were summed to the thermal corrections, calculated at 298.15 K and 1 atm, to obtain an estimative of the activation and reaction Gibbs free energies for each step of the mechanism.

All the calculations mentioned above were performed using Gaussian 09 [280]. The preparation of all input files and subsequent analysis of the output files were accomplished using molUP plugin [316, 369] for VMD.

The electronic energies will be refined through single-point energy calculations using DLPNO-CCSD(T)/CBS. The HL layer will be isolated to perform these calculations using the ORCA software (version 4.1.2) [250]. The SPs will be carried out using the cc-pVDZ and cc-pVTZ basis sets [185], and the cc-pVDZ/C and cc-pVTZ/C correlation fitting basis sets [317, 318], respectively. The combination of the energies obtained with cc-pVDZ |

cc-pVDZ/C and cc-pVTZ | cc-pVTZ/C basis sets for the HL layer are going to be used to extrapolate to the complete basis set (CBS) limit according to the Truhlar's extrapolation approach [319], excepting for the α and β exponents. The values 4.42 and 2.46 will be considered as α and β exponents in the extrapolation according to the ORCA documentation.

1.4.3. Results and Discussion

This work is still being conducted, but there are already results for the five first steps of the catalytic mechanism. Consequently, the following section provides a detailed description of the five initial steps of the γ-elimination of cystathionine catalyzed by the CGL. In addition, it is also presented the mechanistic proposals that are going to be tested for the remaining steps of the mechanism. The activation and reaction Gibbs free energies presented below do not include the refinement by single-point energy calculations with DLPNO-CCSD(T)/CBS.

a. α-Proton Transfer (2 Step)

The first step of the cystathionine γ-elimination catalyzed by the CGL starts with the nucleophilic attack of the Lys^{212} residue to the Hα proton of the substrate (**Figure 51** – Step 1). The NK atom is initially distancing 2.30 Å of the Hα proton of the substate. During step 1, the Hα proton is transferred to the Lys^{212} residue (NK-Cα distance: TS 1.23 Å; P 1.04 Å), and the resulting electron pair is lodged and stabilized by the pyridine group of the PLP cofactor. This reaction requires an activation Gibbs free energy of 19.7 kcal/mol and is endergonic by 18.8 kcal/mol. Both minima and TS were validated through additional vibrational frequency calculations, and only the TS presented one imaginary frequency at 1097.7i cm^{-1}. This frequency has the major contribution of the Hα atom vibration, which corresponds to the atom transferred in this first step of the mechanism. (**Figure 51** – Step 1: TS)

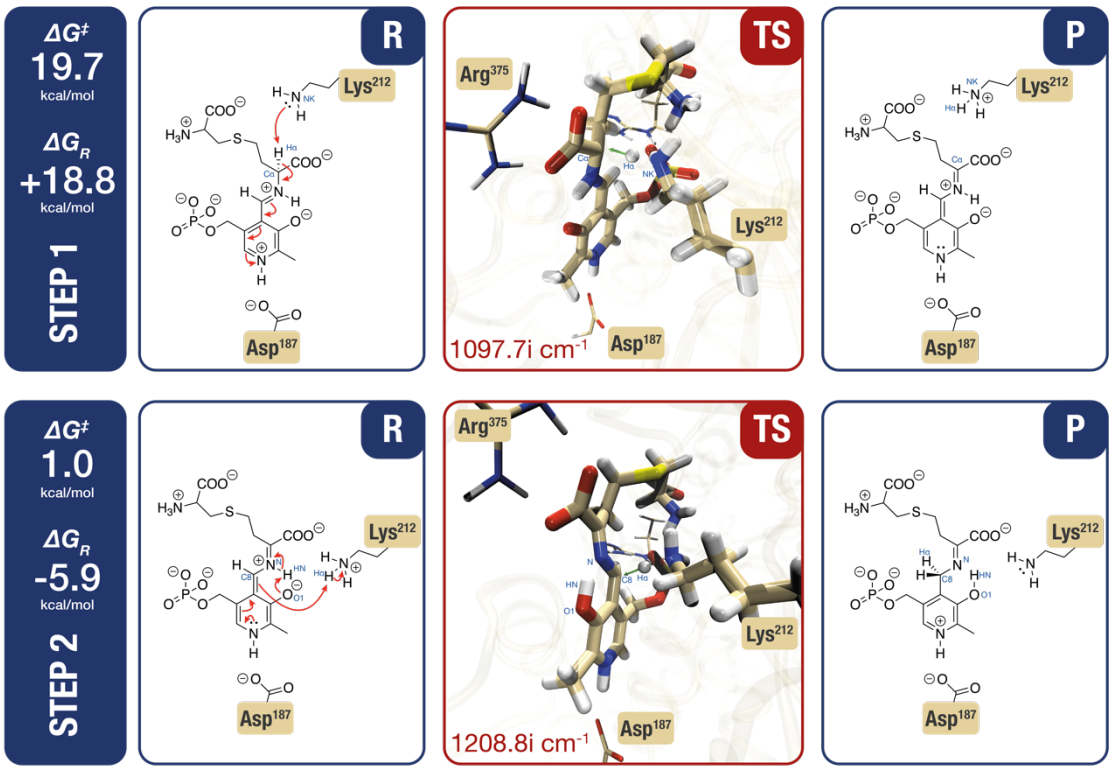


Figure 51. Structures of reactant (R), transition state (TS), and product (P) of the two initial and sequential steps associated with the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism.

This first step of the mechanism agrees with the recently published study on the γ-elimination of the PLP-dependent MGL [29]. According to Zhe Li *et al.* [29], the Lys^{212} that becomes positively charged in step 1 should lose one proton to the phosphate group of the PLP cofactor and then abstract the Hβ proton from the substrate. Two additional mechanistic hypotheses were tested for the second step (Figure 52).

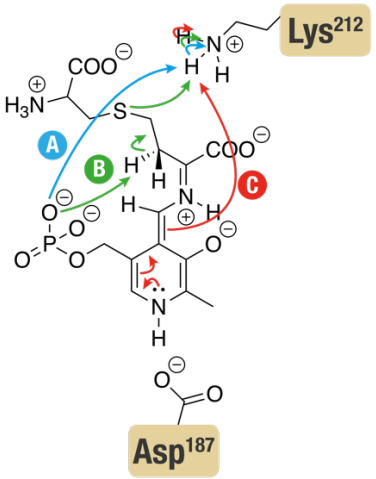


Figure 52. Scheme of the three mechanistic hypotheses tested for the catalytic mechanism of step 2.

Hypothesis A (Figure 52 - A) involves the proton transfer from the Lys^{212} residue to the phosphate group of the PLP as suggested by Zhe Li *et al.* [29]. Our calculations have shown that this reaction is possible with an activation energy of 3.4 kcal/mol and a reaction energy of 2.0 kcal/mol. However, the next step, considering this mechanistic pathway, showed to be impossible with an activation barrier that exceeds 40 kcal/mol and no reaction product is formed. Therefore, according to the calculations performed, the second step might not follow the same mechanism as described for MGL.

Then, another hypothesis was tested (Figure 52 - B), where the Lys^{212} directly protonates the S atom of the substrate simultaneously with the abstraction of the Hβ proton from the phosphate group. The calculations showed that the reaction could not occur under biological conditions since the estimated activation barrier is greater than 35 kcal/mol.

Finally, the proton transfer from Lys^{212} to C8 atom was tested in hypothesis C (Figure 52 - C). In this attempt, the Hα proton abstracted in step 1 is transferred to the C8 atom of the PLP factor. This reaction requires an initially thermodynamically favorable rearrangement of Lys^{212} residue (-2.4 kcal/mol), which rotates ~180 ° across its Cδ-Cε bond. After that rotation, the NK atom distances 2.26 Å from the C8 atom, and the reaction becomes almost spontaneous ($\Delta G^{\ddagger}$ = 1.0 kcal/mol) (Figure 51 – Step 2). Step 2 is exergonic by 5.9 kcal/mol. The reaction is possible through the mobilization of the electron pair generated in step 1 that was lodged in the pyridine ring of the PLP cofactor. At the same time, the HN proton is transferred from the N nitrogen (N-HN distance: R 1.07 Å; TS 1.56 Å; P 1.55 Å) to the O1 oxygen (HN-O1 distance: R 1.58 Å; TS 1.03 Å; P 1.03 Å) of the PLP factor. As in step 1, the minima and TS were validated through frequency calculations, and the TS revealed one imaginary frequency at 1208.8i cm^{-1} (Figure 51 – Step 2: TS).

b. β-Proton Abstraction (1 Step)

At the end of step 2, the Lys^{212} residue restores its initial neutral form and is available to act as a nucleophilic species in another reaction. Consequently, the NK atom of Lys^{212} residue collects the Hβ proton of the substrate (NK-Hβ distance: R 3.09 Å; TS 1.33 Å; P 1.03 Å). The Hβ proton abstraction triggers the formation of a double bond between Cβ and Cα (Cα-Cβ distance: R 1.52 Å; TS 1.43 Å; P 1.37 Å) and the re-protonation of the N nitrogen (Figure 53). This reaction requires an activation Gibbs free energy of 13.3 kcal/mol, and it is endergonic by 7.7 kcal/mol. All minima and TS were validated as in previous steps, and the TS has an imaginary frequency at 1198.6i cm^{-1} involving the vibration of the Hβ atom transferred in this step (Figure 53 - TS).

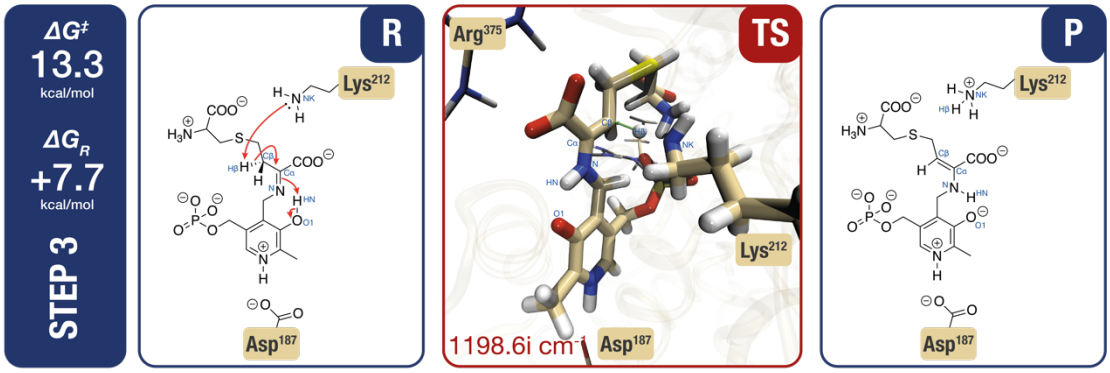


Figure 53. Structures of reactant (R), transition state (TS), and product (P) of the step 3 of the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism.

c. γ-Elimination of Cysteine (1 Step)

At this stage, the Lys^{212} residue is ready to protonate the S atom of the substrate and catalyze the γ-elimination of cysteine. The Hβ atom, covalently bonded to Lys^{212}, is transferred to the S atom of the cystathionine (Hβ-S distance: R 2.76 Å; TS 1.62 Å; P 1.35 Å) (Figure 54). The elimination of cysteine is triggered by the double bond formation between Cγ and Cβ atoms (Cβ-Cγ distance: R 1.49 Å; TS 1.35 Å; P 1.34 Å), that in turn results from the electron density mobilization from the O1 atom. In this process, the HN proton transferred again from O1 oxygen to N nitrogen (Figure 54). The reaction is almost thermoneutral (ΔG_R = 3.5 kcal/mol) and requires an activation Gibbs free energy of 12.3 kcal/mol. This energy barrier is significantly lower than the ones obtained when the direct elimination of cysteine was tried in step 2. This suggests that CGL evolved to catalyze a set of three initial proton transfers in order to prepare the substrate for the γ-elimination in a more efficient way.

The frequency calculations revealed only one imaginary frequency for the TS structure (955.4i cm^{-1}), mainly associated with the vibration of the Hβ atom that is transferred from Lys^{212} to the S atom of the substrate. At the end of this step, one of the final products of the global reaction is formed, cysteine.

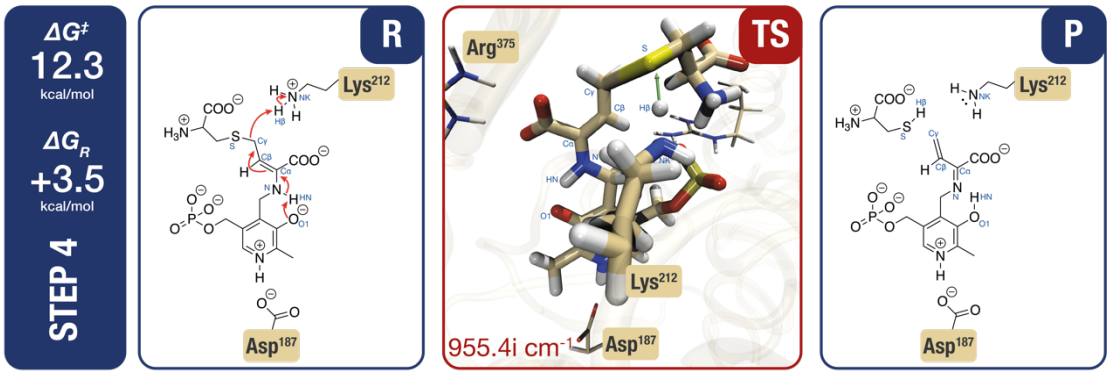


Figure 54. Structures of reactant (R), transition state (TS), and product (P) of the step 4 of the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism.

d. γ-Protonation (3 Steps)

After the elimination of cysteine in step 4, the following steps of the mechanism aim to eliminate the ammonia and α-ketobutyrate molecules. Thereupon, three sequential steps are required to protonate the Cγ of the intermediate generated in step 4.

First, the Lys^{212} recovers the Hα proton that previously has been transferred to the C8 atom of PLP in step 2 (Hα-NK distance: R 1.96 Å, TS 1.38 Å; P 1.03 Å). This reaction follows the reverse mechanism of step 2, and the pyridine group of PLP cofactor stabilizes the electron pair generated. Step 5 is very fast, requiring only an activation Gibbs free energy of 3.9 kcal/mol. This step is slightly exergonic by 2.8 kcal/mol. As in all previous steps, the frequencies calculations only unraveled one imaginary frequency for the TS, corresponding to the vibration of the atoms involved in the reaction.

At the end of step 5, the Lys^{212} residue assumes a positively charged form and can protonate the Cγ atom of the intermediate. The direct protonation of Cγ carbon from Lys^{212} residue was tested, but it proved impossible. Such reaction required an activation energy greater than 35 kcal/mol. Therefore, the protonation via cysteine was tested.

In this mechanistic hypothesis, the released cysteine molecule transfers its Hβ proton to the Cγ atom (Hβ-Cγ distance: R 2.35 Å; TS 1.39 Å; P 1.10 Å). This step is possible because of the electron pair delocalization from the pyridine group of the PLP cofactor to the Cγ atom. In this process, the Cγ-Cβ (R 1.35 Å; P 1.40 Å), Cα-N (R 1.32 Å; P 1.41 Å), and C8-C7 (R 1.42 Å; P 1.47 Å) become single bonds, whereas the Cβ-Cα (R 1.46 Å; P 1.36 Å) and N-C8 (R 1.36 Å; P 1.30 Å) become double bonds. This reaction is almost thermoneutral (ΔE_R = +0.6 kcal/mol) and has an activation energy of 9.1 kcal/mol.

After the protonation of the Cγ atom, the cysteine molecule becomes negatively charged and occupies a position near the positively charged Lys^{212} residue (Hα-S distance: R 2.02 Å; TS 1.63 Å; P 1.35 Å). Consequently, the Hα proton is quickly transferred from Lys^{212} to the S atom of the cysteine. The reaction only requires an activation energy of 2.3 kcal/mol, and the product is 5.1 kcal/mol more stable than the reactant of this seventh step.

The results for steps 6 and 7 only considered the electronic energies obtained from the geometry optimizations. The frequency calculations are still being calculated, as well as the remaining steps of the mechanism until the release of the α-ketobutyrate and ammonia molecules.

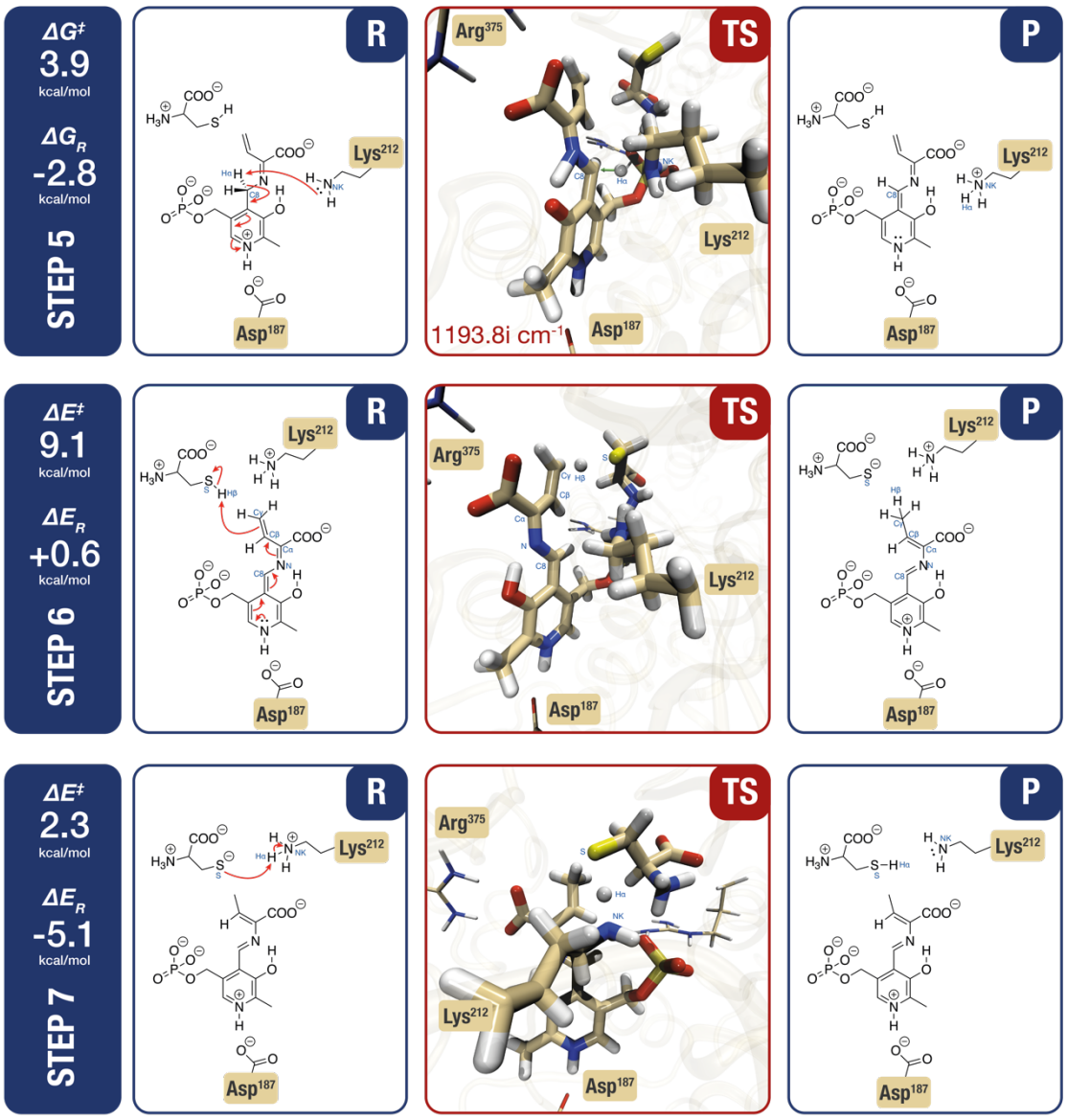


Figure 55. Structures of reactant (R), transition state (TS), and product (P) of steps 5, 6, and 7 of the γ-elimination mechanism catalyzed by the CGL. Wedge-Dash representation of the structures of reactant and product. Three-dimensional representation of the TS structure. The green arrows correspond to the major vibrational frequency associated with the imaginary frequency of the TS. The computed activation and reaction energies are also shown for each step of the mechanism.

1.4.4. Conclusion

This computational study of the γ-elimination of cystathionine by CGL is still ongoing. However, some conclusions can be taken from the first batch of results.

In particular, the results seem to identify the first step as the rate-limiting step of the entire mechanism. Although further energy refinements by single-point energy calculations are required, all other computed steps have substantially lower activation barriers.

So far, Lys^{212} plays a critical role in all steps of the catalytic mechanism where it is involved in the cleavage and formation of covalent bonds, except for step 6. Lys^{212} assumes the role of a base in steps 1, 3, and 5, but it also acts as an acid in steps 2 and 6. Therefore, the unknown basis and acidic species discussed in the literature are, in fact, assumed by the Lys^{212} residue. The capability of Lys^{212} residue exchange between a base and an acid is associated with an uncommon feature of the PLP cofactor. Most of the PLP-dependent enzymes employ the PLP cofactor to orientate the substrate (usually an amino acid) in the most suitable position to react and to act as an electron sink to promote the elimination of atoms or molecular groups bonded to the Cα atom of the substrate. In the case of CGL, the PLP cofactor also plays an interesting role in transiently storing protons between steps of the catalytic mechanism. In step 2, the Lys^{212} residue transfers a proton to the C8 carbon of the PLP cofactor. This allows the Lys^{212} residue to restore its base character. This proton is kept by the PLP cofactor until it is required again, and the Lys^{212} can easily get that proton back in step 5. Moreover, both steps (2 and 5) are very fast (low activation barriers) and thermodynamically favorable. This unique behavior does not seem to be common to other PLP-dependent γ-elimination enzymes, as MGL, where transient protons are stored at the phosphate group of the PLP cofactor [29]. Alternatively, this difference can result from the different substrates (methionine vs. cystathionine) rather than the enzyme itself. In the case of cystathionine, the terminal positively charged amine group establishes a strong interaction with the negatively charged phosphate group of the PLP cofactor. Consequently, the phosphate group may be less available to accept protons during the different steps of the catalytic mechanism.

It was also noticed that the cysteine molecule released as a product of the reaction act as a crucial catalytic player in step 6. The protonation of the Cγ carbon of the intermediate turned impossible when the protonation occurs directly from the Lys^{212} residue. In this scenario, one product actively participates in the mechanism of subsequent steps.

Once the study of this mechanism is finished, a detailed description of the γ-elimination mechanism of cystathionine by CGL will be provided. Moreover, it will unravel some new features of PLP as a cofactor, namely its capability to store protons transiently.

1.5. Chemistry of PLP-dependent enzymes

The main goal of this Ph.D. project was the description of the catalytic mechanism of PLP-dependent enzymes with an atomistic and electronic level of detail. The previous sections of this chis chapter C disclosed the catalytic mechanism of decarboxylation, α-elimination, and γ-elimination (ongoing) reactions catalyzed by PLP-dependent enzymes.

All these PLP-dependent enzymes share similar mechanisms for the enzymatic activation and transimination reactions (Figure 5). After that, each enzyme catalyzes a totally distinct reaction following different catalytic mechanisms. However, the study of these mechanisms revealed some particularities common to PLP-dependent enzymes. Besides, some chemical features of PLP were disclosed and explained in terms of how enzymes can employ the same cofactor to catalyze so much distinct reactions.

In this section, the most important features of the PLP cofactor and PLP-dependent enzymes are analyzed and described.

1.5.1. Dunathan's theory, Intermolecular Interactions, and Electron Sink

HDC, SHMT, and CGL are PLP-dependent enzymes that catalyze different reactions: decarboxylation, α-elimination, and γ-elimination. According to each enzyme's catalytic mechanism, the first step of the mechanism involves the elimination of a different atom or molecular group from the substrate. In the HDC, the first step consists in the elimination of the carboxylate group into carbon dioxide. The SHMT catalyzes the sidechain elimination from serine. In the case of CGL, the proton bonded to the Cα is abstracted in the first step. These three enzymes use PLP cofactor to eliminate different groups of the respective amino acid substrate. Although the substrates share a similar structure, where the carboxylate group and the Hα are common to all of them, the eliminated group is different for each enzyme.

From the analysis of the relative positioning of the substrate's groups, the results showed that the elimination always occurs perpendicular to the pyridine ring of the PLP cofactor (Figure 56). According to the optimized structures of each reactant, each enzyme constrains the orientation of the substrate to orientate the target group in a perpendicular position. The HDC positions the carboxylate group in a 78° relative to the pyridine ring. The SHMT orientates the Cα-Cβ bond to become at 98°, whereas the CGL makes the Hα at 113° relative to the PLP cofactor plane. (Figure 56 - Bottom)

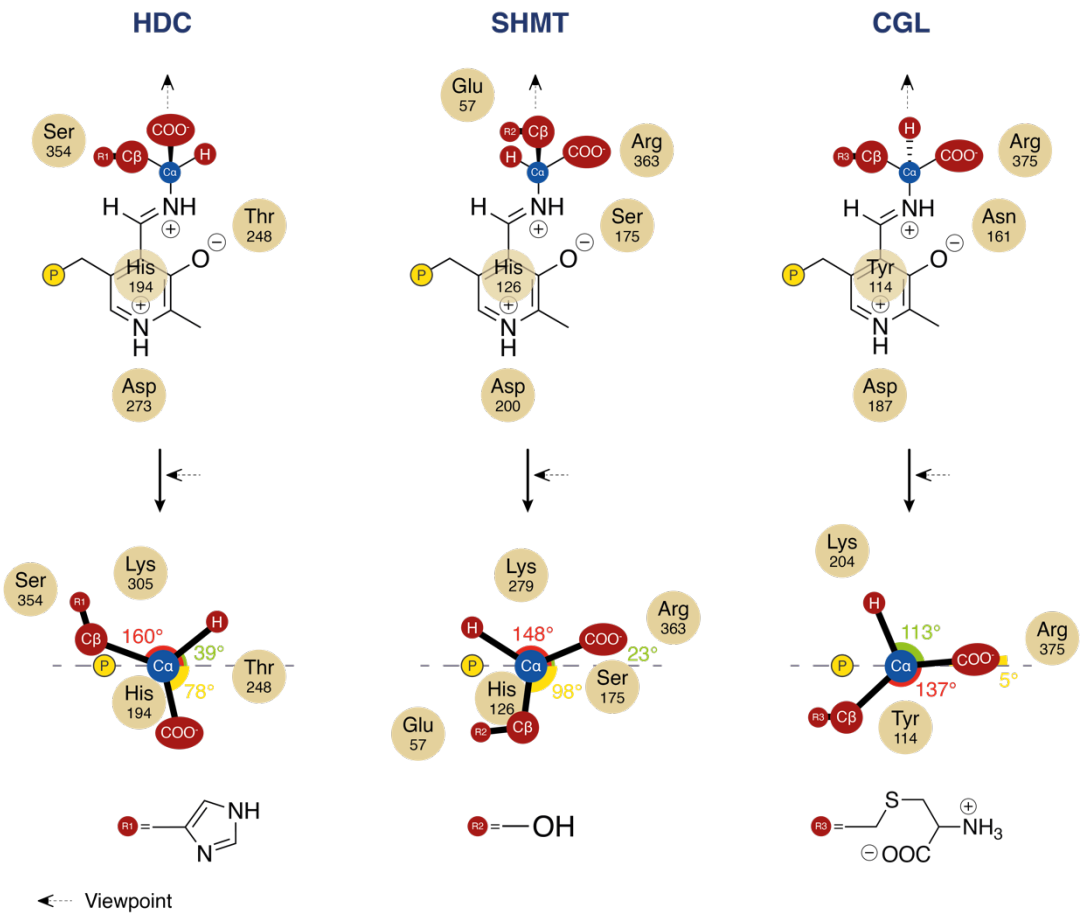


Figure 56. Key interactions of the EA with the active site of the HDC, SHMT, and CGL enzymes. (Bottom) Top view of the EA with labels for the angles of each atom/group bonded to the Cα of the substrate. The angle is measured starting from the pyridine's plane.

This evidence was observed for all PLP-dependent enzymes studied in this project, and it agrees with Dunathan's theory [295, 296]. According to Dunathan's proposal, the covalent bond that is cleaved in PLP-dependent enzymes should be perpendicular to the PLP plane. When the target σ bond (Cα-Cβ, Cα-C_{COO}, or Cα-Hα) is perpendicular to the PLP cofactor, it occupies the same region of the π-bonds that compose the conjugated system of the PLP cofactor. Consequently, the electron density of the σ bond can be highly delocalized throughout the PLP cofactor. The electron delocalization weakens the covalent bond, making the reaction faster and thermodynamically more favorable.

The appropriate orientation of the target atom/group is intrinsically associated with key interactions of the EA with the amino acid residues of the active site of each enzyme. In the case of SHMT and CGL, Arg residues play a crucial role in this correct positioning of the substrate. The Arg^{363} residue of the SHMT makes an ionic bridge with the carboxylate group of the serine (substrate), placing the carboxylate group in a way that the Cα-Cβ bond is forced to assume an almost perpendicular position. In the case of SHMT, Glu^{57}

is also important for the orientation of the leaving group since it establishes a hydrogen bond with the hydroxyl group of the substrate (Figure 56 - SHMT). The situation is similar for the CGL enzyme, where the Arg^{375} residue orientates the carboxylate group to a similar position but on the opposite side of the PLP cofactor. In this way, the Hα proton becomes almost perpendicular and points to the Lys^{204} residue (Figure 56 - CGL). Naturally, the HDC is a different case where the carboxylate group cannot be stabilized by the Arg residue since it is the target group for the elimination. Consequently, the enzyme evolved to stabilize and accommodate the sidechain of the substrate. The stabilization of the imidazole group (in the case of HDC) orientates the carboxylate group to the perpendicular position to catalyze its release (Figure 56 - HDC).

In sum, these results show that Arg residues play an important role in the correct orientation of the substrate for the catalysis through a strong interaction with the carboxylate group. Only for decarboxylases, the approach needs to be different since the eliminated group is the carboxylate itself. In this situation, the sidechain of the substrate is accommodated in a pocket that plays the role of the Arg residue.

In addition, three other important and commonly observed residues were identified. In all enzymes, a non-bulky residue capable of establishing hydrogen bonds assumes a position near the oxo group of the PLP cofactor. Thr^{248}, Ser^{175}, and Asn^{161} establish hydrogen bonds with the PLP cofactor in HDC, SHMT, and CGL enzymes, respectively (Figure 56).

Also observed in all studied enzymes, a His or Tyr residue occupies a strategic position where it interacts with the pyridine group of the PLP cofactor. The π-π interaction between the pyridine group and the imidazole or phenyl groups of the His or Tyr residues, respectively, provides a suitable pocket to orientate the PLP cofactor. Moreover, this interaction prevents possible rotation of the PLP cofactor and stabilizes the negative charge generated after the elimination reaction. The generated excess of electron density is delocalized across the PLP cofactor, and this residue can also contribute to such stabilization (Figure 56).

Finally, an Asp residue is found at the bottom of all PLP cofactors within the active site of each enzyme. This negatively charged residue makes a hydrogen bond with the protonated nitrogen of the pyridine group. In some extension, the Asp residue stabilizes the positive charge lodged in the pyridine group. However, the negative charge of the Asp residue is also stabilized by a set of other hydrogen bonds. This stabilization of the Asp residue is crucial to prevent an excessive stabilization of the positive charge of the PLP cofactor. In such hypothetic case, the polarization of the PLP cofactor will vanish,

and the cleavage of the target bond could be compromised. On the other hand, the lack of such interaction could cause an over-stabilization of electron pair, precluding its availability for a posterior step where a different atom or molecular group binds the Cα atom (Figure 56).

The ability of PLP to promote the cleavage of a covalent bond, to transiently stabilize the resulting electron density, and to make it available for establishing new covalent bonds results from balanced interactions between the cofactor and the enzyme. In part, the catalytic power of PLP is due to its non-permanent electron sink property. The electrostatic potential surfaces computed for the steps of HDC (Figure 22) and SHMT (Figure 29 and Figure 37) catalytic mechanisms show that effect. In the elimination step, the electron density increases in the pyridine region of the PLP cofactor. When those electrons are requested for establishing a new bond, the electron density decreases at this region since the electrons are requested for the new covalent bond.

1.5.2. Substrate Specificity

There are 224 different PLP-dependent enzymes that catalyze a myriad of different reactions. The main targets of PLP-dependent enzymes, in terms of substrate, are amino acids. There are also some PLP-dependent enzymes that catalyze the same type of reaction but towards different substrates. In particular, HDC and AADC are enzymes that catalyze the decarboxylation of different amino acids. HDC is specific for L-histidine, whereas AADC can catalyze the decarboxylation of L-DOPA, L-phenylalanine, L-tyrosine, L-histidine, L-tryptophan, and 5-hydroxytryptophan.

When the structures of HDC and AADC are superimposed (Figure 57), the key residues of the active site are the same for both enzymes. Particularly, the Asp residue that interacts with the pyridine group of the cofactor, the His residue that makes a π-π interaction with the pyridine group of the PLP cofactor, and the Thr residue that stabilizes the oxo group of the cofactor.

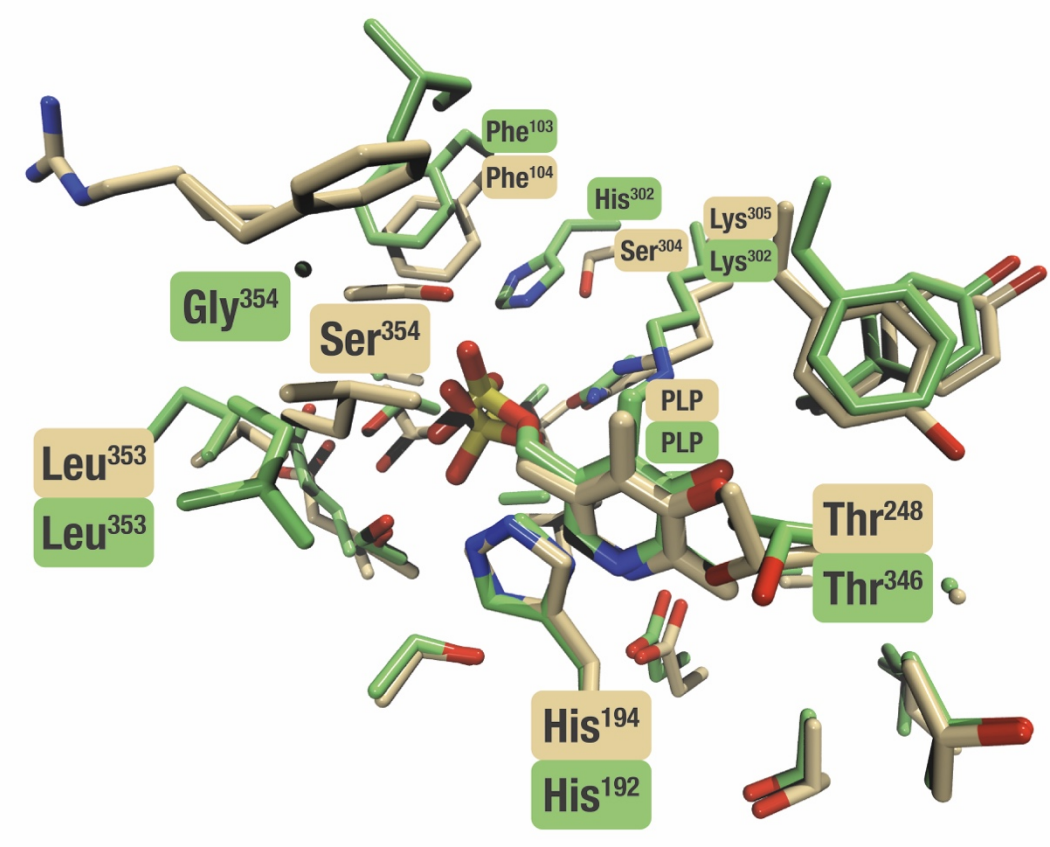


Figure 57. Structural comparison of the active sites of HDC and AADC. The X-ray crystallographic structures of (wheat) HDC (PDB code 4E1O) and (green) AADC (PDB code 1JS6) were superimposed. The key active site residues are labelled and colored according to the respective structure.

Although both enzymes evolved to catalyze the same type of mechanism using the same cofactor and similar active site moiety, AADC is substantially more promiscuous than HDC. Since the only difference between the substrates is the sidechain, these two enzymes should have differences regarding the pocket that engulfs the sidechain. A close analysis of the pocket reveals a key residue involved in the specificity of HDC to L-histidine. The Ser^{354} residue makes a hydrogen bond with one of the nitrogen atoms of the imidazole group of the substrate. In AADC, this position is occupied by a Gly residue which turns the pocket bigger and less specific. Consequently, AADC can accommodate different molecular groups, whereas HDC is exclusive for L-histidine. The importance of Ser^{354} has already been reported in a structural study, where the specificity of HDC is determined by this residue [261]. In that work, the mutation of Ser^{354} by a Gly causes a decrease in HDC's affinity to L-histidine. This mutation also turns HDC suitable to accept and catalyze the decarboxylation of L-DOPA [261]. In sum, the Ser^{354} residue is the key difference between HDC and AADC in terms of which substrates can be decarboxylated by both enzymes.

1.5.3. Proton storage

The study of possible strategies to improve SHMT for industrial application showed that the mutant forms (Glu57Gln and Glu57Leu) are able to catalyze the retroaldol cleavage using one oxygen atom from the phosphate group of PLP as a base. In such a mechanism, the proton that is abstracted from the Glu^{57} residue in the wild-type enzyme is transferred to the phosphate group of PLP via a water molecule. In the next step, the proton is transferred again from the phosphate group to the Cα atom of the quinonoid intermediate. Therefore, the phosphate group of PLP assumes the role of a base during the first step and an acid in step 2. This evidence characterizes the phosphate group as more than an anchor of PLP within the active site of the enzyme. In fact, the phosphate group can also be catalytic active and momentarily store a proton between steps of the catalytic mechanism.

In the case of CGL, a similar situation occurs, where the Lys^{204} residue captures the Hα proton, but it needs to be "stored" by another species to restore nucleophilic character of the Lys residue. In the assessment of different mechanistic proposals, our calculations showed that the phosphate group could accept that proton, but this would compromise the next step, making the entire catalytic cycle unfeasible. Then, it was tested the Hα proton transference from the Lys^{204} residue to the C8 atom of the PLP cofactor. The calculation shows a very (kinetic- and thermodynamically) favorable reaction. Moreover, the reverse reaction is also very fast, meaning that the proton can be easily transferred back to the Lys^{204} residue. The obtained results show that the PLP cofactor can also "store" a proton at the C8 position, making it available when requested later in the catalytic cycle.

In addition to all already known features of the PLP cofactor, the calculations in this project show that PLP can act as a base and transiently store protons through either the phosphate group or the C8 atom.

1.5.4. The Phosphate Group

As already mentioned, the PLP is the catalytically active form of vitamin B_6, which means that the vitamin molecule needs to undergo a set of reactions to become ready for its catalytic purpose. One important step is the oxidation of the hydroxyl group, by the PNPOx enzyme, to produce an aldehyde (Figure 3). The aldehyde group is significantly more reactive with the primary amine groups of the substrates of PLP-dependent enzymes. Thus, this modification enhances the reactivity of the PLP cofactor towards the substrate molecules.

Nonetheless, this modification is not enough to turn vitamin B_6 ready for catalysis. Vitamin B_6 needs also to be phosphorylated by the PLK (Figure 3). The PLK is a kinase that transfers a phosphate group to the 5' hydroxyl group of vitamin B_6. This phosphate group plays a critical role in positing the PLP cofactor within the enzyme's active site. In all PLP-dependent enzymes, the PLP cofactor is tightly anchored in the active site through a set of hydrogen bonds and salt bridge interactions with residues from the active site. Arg and/or Lys residues appeared to establish strong ionic interactions with the negatively charged oxygen atoms of the phosphate group. Moreover, different residues participate in the phosphate group stabilization through hydrogen bonds either with their polar sidechains or the backbone amine groups.

The role of the PLP phosphate group as an anchor seems to be ubiquitous to all PLP-dependent enzymes. However, in some cases, the phosphate group can also act as a base and directly participate in chemical reactions during the catalytic mechanism, as previously mentioned.

1.5.5. The Pyridine Ring

The pyridine ring of the PLP is the molecular region most responsible for the electron sink feature of this cofactor. The electronegativity of the nitrogen atom withdraws the electron density of the ring, contributing to its electron deficit. Moreover, in the studied enzymes, this effect is even more drastic since the nitrogen atom is positively charged and interacts, through a hydrogen bond, with a nearby Asp residue. As aforementioned, this interaction is important to balance the electron sink property of PLP, facilitating the cleavage of Cα-R bonds of the substrate without compromise the formation of new ones. Further investigation will be conducted to assess the impact of the nitrogen atom on the overall reactivity of different PLP-dependent enzymes. To do so, the nitrogen atom would be substituted by a carbon atom to understand the importance of this specific atom in the PLP cofactor. Moreover, different substitutions could be performed to look for more efficient cofactor analogs with potential industrial application. For example, the pyridine moiety could be replaced by a phosphorine ring, substituting the nitrogen atom by a phosphorus atom.

There are also two substituent groups at the 2 and 3 positions of the pyridine group.

The 3-oxo group is crucial in the interaction with the N-H of the imine bond. This interaction stabilizes the positive charge at the nitrogen atom of the imine bond. The performed calculations have shown that, in some reactions, the proton is transferred to the oxo group and then to the nitrogen atom again. This highly labile proton can easily

exchange between these two atoms. The significance of the 3-oxo group becomes evident in the study of the catalytic mechanism of HDC. The crystallographic structure of HDC (PDB code: 4E1O) includes structural information about the external aldimine of an inhibitor very similar to the natural substrate. In that structure, the imine bond is pointing towards the phosphate group instead of the 3-oxo group of PLP. Consequently, all the attempts to eliminate the carboxylate group of histidine resulted in prohibitive activation barriers. As soon as the imine bond was rotated to occupy the position commonly known for all other PLP-dependent enzymes, the interaction of the NH and oxo groups was restored and the estimated barriers for the decarboxylation closely agreed with the experimental kinetic data. This experiment provides factual evidence that the 3-oxo group is important for the catalytic process through the interaction with the protonated nitrogen atom of the imine bond.

Regarding position 2 of the pyridine ring, the impact of the methyl group is not known. Therefore, further calculations are going to be performed to disclose the relevance of such group, in particular, repeating the calculations in the absence of this methyl group. The methyl group is an electron donor through inductive effect, slightly increasing the electron density of the pyridine ring. Therefore, its presence can attenuate the electron-withdrawing power of the pyridine group, allowing a good commitment between the catalytic efficiency in cleaving and forming covalent bonds with the Cα atom of the substrate.

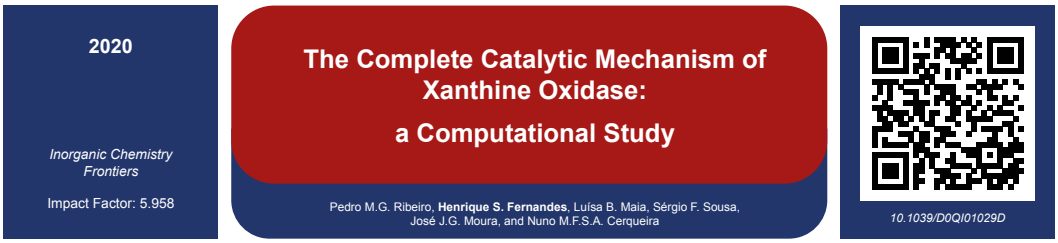
2. Mo-dependent Enzymes

This page was intentionally left blank.

2.1. The Catalytic Mechanism of Xanthine Oxidase

The catalytic mechanism of the xanthine oxidation into uric acid by xanthine oxidase (XO) was studied using a QM/MM approach [195].

The calculations were started by Pedro M.G. Ribeiro. Then, Henrique S. Fernandes finished the calculations, analyzed the results and wrote the article with contributions of all the co-authors. The work was published in the Inorganic Chemistry Frontiers journal, and the following content is almost a transcription of the published version.



In this work, QM/MM methods were used to study the full catalytic mechanism of XO. XO catalyzes the conversion of XAN to URC, in the presence of a molybdenum cofactor (Moco). The mechanism occurs through four reaction steps. Initially, the proton from the hydroxyl group of Moco passes to Glu^{1261} and the activated hydroxyl group makes a nucleophilic attack on XAN. Then, a hydride is transferred from the tetrahedral intermediate to the sulfur atom of the Moco, reducing Mo(VI) to Mo(IV). In the third step, one molecule of URC is formed through its protonation by Arg^{880}. Once this reaction is complete, FAD is reduced to $FADH_2$, oxidizing Mo(IV) to its initial oxidation state of Mo(VI). The enzymatic turnover is achieved with the reaction of one water molecule with the Moco. The rate-limiting step of the full catalytic mechanism is the hydride transfer that requires a free activation barrier of 16.9 kcal/mol, which closely agrees with the experimental k_{cat} value (18.3 s^{-1}), which corresponds to approximately 15.7 kcal/mol. This work also elucidates the key role played by Arg^{880} in the catalytic mechanism and the importance of Glu^{802} in the binding of the substrate. Both residues were previously shown to be important by mutagenesis studies, but their role was still not clearly understood. Additionally, it was observed that the presence of a tunnel of water molecules located close to Moco and Glu^{1261} is important for the enzymatic turnover. The determined transition state structures can now be used to help the development of transition-state analog inhibitors targeting XO.

This page was intentionally left blank.

2.1.1. Introduction

XO, E.C. 1.17.3.2, is a Mo-dependent enzyme involved in the catalysis of the two rate-limiting steps in the degradation of purine nucleotides. XO catalyzes the oxidation of hypoxanthine (HPA) into xanthine (XAN) and XAN into uric acid (URC). The full process requires two water molecules and two dioxygen molecules that are converted into two hydrogen peroxide molecules [371] (Figure 58). Structurally, XO is a homodimer enzyme-containing 1333 amino acid residues per subunit.

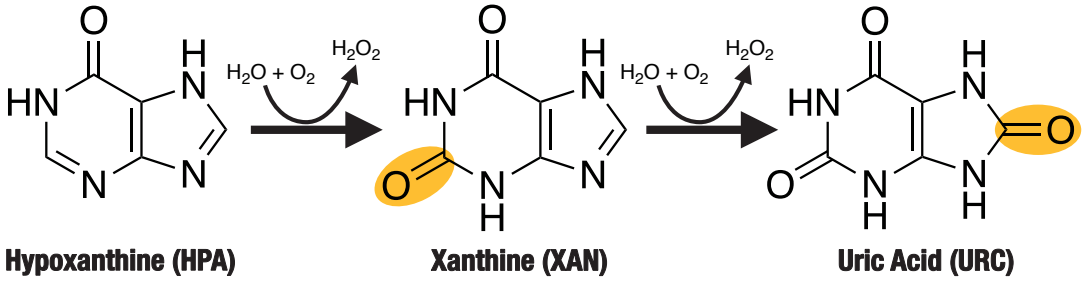


Figure 58. Reactions catalyzed by XO: (1) oxidation of HPA to XAN, and (2) oxidation of XAN to URC.

XO has been extensively studied throughout the years due to its relationship with some diseases like gout, the formation of uric acid stones, and xanthinuria. Gout and uric acid stones formation (hyperuricemia) [94-99] are caused by intense activity of XO that increases the levels of URC. Consequently, uric acid tends to precipitate in joints, skin, and blood capillaries in the form of needle-like crystals causing pain [99]. URC can also originate kidney stones in the form of sodium urate crystals. Xanthinuria is a rare genetic disorder that is associated with low XO activity and that can lead to high levels of circulating XAN and is responsible by some disease conditions like renal failure [372, 373].

The main research is devoted to the search for XO inhibitors that could be used to treat diseases like gout. Actually, there are already a few commercial inhibitors used in clinic: e.g. allopurinol [374, 375] and 6-mercaptopurine [376]. However, the seek for better new inhibitors continues [377-385], largely because of the increasing in gout incidence [98, 99] that accounted for almost 6 million people in 2013 [386].

A lot of information regarding the catalytic mechanism of XO has been obtained through the years. Kinetic parameters of XO were determined experimentally, revealing a reaction rate of 30.0 s^{-1} for the conversion of HPA into XAN, and 18.3 s^{-1} for the oxidation of XAN into URC (pH 8.5 and 25 °C) [72]. In terms of mutagenesis studies, Yamaguchi and collaborators [72] addressed the impact of the mutation of two residues at the active site of XO: Glu^{802} and Arg^{880} residues. The $Glu^{802}Val$ mutant showed a k_{cat} reduction to

1.35 s^{-1}, while the Arg880Met completely inactivates the enzyme. In spite of this knowledge, the specific steps involved in the catalytic mechanism are still poorly understood. Currently, there are two main proposals: one suggested by Yamaguchi and coworkers [72], based on the kinetic data from mutations and X-ray structures; and a second one based on the QM and QM/MM work of Metz and Thiel [73, 74] (Figure 59).

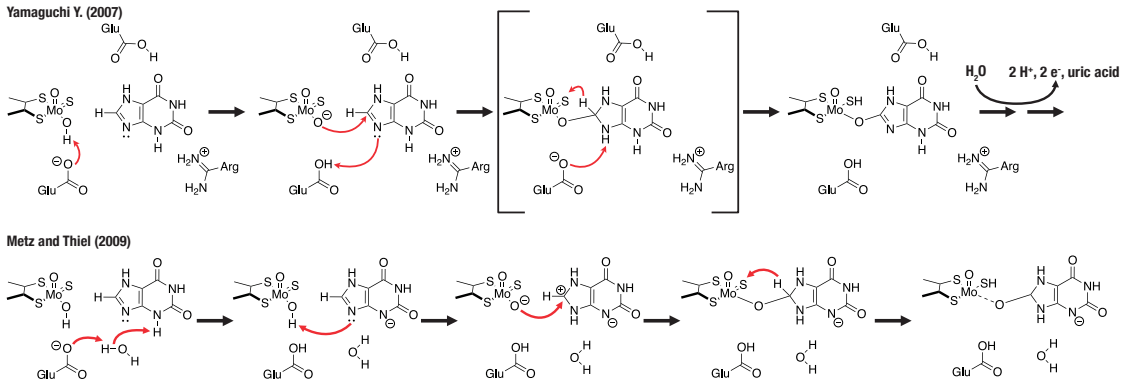


Figure 59. Proposals of Yamaguchi Y. (2007) [72] and Metz and Thiel (2009) [73, 74] for the catalytic mechanism of XO.

Although the Yamaguchi's proposal provides a simple explanation about how the mechanism occurs, it does not provide insights about the importance of a crystallographic water molecule placed between Glu^{1261} and Arg^{880}. Moreover, the mechanism does not explain the importance of Arg^{880} since its mutation completely inactivates the enzyme.

In the proposal of Metz and Thiel, there is no explanation for the release of the product (URC), and the role of Arg^{880} is not completely clarified.

Those are the main pitfalls of the previously available mechanism proposals that were considered in this work that aims to provide, for the first time, a complete and atomistic detailed description of the catalytic mechanism of XO.

2.1.2. Methodology

a. Model Preparation

Among the structures of XO available on PDB, the structure deposited under the code 3AMZ [387] was used in this study. Although this structure is from bovine, the structure was solved with a good resolution (2.1 Å) and is the only XO structure co-crystallized with the product of the reaction in the active site. In fact, the existing structures of the human XO have worst resolutions (2.6 Å [72] and 3.59 Å [388]) and do not include the substrate, product or analogues. Moreover, the sequence of the bovine form of XO has a similarity identity of 89.65% when compared with the human form (BLAST) [314, 315]. In particular, the structure of both active sites is perfectly superimposable (Figure 60).

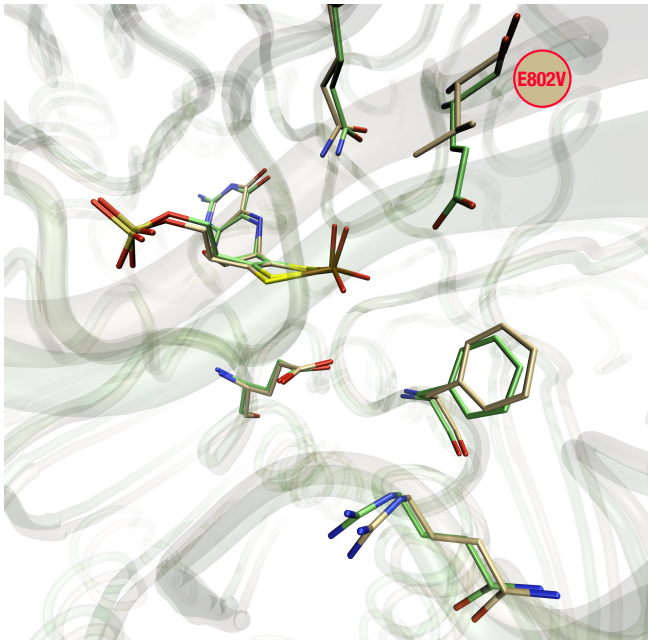


Figure 60. Overlapping of the structures of the active site residues of bovine and human forms of xanthine oxidase.

The gaps present in the PDB structure 3AMZ, 166-191, 520-537, and 1321-1325, were filled using homology modeling (SWISS-model software) [389-392]. The homology modeling was only used to fill the gaps in the PDB structure, accounting only about 4% of the entire protein sequence. Additionally, those gaps correspond to flexible loops in the enzyme's surface and are considerable far from the active site (Figure 61).

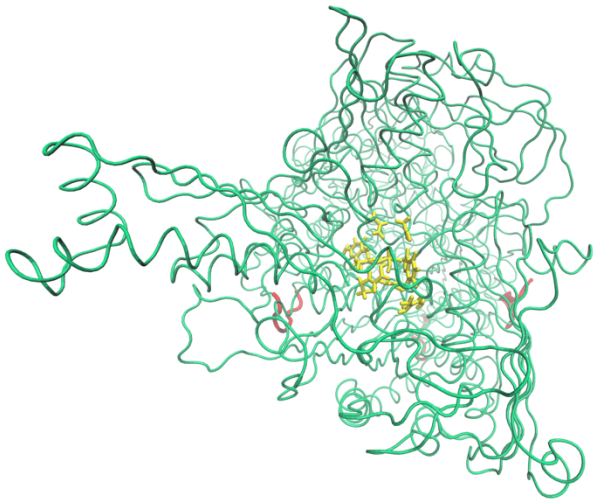


Figure 61. Representation of the XO structure with the three regions modeled by homology highlighted in red. The active site residues and pterin cofactor are represented in licorice colored by yellow.

The final structure contains the entire homodimer, where each subunit includes two [2Fe-2S] clusters, a flavin adenine nucleotide (FAD) molecule, a molybdopterin cofactor (Moco), and a substrate (XAN) molecule.

b. MD Simulations

A 22 ns molecular dynamics (MD) simulation was conducted to equilibrate the system and to provide a sample for choosing the initial structure for the study of the catalytic mechanism. All the energy minimizations and MD simulations were conducted using GAFF [148] and ff99SB [149] force fields. The molecular mechanics (MM) parameters for Moco were obtained from Ferreira P. and coworkers publication [393], whereas the parameters for [2Fe-2S], FAD, and XAN molecules were obtained through geometry optimization (HF/6-31G(d)) for charge determination and atomic topology assignment by antechamber software from AMBER12 package [277]. The protonation of all amino acid residues of the protein was assessed using PROPKA [275, 276] software.

Four sequential minimization steps were performed: (1) minimization of water molecules, (2) hydrogens, (3) all atoms except backbone, and finally (4) the entire system. The MD simulation was performed using PMEMD software from AMBER12 package [277], a solvation box of TIP3P [274] type of water molecules, cutting-off non-bond interactions at 10.0 Å, and an integration time-step of 1 fs at the temperature of 310.15 K and pressure of 1.0 bar using the Langevin's thermostat [394] and Berendsen's barostat [162].

c. QM/MM

The initial structure for the QM/MM calculations was retrieved as a representative structure of the most populated cluster from the MD simulation [248]. The model was built using molUP [316] plugin for VMD [292] software. The QM/MM model includes the full protein, the substrate (XAN), the Moco, and a 4.0 Å coat of water molecules around the full protein. The coat of water molecules mimics the solvent effect (water) around the enzyme, through intermolecular interactions and limiting the free movements of the solvent-exposed parts of the enzyme. The total system is composed of about 24,500 atoms that were treated using two different theoretical methods following a 2-layers subtractive ONIOM QM/MM approach [254, 366, 368]: the HL layer, which was calculated with DFT, and the LL layer, which was calculated with MM. The HL layer included the Moco, a water molecule (two in the last step), the XAN (except for the last step) molecule, and the Ala^{1043}, Ser^{1044}, Arg^{844}, Glu^{1225}, Glu^{766}, and Gln^{731} amino acid residues, accounting for 84 to 97 atoms (Figure 62). The remaining atoms of the system were included in the LL layer. Hydrogen atoms were used as link atoms to complete the valence of the bonds crossing between the two layers.

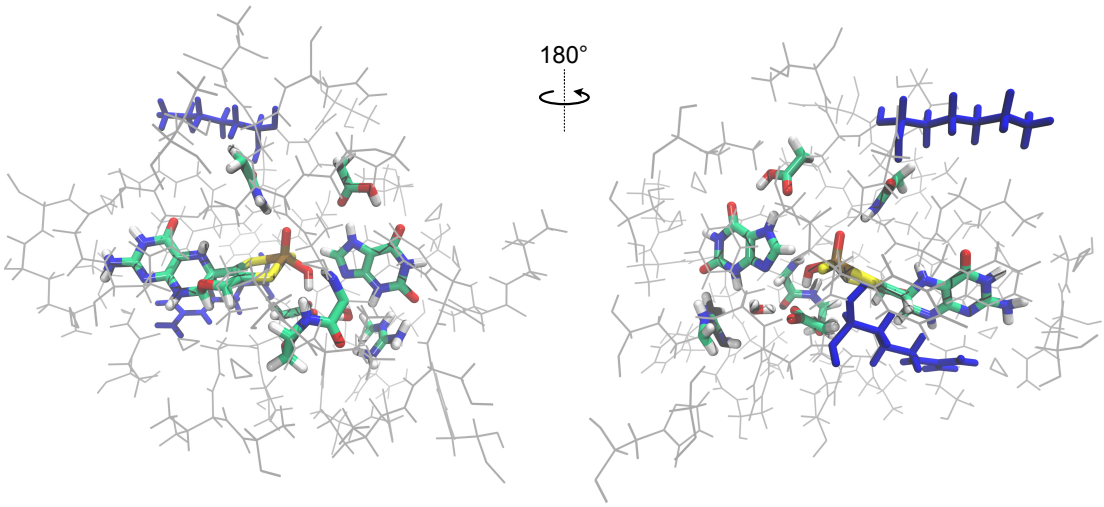


Figure 62. Representation of all the atoms included in the high-level (HL) layer treated using quantum mechanics (licorice colored by element). The residues that are up to 5 Å from the HL layer were represented in grey lines, and the charged residues within that shell are highlighted in blue. A 180 ° rotation across the z-axis was applied from the left to the right images.

The geometry optimization of the HL layer was performed using the B3LYP [330] functional (DFT). The 6-31G(d) [182, 282-285] basis set and LanL2dz [395] pseudo-potential (Mo atom) were employed for the geometry optimizations as available in Gaussian 09 [280]. During the catalytic mechanism, the oxidation state of Mo atom changes from +6 to +4, and back to +6 at the enzymatic turnover. The +6 state

corresponds to a Mo atom with no valence electrons (Mo(VI): [Kr] $4d^0$). In the +4 state, Mo atom assumes the Mo(IV): [Kr] $4d^2$ configuration, where two electrons are paired. Therefore, the spin multiplicity was set 1 for all the calculations.

The reactional space was explored through linear transit scans along with the reaction coordinates implicated in each step of the mechanism. Then, the TSs were optimized, using the structures of higher energy obtained with the previous scans. IRC calculations were conducted to obtain the reactant and product for each step. The TSs were confirmed by vibrational frequency calculations, resulting in a single imaginary frequency with the correct transition vectors assigned. The vibrational frequency calculations were also conducted for the minima, confirming no imaginary frequencies. The ZPE, thermal, and entropic energy corrections were estimated at 310.15 K and 1.0 bar during the frequency calculation of TS and related minima structures. The activation and reaction Gibbs free energies, for each step, were calculated through the difference between the computed Gibbs free energies of TS or product and the reactant, respectively.

The final energies were refined through SP energy calculations using a more complete basis set: 6-311++G(3df,3pd) [281, 284, 287-290].

All the calculations were accomplished using Gaussian 09 software [280], whereas the preparation of input files and analysis of results were made using molUP [316, 369] plugin for VMD [292].

2.1.3. Results and Discussion

Based on the calculations performed and in the comparison with the available experimental data, the catalytic mechanism of XO occurs in four steps involving the oxidation of XAN (Figure 63) to URC and subsequent enzymatic turnover. During the two initial steps, a XAN molecule is oxidized to an anionic form of URC (URT). In the third step, the URT is protonated before being released as URC. Finally, the Moco (Figure 63) is oxidized, and a water molecule becomes coordinated to Mo, closing the catalytic cycle, and making XO ready to catalyze a new cycle.

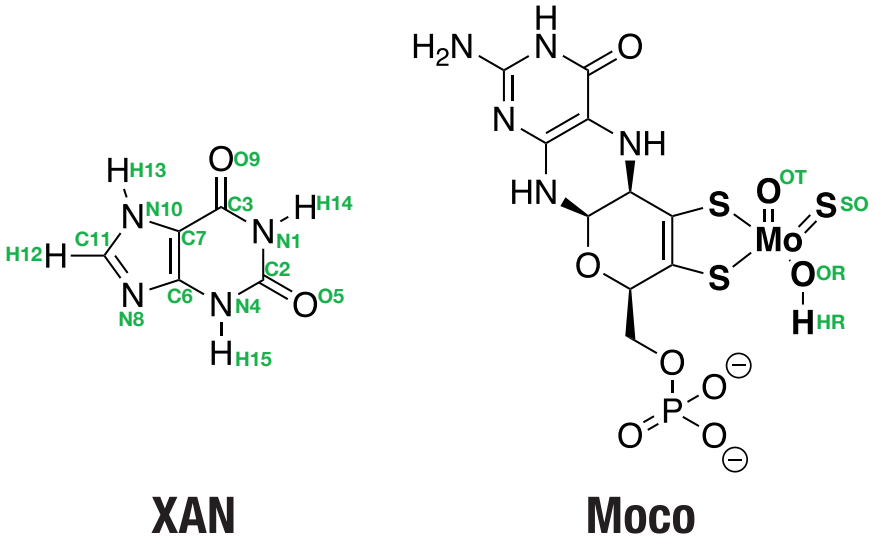


Figure 63. Wedge-dash representation of XAN and Moco molecules.

a. *Oxidation of XAN*

The oxidation of XAN to URT is the main stage of the entire catalytic mechanism, and it happens in two sequential steps.

In the first step, Glu^{1261} abstracts the proton (HR) from the hydroxyl group that is coordinated to Mo(VI) ion (R: 1.88 Å to P: 1.00 Å in the optimized minima), enhancing the nucleophilic character of the OR oxygen atom. Simultaneously, the OR oxygen makes a nucleophilic attack to carbon C11 of the XAN molecule (R: 2.59 Å; P: 1.46 Å). The negative charge, which is transferred to the XAN molecule, is stabilized by resonance through the conjugated π-system involving C11, N8, C7, C6, C3, and N4 atoms (Figure 64). The charge transfer from Glu^{1261} to XAN is corroborated by the analysis of the electrostatic potential surfaces represented in Figure 65.

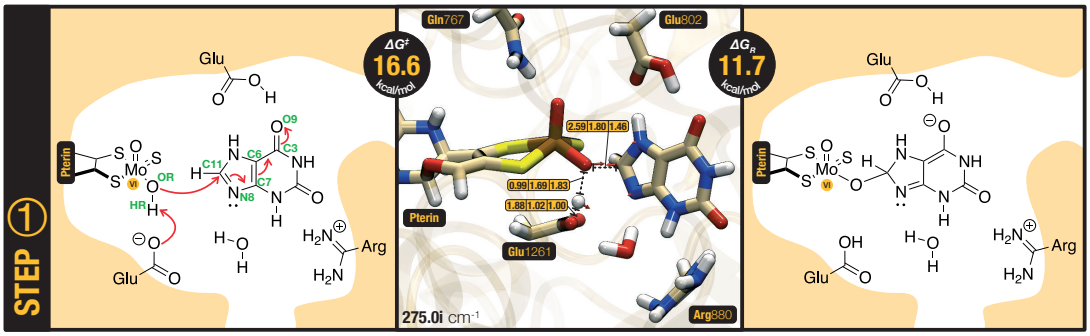


Figure 64. Tridimensional representation of the active site residues, Moco, and XAN molecule at the active site of XO in the TS of step 1. The red arrows represent the vectors assigned to the atoms' vibrations associated with the imaginary frequency of the TS structure. Relevant interatomic distances for reactant, TS, and product are labeled. Wedge-dash representation of the reaction and product of the reaction. The activation and reaction Gibbs free energies for this step, and imaginary frequency assigned to the TS are shown.

In this step, Glu^{802} plays a key role in the stabilization of the negative charge that is generated in the transition state structure. The strengthening of this interaction is supported by the decrease of the hydrogen bond distance from 2.31 Å to 1.86 Å in the optimized structures.

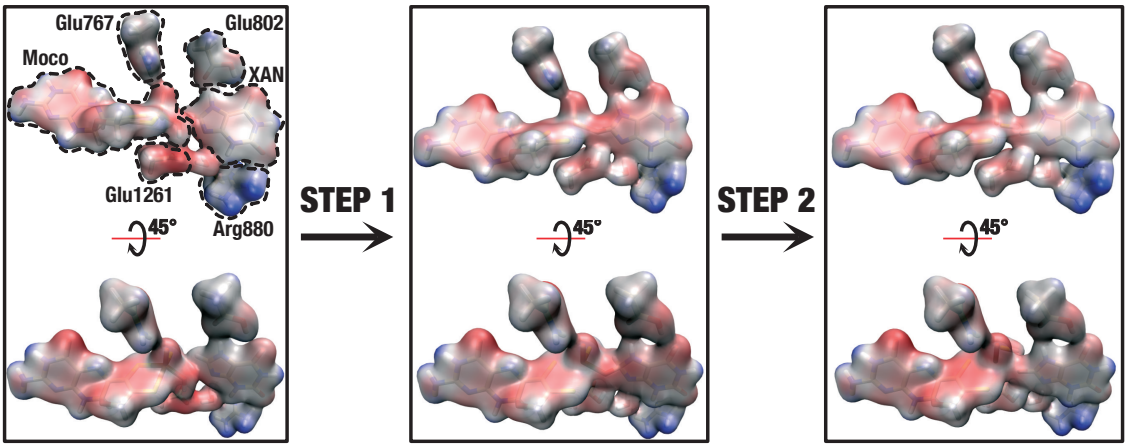


Figure 65. Electrostatic potential mapped in the surface of the molecules involved in the reaction of steps 1 and 2.

The important role played by Glu^{802} goes in line with the experimental result obtained by Yamaguchi and coworkers [72], since the Glu^{802}Val mutation decreases the kinetics of the reaction (k_{cat} (WT) = 18.3 s^{-1} vs. k_{cat} (Glu802Val) = 1.35 s^{-1}). Glu^{802} seems to be important to position the XAN molecule in the active site but also to stabilize the intermediate generated after the first step.

The TS of this reaction was confirmed by a single imaginary frequency at 275.0i cm^{-1}. This reaction requires an activation Gibbs free energy of 16.6 kcal/mol and is endergonic in 11.7 kcal/mol.

The second step of the oxidation of XAN involves the hydride transfer (H12) from carbon C11 to sulfur SO (R: 2.25 Å; P: 1.35 Å). The results suggest that this reaction is favored by the covalent bond established between oxygen OR and carbon C11 that decreases the distance between the hydrogen H12 to the sulfur SO that is coordinated to Moco. Consequently, the H12 hydride transfer is favored by the spatial proximity of the intermediate and the Moco, and the high electron density of the product of step 1, as represented in **Figure 65**.

Although the overall reaction involves the transfer of two electrons and one proton (hydride) from XAN to Moco, this process can occur through three different mechanisms: a hydride transfer, a hydrogen-one electron coupled transfer, or proton-two electrons coupled transfer.

The hydride transfer is the simplest case where the H12 and the two electrons are passed directly to sulfur SO. In this process, it is expected that H12 assumes a negative charge in the TS.

The hydrogen-one electron coupled transfer occurs when the H12 and one electron are passed to sulfur SO, and the second electron is transferred through the oxo-bridge established between Moco and XAN (Mo-OR-C11). In this case, the hydrogen H12 would assume a neutral charge during the TS.

Finally, the proton-two electrons coupled transfer involved the transfer of a proton (H12) directly to sulfur SO, and the two electrons are transferred to Moco via the oxo-bridge between Moco and XAN.

According to our calculations, the transference should follow a hydrogen-one electron coupled mechanism, since the H12 hydrogen assumes an almost neutral atomic charge (Merz-Singh-Kollman) during the TS (charge: -0.019 a.u.)

During this second step, the Mo ion is reduced from oxidation number Mo(VI) to Mo(IV), triggering the cleavage of the bond between the Mo ion and oxygen OR (R: 2.00 Å; P: 2.20 Å) and consequent release of the URT to the active site (**Figure 66**).

This second step is considerable faster than step 1, accounting for an activation Gibbs free energy of 5.2 kcal/mol, and it is also a thermodynamically more favorable reaction (ΔG_R = -20.1 kcal/mol). The energies involved in this step makes the reverse reaction not possible under physiological conditions (due to the high activation energies for the reverse reaction), ensuring the continuity of the catalytic process.

The TS of this reaction is characterized by one imaginary frequency, at 1028.7i cm^{-1}, corresponding to the vibration of the atoms involved in the reaction.

The results demonstrate that after this reaction, URT is no longer bonded to the Mo ion, but remains in the active site waiting for its protonation on N8 atom.

During this second step, the electron density that was lodged at XAN in the first step is now transferred to the Moco, where it is deeply stabilized (as can be seen on Figure 65), contributing to an extremely favorable reaction.

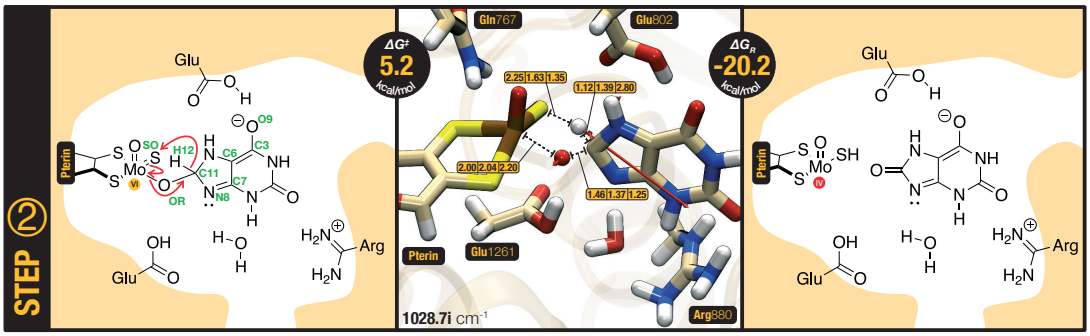


Figure 66. Tridimensional representation of the active site residues, Moco, and XAN molecule at the active site of XO in the TS of step 2. The red arrows represent the vectors assigned to the atoms' vibrations associated with the imaginary frequency of the TS structure. Relevant interatomic distances for reactant, TS, and product are labeled. Wedge-dash representation of the reaction and product of the reaction. The activation and reaction Gibbs free energies for this step, and imaginary frequency assigned to the TS are shown.

b. Protonation of the URT

The protonation of the URT precedes the release of the product (URC) from the active site. For this step, two hypotheses were considered and evaluated: (1) protonation by the Arg^{880} residue, and (2) protonation by the Glu^{1261} residue, both involving the same water molecule (Figure 67).

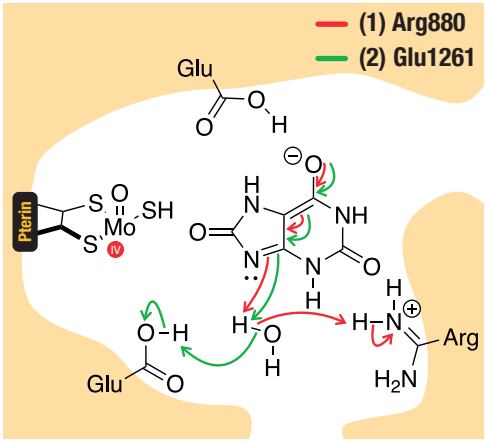


Figure 67. Representation of the two hypotheses for the catalytic mechanism of step 3. (Red) Protonation of URC molecule by (Red) Arg880 residue or (Green) Glu^{1261} residue.

According to the computed energies for the two hypotheses, the protonation of URT can be accomplished by Arg^{880} or Glu^{1261} residues. The activation Gibbs free energies for the protonation by Arg^{880} and Glu^{1261} are 4.6 kcal/mol and 1.8 kcal/mol, respectively. The main difference between both reactions is that the protonation promoted by Arg^{880} originates a more stable product (ΔG_R = -26.6 kcal/mol) when compared with the protonation mediated by Glu^{1261} (ΔG_R = +2.6 kcal/mol). In sum, the Glu^{1261} hypothesis is kinetically more favorable whereas the Arg^{880} one is thermodynamically more advantageous. However, we also observed that the product of the protonation by Glu^{1261} can be easily converted into the product of Arg^{880}-mediated protonation. Consequently, the reaction where the URT is protonated by Glu^{1261} can reach the same final product after a subsequent proton transfer from Glu^{1261} to Arg^{880}.

Although both hypotheses are possible and interconvertible, it is more likely that XO follows the shortest and thermodynamically more favorable process, in which Arg^{880} is involved in the protonation of the URC molecule.

In fact, this process, which involves a proton transfer from Arg^{880} to nitrogen N8 of URT mediated by a water molecule (as described in Figure 68), is very exergonic (ΔG_R = -26.6 kcal/mol), contributing (together with step 2) to the irreversibility of the XAN oxidation to URC. At this stage, URC becomes ready to be released from the active site as the final product of the reaction.

The simulations performed for this step provide an explanation for the importance of Arg^{880} for the mechanism, as suggested by the mutagenesis studies in which the Arg880Met mutation inactivates the enzyme [72]. The results of the present work show that Arg^{880} is the critical residue involved in the final protonation of URC and its release from the active site.

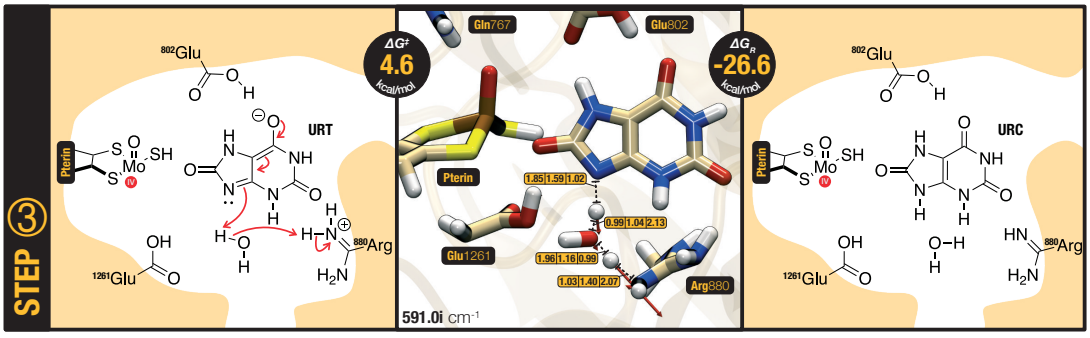


Figure 68. Tridimensional representation of the active site residues, Moco, and URC molecule at the active site of XO in the TS of step 3. The red arrows represent the vectors assigned to the atoms' vibrations associated with the imaginary frequency of the TS structure. Relevant interatomic distances for reactant, TS, and product are labeled. Wedge-dash representation of the reaction and product of the reaction. The activation and reaction Gibbs free energies for this step, and imaginary frequency assigned to the TS are shown.

c. *Enzymatic Turnover*

According to the calculations performed, at the end of step 3, the URC is completely formed and is released from the active site of the enzyme.

At this moment, based on experimental evidence, the Mo ion is oxidized from Mo(IV) to Mo(VI). The two resulting electrons are transferred over ~24 Å through two [2Fe-2S] clusters until they reach the FAD cofactor. FAD acts as an electron acceptor species, being reduced to $FADH_2$ [42, 76] (Figure 69).

Once the URC molecule is generated, the next step of the catalytic mechanism involves the oxidation of the cofactor and reaction of Moco with one water molecule. This can happen in several ways: i) the water molecule binds to Mo(IV) and only afterwards occurs the oxidation process (Mo(IV) to Mo(VI)); ii) first occurs the reduction of the cofactor from Mo(IV) to Mo(VI) and only after, the reaction with the water molecule takes place, or iii) the reduction of the cofactor occurs in a stepwise manner and involves the formation of transient Mo(V) species, that react with the water molecule.

Our calculations have shown that the water molecule can only coordinate with the Mo cofactor with Mo(VI) species. The same reaction cannot occur with Mo(IV) or Mo(V) species. This goes in line with the theoretical results obtained in a recent study involving hAOX that has the same cofactor and a similar active site [396].

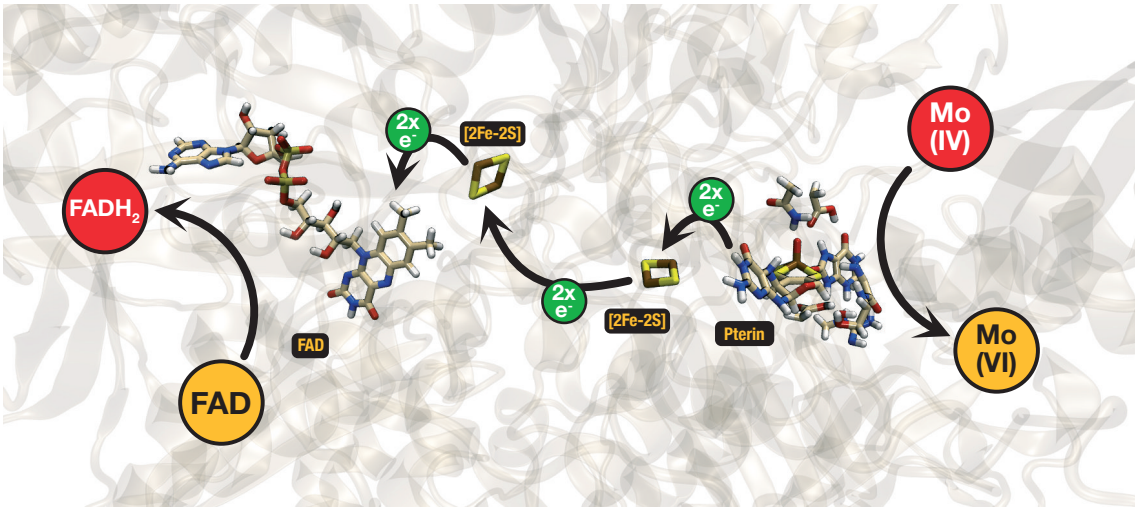


Figure 69. Tridimensional representation of the active site residues, the two [2Fe-2S] clusters, and the FAD cofactor. Scheme of the oxidation of the Mo ion, the consequent flow of electrons through the [2Fe-2S] clusters, and reduction of the FAD cofactor into $FADH_2$.

Once the Moco is oxidized again to Mo(VI), a water molecule can diffuse to the Moco's proximity, and bind the Mo ion. To assess this hypothesis, the movement of the water molecules in the MD simulations was analyzed. The MD simulations performed in this work show that this water molecule can be supplied by a water tunnel located between

Arg^{880} and Glu^{1261} residues (Figure 70). The tunnel has approximately 18 Å in length and allows water molecules to flow inside the active site. Consequently, water molecules can occupy the active site simultaneously with the release of the URC molecule.

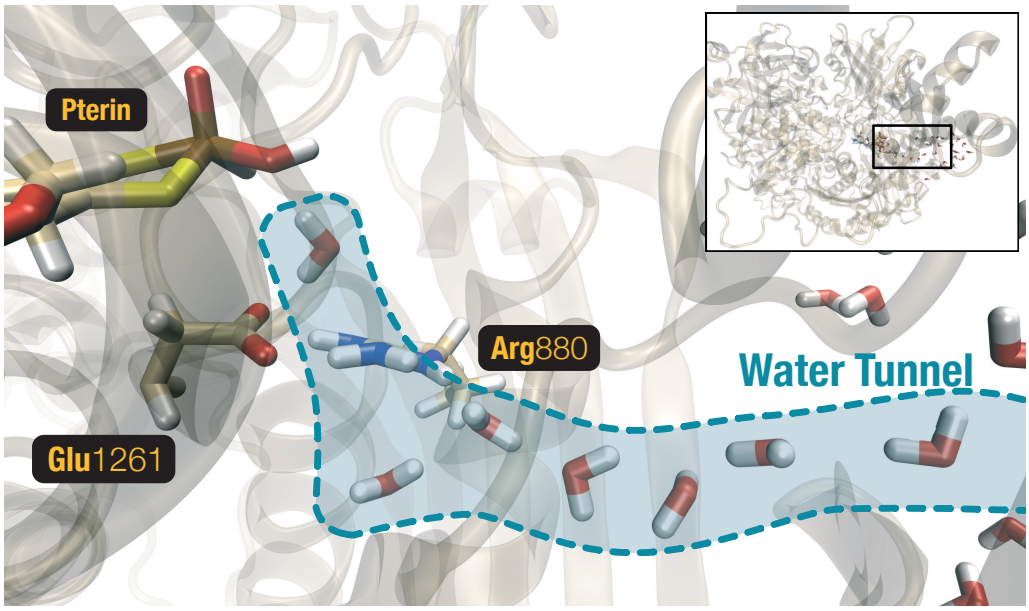


Figure 70. Tridimensional representation of the Moco, Glu^{1261}, Arg^{880}, and water molecules filling the protein tunnel from the solvent to the bottom region of the active site obtained from the MD simulations.

Once the water molecule fills the active site of XO, there are two possibilities for its reaction with the Mo ion: i) the water molecule binds directly the Mo with Glu^{1261} protonated or ii) the Glu^{1261} is deprotonated first and only afterwards one water molecule binds to the Mo ion. It worth to mention that nothing is known how the proton is removed from the Glu^{1261}, but it is believed to involve the water molecules present on the XO active site once URC leaves the active site.

The two hypotheses were explored. The calculated free energies shown that first scenario is, from the kinetic and thermodynamic point of view, less favored ($\Delta G^{\ddagger}$ = 17.6 kcal/mol; ΔG_R = 15.2 kcal/mol) when compared with the second scenario in which the reaction is spontaneously (ΔG_R = -3.3 kcal/mol). In the end of this reaction the water molecule becomes coordinated to the Mo ion (2.20 Å) and stabilized by one hydrogen bond provided by Glu^{1261} (1.50 Å) (Figure 71).

The formation of this interaction justifies the more favorable nature of the reaction when Glu^{1261} is deprotonated. The transient positive charge developed on the water molecule is compensated (-0.04 a.u. versus 0.09 a.u. in the products) with the increase of the negative charge on Moco cofactor. Consequently, both pterin's sulfur atoms become more distant from the Mo ion ([Pterin]S-Mo R: 2.35 Å and 2.34 Å; P: 2.43 Å and 2.43 Å) (Figure 71).

The results demonstrate that Glu^{1261} needs to lose its proton first in order to promote a spontaneous binding of the water molecule to the Mo ion. Afterward, the bonded water molecule loses one of its protons to a water molecule of the active site, completing the enzymatic turnover (Figure 71).

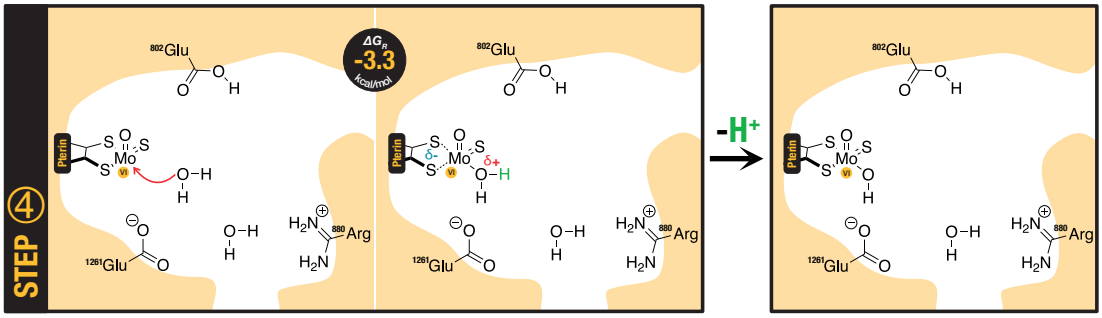


Figure 71. Wedge-dash representation of the reaction and product of step 4, and consequent deprotonation of the water molecule closing the enzymatic cycle. The reaction Gibbs free energy for this step is shown.

d. Energy Profile

The complete energy profile (Figure 72) obtained for the mechanism of XO shows two initial energy-demanding steps, where the oxo group is transferred to the XAN molecule, followed by the hydride transfer that impossibilities the reverse reaction. The hydride transfer results in a very stable product (ΔG_R = -20.2 kcal/mol) that makes the reverse reaction impossible to occur under biological conditions. Furthermore, the subsequent step is even more exergonic, agreeing with the fact that this reaction should be irreversible. This idea is supported by the fact that, in biological conditions, high concentrations of URC are accumulated and are not converted again in XAN. The high stabilization of the URC in the enzyme and the impairment of the reverse reaction explains how it was possible to obtain the crystallographic structure of such intermediate using a reduced form of the XO [387]. This stabilization allows URC to stay in the active site of XO enough time, prevents the reaction with water, and allow its isolation in X-ray crystallography.

In the last step, a water molecule coordinates the Mo ion closing the catalytic cycle.

The two steps involved in the oxidation of the XAN molecule are the rate-limiting steps of the entire mechanism. The barrier of 16.9 kcal/mol of the step 2 closely agrees with the predicted value (15.7 kcal/mol) based on the available kinetic parameters [72].

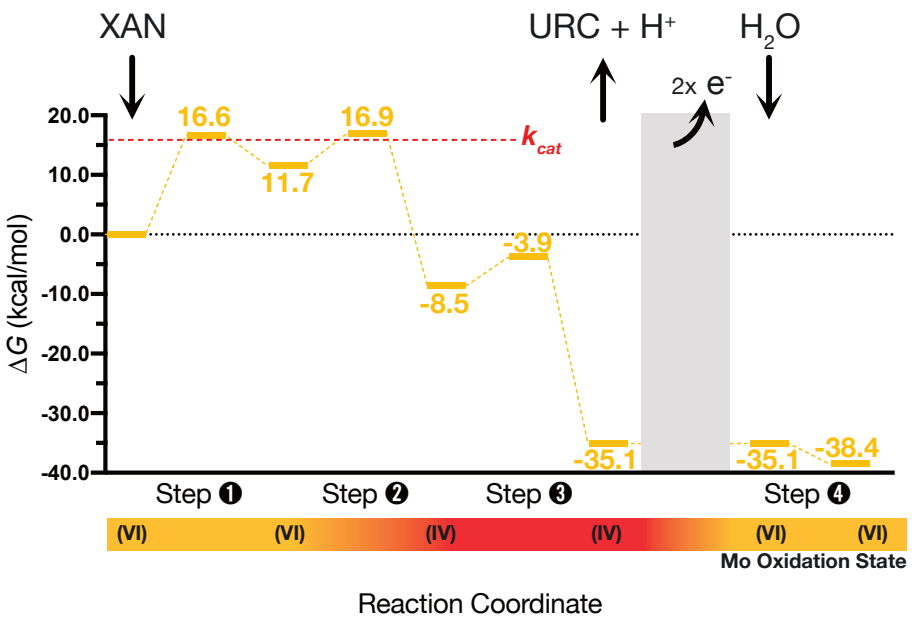


Figure 72. Complete energy profile (B3LYP/6-311++G(3df,3pd),LanL2DZ:ff99SB) for the catalysis of the oxidation of XAN into URC by XO.

2.1.4. Conclusion

In this work, the catalytic mechanism of the oxidation of XAN to URC by XO was studied using a computational ONIOM QM/MM approach.

The results provide, for the first time, a detailed and complete description of the entire catalytic mechanism of XO, completing previous works from Yamaguchi Y. [72], and Metz and Thiel [73, 74]. Moreover, it was possible to rationalize the role of Arg^{880} and Glu^{802} that have been previously shown as important by mutagenesis studies [72].

The complete catalytic mechanism is summarized in Figure 73. Initially, one XAN molecule enters the active site of XO and undergoes a nucleophilic attack by a hydroxyl group coordinated to Moco (Figure 73 - 1). This attack is enhanced by the Glu^{1261} when it abstracts the proton from the hydroxyl group. Then, a hydride is transferred from the previous tetrahedral intermediate to the sulfur atom coordinated to the Mo ion (Figure 73 - 2). At this moment, the Mo ion is reduced from Mo(VI) to Mo(IV), and the URT molecule is formed. Finally, Arg^{880} residue protonates the URT molecule triggering the release of the URC molecule from the active site to the solvent (Figure 73 - 3).

At this stage, FAD is reduced to $FADH_2$, oxidizing the Mo(IV) to its initial oxidation state of Mo(VI), promoting the deprotonation of the sulfur SO. This process is only possible

due to two [2Fe-2S] clusters that establish a connection between Moco and FAD cofactors for the electrons flow.

Afterward, a water molecule diffuses from a water tunnel settled between Glu^{1261} and Arg^{880} into the active site where it coordinates the Mo ion (Figure 73 - 4). Finally, the coordinated water molecule loses one proton to the solvent, and the enzymatic turnover is accomplished.

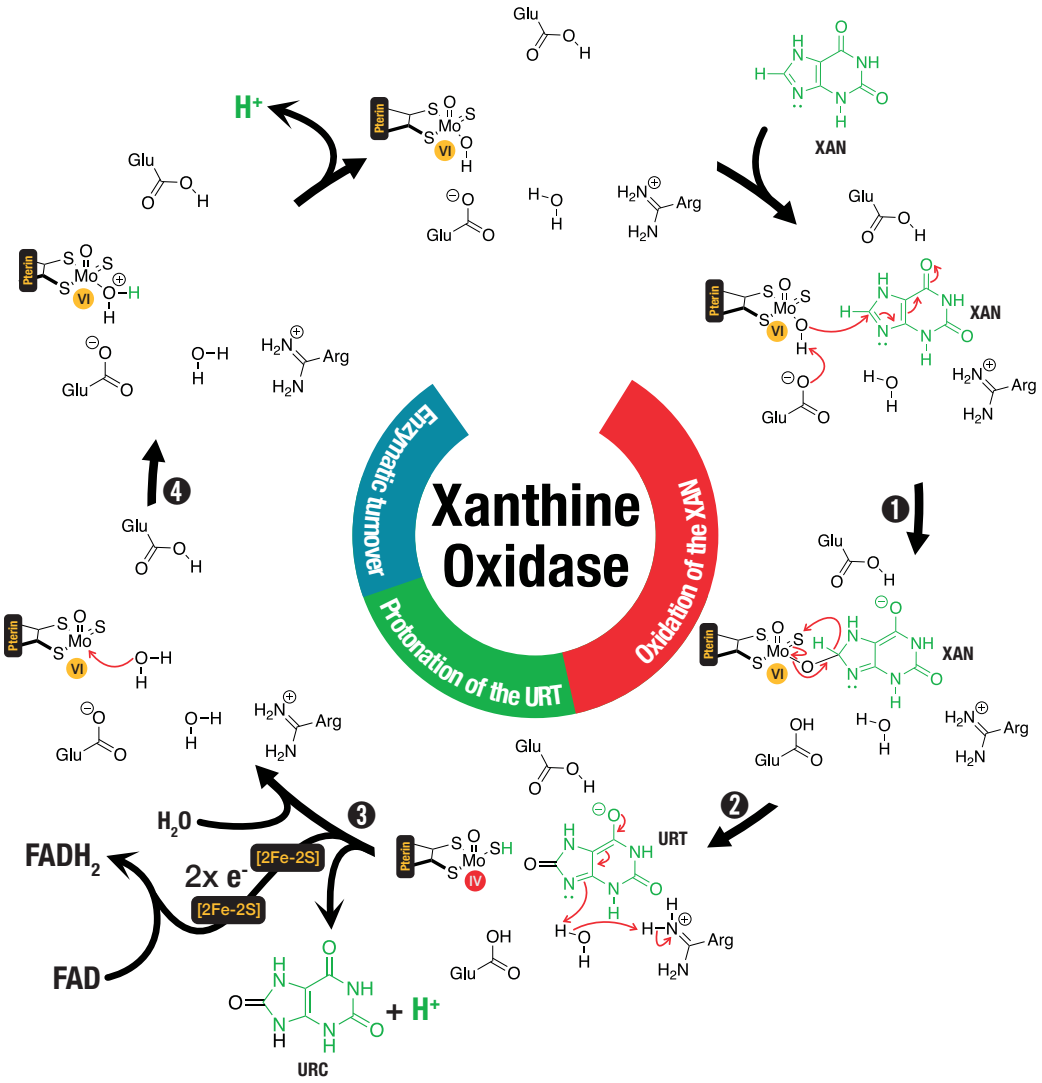


Figure 73. Overview of the catalytic mechanism of the XAN oxidation into URC by XO.

The description of the catalytic mechanism presented in this work provides new insights into the catalytic behavior of this type of enzymes and the TS structures can now be used to help the development of transition state analogues inhibitors targeting XO.

2.2. Nitrate and Nitrite Reductase Activity of Xanthine Oxidase

Although the main role of XO in the human body is oxidizing hypoxanthine and xanthine into xanthine and uric acid, respectively, some experiments have shown that XO possess reductase activity under certain conditions. In particular, XO has been reported as capable of reducing nitrate into nitrite and nitrite into nitric oxide. A QM/MM study was conducted to depict the catalytic mechanism behind these reactions, using the contextual knowledge acquired with the study of the natural mechanism of XO. This work was conducted in strict collaboration with Luísa B. Maia and Professor José J.G. Moura from the NOVA University, in Lisbon.

All the calculations were performed by Henrique S. Fernandes, as well as, the writing of the article, which was revised considering the contributions of all the authors. The work is currently is preparation for submission to the Inorganic Chemistry Frontiers journal.

This work was presented in the following conference:

- (Oral Communication) 24th European Nitrogen Cycle Meeting, 2019, Lisbon, Portugal

This page was intentionally left blank.

2.2.1. Introduction

As mentioned before, XO is a key enzyme in the elimination process of purines from the human body, through the catalysis of the two rate-limiting steps: oxidation of hypoxanthine into xanthine and xanthine into uric acid. The catalytic mechanism behind this last reaction was studied [195] and discussed in the previous section of this thesis. The results revealed that in each catalytic cycle, one molecule of hypoxanthine or xanthine is oxidized into xanthine or uric acid, respectively. Consequently, the Mo ion becomes reduced from Mo(VI) to Mo(IV). Since Mo ion is not the final electron acceptor, the pair of electrons must travel through two 2Fe-2S clusters to a FAD molecule where it becomes reduced to $FADH_2$. (**Figure 74** - Green) Finally, the $FADH_2$ reduces one oxygen molecule, which is the final electron acceptor of the entire mechanism. The process tends to occur flawlessly under normal cell conditions. However, in certain stress conditions, as is the case of an ischemic stroke, the tissues become deprived of oxygen.

In this situation, cells undergo a cascade of events leading to an increased oxidative stress and an increase of the environment acidity (low pH) (**Figure 74** - Red). In such conditions, the normal catalytic cycle of XO cannot be completed due to the lack of oxygen that impairs the oxidation of the $FADH_2$ and therefore the enzymatic turnover. Experimental studies have shown that, under these stress conditions, XO can show nitrate and nitrite reductase activity, catalyzing the reduction of nitrate into nitrite and nitrite into nitric oxide, respectively [75-78]. The NO production, which occurs in these situations, provides an important protective role against the ischemic damage [79-83].

This alternative mechanism of XO appears as secondary mechanism that can be trigger in the absence of oxygen. Despite the initial finding of a reductase activity of XO [77, 84-92], there are no further knowledge about the mechanism behind these reactions.

In this work, the nitrate and nitrite reductase activity of XO were studied, with an atomistic level of detail, using a computational QM/MM approach:

1. Reduction of Nitrate (NO_3^-) to Nitrite (NO_2^-);
2. Reduction of Nitrite (NO_2^-) to Nitric Oxide (•NO).

The conversion of nitrate into nitrite involves a two-electron reduction of nitrate, oxidizing the Mo ion directly from Mo(IV) to Mo(VI). Therefore, per each hypoxanthine or xanthine oxidation cycle, one nitrate molecule can be reduced to complete the alternative enzymatic turnover (**Figure 74**).

On the other hand, since nitrite is converted into nitric oxide through a one-electron reduction, two nitrite molecules are required to be reduced in order to achieve the same outcome (Figure 74).

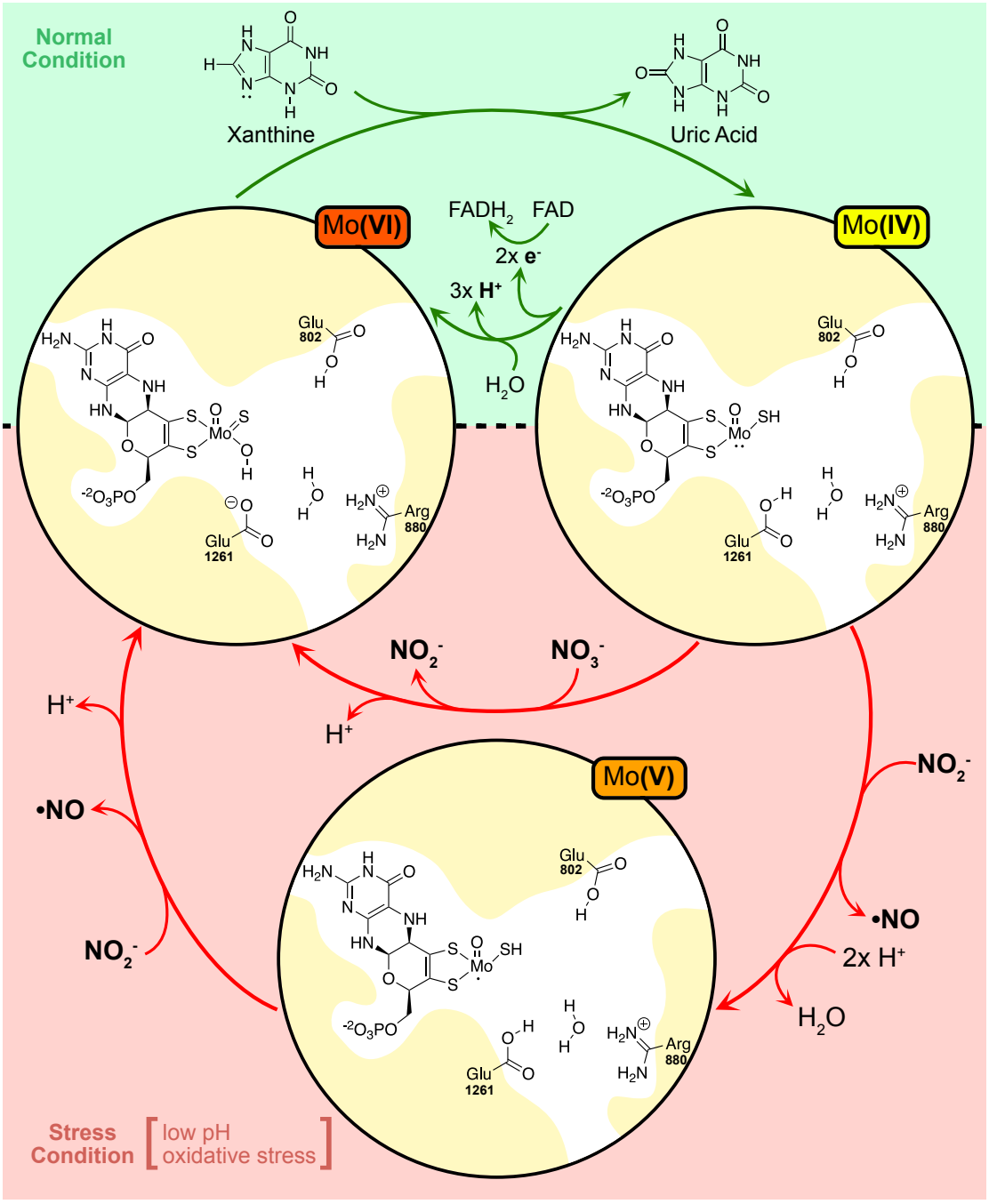


Figure 74. Schematic representation of the catalytic cycle of XO under (green) normal and (red) stress conditions.

2.1.1. Methodology

a. Model Preparation

The QM/MM model used in this study was based on the QM/MM model used in the study of the catalytic mechanism presented in the previous work (Chapter C - 2.1. page 181). In addition, the same computational protocol was employed in the study of this mechanism.

Therefore, the product of the third step of the catalytic mechanism of XAN oxidation (Figure 68) was used as starting structure for the study of this mechanism. To do so, one molecule of nitrite or nitrate was added to the active site, depending on the reaction that was studied (nitrite or nitrate reductase activity of XO).

b. QM/MM

The QM/MM model was prepared and adjusted using molUP [316] plugin for VMD [292] software, as happened for the study of the catalytic mechanism of XAN oxidation by XO. Accordingly, the model includes the full monomer of XO, the Moco, and a 4.0 Å coat of water molecules that surround the protein. The system is composed of approximately 24,500 atoms that were treated using two different theoretical methods following a 2-layers subtractive ONIOM QM/MM approach [254, 366, 368]: the HL layer, which was calculated with DFT, and the LL layer, which was calculated with MM. The HL layer included the Moco, three water molecules, a nitrate or nitrite molecule, and the Arg^{844}, Glu^{1225}, Glu^{766}, Ser^{1164}, Arg^{913} (backbone), Gly^{913} (backbone), and Gln^{731} amino acid residues. The remaining atoms of the system were included in the LL layer. Hydrogen atoms were used as link atoms to complete the valence of the bonds crossing between the two layers.

The geometry optimization of the HL layer was performed using the B3LYP [330] functional, the 6-31G(d) [182, 282-285] basis set, and the LanL2dz [395] pseudo-potential for Mo atom. The geometry optimizations, frequency calculations, and single-point energy calculations were performed using Gaussian 09 [280].

The exploration of the reactional space and respective characterization of all the minima and TSs follows exactly the same approach of the study of the catalytic mechanism of the XAN oxidation reaction by XO (Chapter C - 2.1.2. page 185)

2.1.2. Results and Discussion

a. Reduction of Nitrate (NO_3^-) to Nitrite (NO_2^-)

The reduction of nitrate or nitrite requires a prior reduction of the Mo cofactor of XO. In normal conditions, this reduction is accomplished through the oxidation of a hypoxanthine or xanthine molecule into xanthine or uric acid, respectively. Thus, the product of step 3 of the catalytic mechanism of either of these substrates (hypoxanthine or xanthine) is the starting point for the nitrate/nitrite reduction. Therefore, and as referred before, the initial structure was built based on the product obtained on the previous work [195].

From that structure, the uric acid molecule (product) was removed, and PROPKA [276] software was used to predict the protonation state of all the titratable residues of the protein. Particular attention was given to the active site residues Glu^{802}, Glu^{1261}, and Arg^{880}, that play a key role in catalysis. The computed pK_a values for these three residues were 9.49, 6.14, and 9.91, respectively. Accordingly, Arg^{880} was set positively charged, Glu^{802} was set neutral, and some doubts arose regarding the protonation state of Glu^{1261}. According to the experimental evidence [77], the efficiency of XO to oxidize nitrite is maximized at pH 6.3. Since the pK_a value calculated for Glu^{1261} is very close to the ideal pH conditions for the reaction, both protonation states of this residue were tested.

The theoretical calculations showed that the binding of nitrate or nitrite to the Mo ion is impossible when Glu^{1261} is negatively charged. This means that nitrate and nitrite binding to the Moco is favored by the neutral form of Glu^{1261}. Therefore, Glu^{1261} was set neutral for the study of the reductase mechanism of XO.

In the case of nitrate reduction, a nitrate molecule was placed in the active site of XO, and the structure was optimized (Figure 75 A). Afterward, the distance between the O1 of nitrate and MO ion was incrementally deceased (from 3.99 Å in the R to 2.27 Å in the P), simulating the coordination of nitrate to Mo ion. The calculations revealed a spontaneous process, exergonic by 6.9 kcal/mol (no transition state could be identified and characterized). During this step, Glu^{802} guides the nitrate molecule to the Mo ion through a hydrogen bond that it establishes with the O3 of nitrate (R: 1.76 Å - P: 1.68 Å). A nearby water molecule is also involved in this process, establishing a hydrogen bond with O1 of nitrate in the reactant (1.72 Å) that progressively fades away and is replaced by another hydrogen bond with the O2 of nitrate (Figure 75 B).

Once the nitrate molecule becomes bonded to Mo ion, occurs the cleavage of the O1-N bond, from which results the formation of nitrite.

The computed structures revealed a very favorable mechanism for the reduction of nitrate into nitrite, with an activation and reaction Gibbs-free energies of 9.7 kcal/mol and -40.0 kcal/mol, respectively. The high exergonicity of this reaction makes the reverse step impossible, under physiological conditions, suggesting that nitrite cannot be oxidized to nitrate by XO.

At the TS (**Figure 75** C), the O1-N bond cleavage occurs at the bond length of 1.53 Å. The structure of the TS was validated by vibration frequency analysis, revealing a single imaginary frequency (634.5i cm^{-1}) related to the vibration of the O1 and N atoms of the nitrate molecule.

Once nitrite is formed (**Figure 75** D), it can easily diffuse out of the active site. At the same time, it is proposed that the proton from the SO atom is lost to the solvent, and the O1 atom abstracts the proton from the Glu^{1261} (**Figure 75** E). In the end of this process, the enzymatic turnover takes place and XO is ready to oxidize another hypoxanthine or xanthine molecule (**Figure 74**).

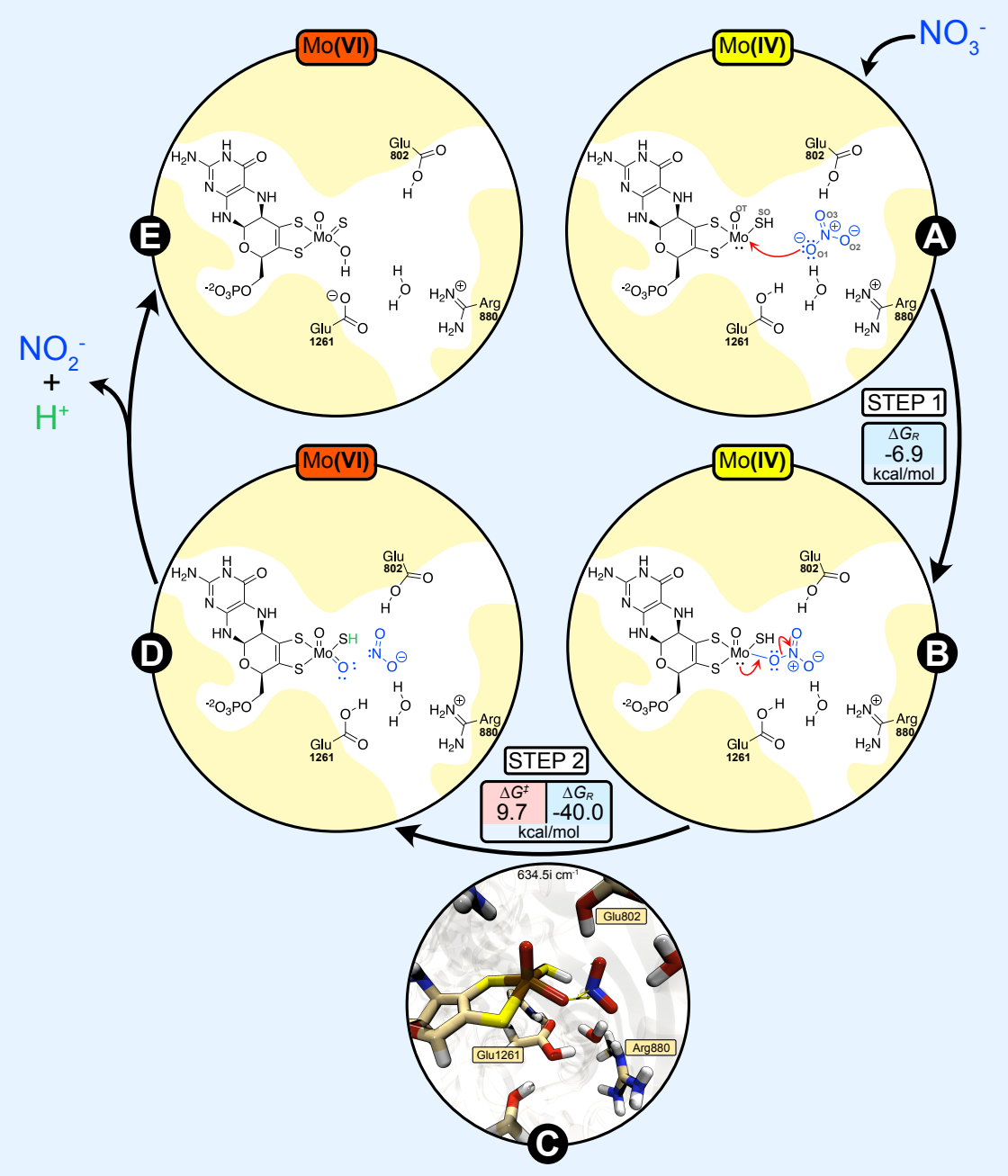


Figure 75. Schematic representation of the mechanism of reduction of nitrate to nitrite by XO. The wedge-dash schemes illustrate the proposed mechanism based on the calculations of this work. The computed Gibbs-free energies and an illustration of the TS of step 2 are also included.

b. Reduction of Nitrite (NO_2^-) to Nitric Oxide (•NO)

Similarly, to the nitrate reduction study, the starting point for the calculations was the same, but the nitrate molecule was replaced by a nitrite molecule.

1st Stage

After the initial structure optimization, it was observed that the nitrite molecule is anchored in the active site by two hydrogen bonds established with a nearby water molecule and Glu^{1261} (Figure 76 A).

The coordination of the nitrite molecule to the Mo ion was tested by a relaxed PES scan where the O1-Mo distance was incrementally decreased from 3.95 Å to 2.11 Å. The coordination occurs spontaneously with a reaction Gibbs-free energy of -1.7 kcal/mol (Figure 76 B') (no transition state was possible to characterize or identify for this reaction).

The next step (Step 2 of the 1st Stage) involves the rupture of the O1-N bond from which results a singlet or a triplet system. Theoretical calculations of the singlet state structure revealed an impossible reaction under physiological conditions, since the activation barrier overcame the 35 kcal/mol, originating a highly unstable product (Figure 77). In contrast, the reaction in the triplet state is very favorable requiring an activation Gibbs-free energy of 7.7 kcal/mol. Moreover, in the triplet state, the reaction is almost thermoneutral, accounting for a reaction Gibbs-free energy of +1.6 kcal/mol. Since both PESs are independent and no intersections between them are observed (Figure 77), it is expected that the reactant of step 2 is more stable in the triplet state than the singlet state.

In sum, the nitrite molecule binds to the Mo ion in the singlet state (Figure 76 B'), and then the resulting structure easily falls into the triplet state (Figure 76 B), which is substantially more stable (by 25.9 kcal/mol). The Glu^{802} residue, which was establishing a hydrogen bond with the nitrite molecule during the step 1, leaves this interaction when the spin multiplicity changes from 1 to 3. In the triplet state, the Glu^{802} establishes a hydrogen bond with the OT atom of the Moco (1.89 Å). These interactions are kept during the entire step 2.

During this process also occurs the oxidation of the Mo ion, from Mo(IV) to Mo(V), (Figure 76 C).

The TS structure of this step was validated by vibrational frequency calculation revealing a single imaginary frequency of 364.1 cm^{-1}, involving the O1 and N atoms. The O1-N

bond cleaves at the bond length of 1.98 Å, after a 0.57 Å displacement compared to the reactant.

In the end, a nitric oxide molecule is ready to be released from the active site, and the Mo ion rests in its Mo(V) oxidation state (**Figure 76** D). Due to the highly acidic environment, two protons from the solvent can reach the O1 atom and protonate it. In that situation, the O1 atom can be released as a water molecule, liberating a position for a new nitrite molecule (**Figure 76** E).

Once new nitrite molecule enters the active site, XO is ready to catalyze another cycle of the one-electron reduction of nitrite to nitric oxide, completing this salvage pathway for the standard enzymatic turnover.

Analogously, the binding of the nitrite to the Mo ion was assessed using the same relaxed scan approach. As previously observed, this process was also spontaneous and exergonic in 19.4 kcal/mol (**Figure 76** F).

Afterwards, Mo(V) ion is oxidized to Mo(VI), triggering the formation of nitric oxide. This reaction is substantially slower than all the previous ones, requiring an activation Gibbs-free energy of 19.4 kcal/mol. Moreover, the product is also less stabilized than the previous ones, since the reaction Gibbs-free energy of this reaction is +13.9 kcal/mol.

The TS was validated by frequency calculations, revealing a single imaginary frequency (436.2i cm^{-1}) corresponding to the vibration of O1 and N atoms involved in this step (**Figure 76** G).

Once nitric oxide is formed, and Mo ion is oxidized to Mo(VI) (**Figure 76** H), the nitric oxide can diffuse out of the active site, the proton from the SO atom is lost to the solvent, and the Glu^{1261} protonates the oxo group that is coordinated to the Mo ion (**Figure 76** I). In this stage, the enzyme completes its enzymatic turnover and becomes ready to oxidize a new hypoxanthine or xanthine molecule (**Figure 74**).

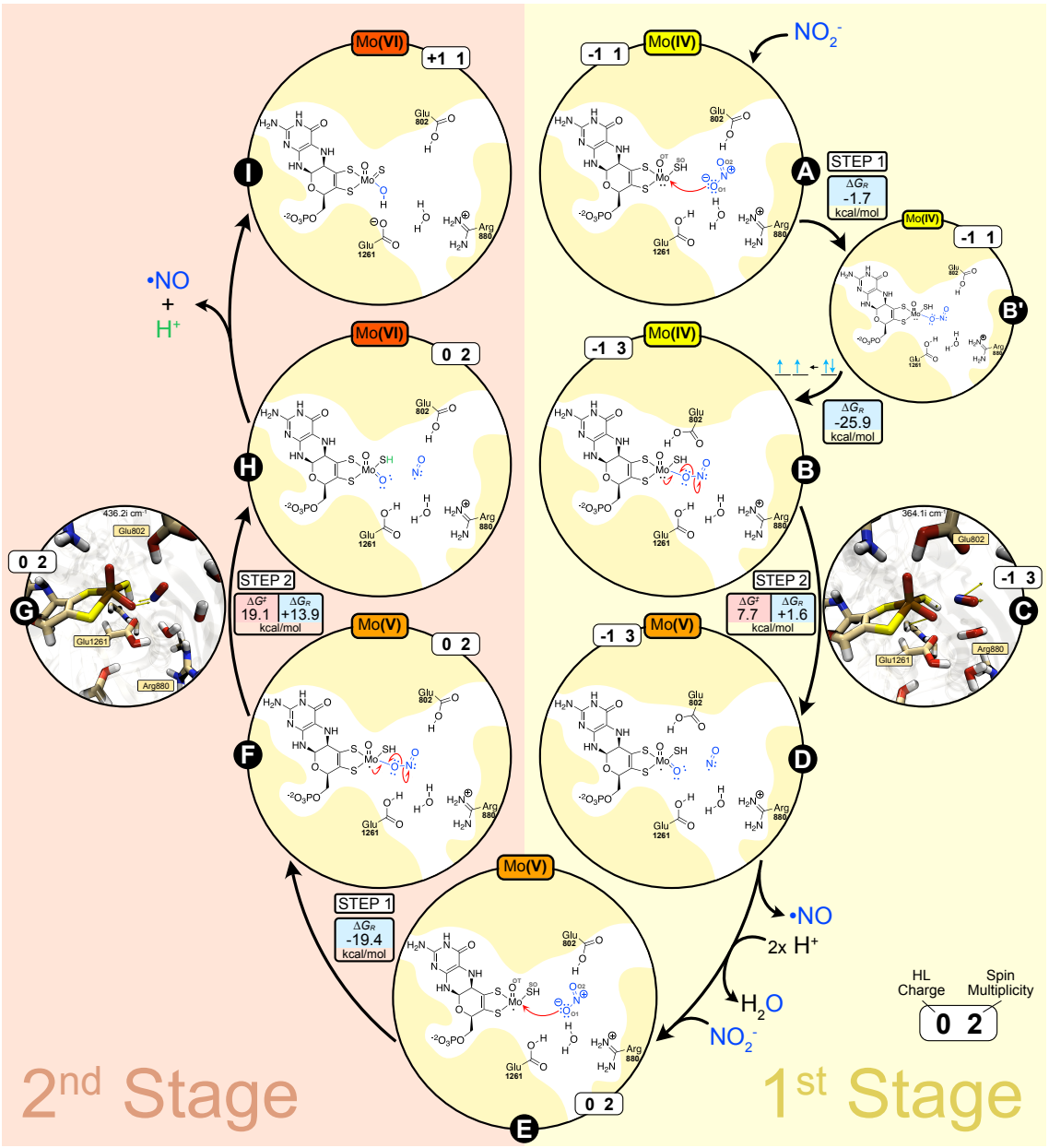


Figure 76. Schematic representation of the mechanism of reduction of nitrite to nitric oxide by XO. Each stage corresponds to a single-electron reduction of nitrite to nitric oxide. The wedge-dash schemes illustrate the proposed mechanism based on the calculations of this work. The computed Gibbs-free energies and an illustration of the TS of step 2 are also included. Colored circles enclose the total charge of the HL and different colors correspond to different spin multiplicities (1, 2, and 3).

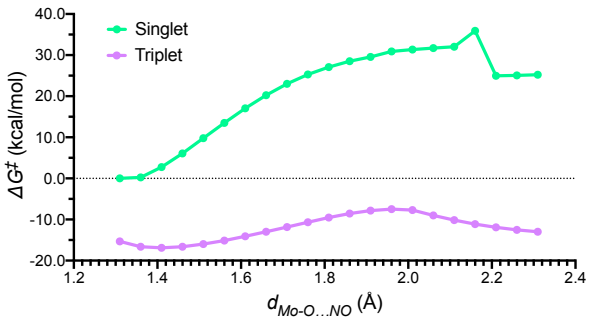


Figure 77. Energy profile for the PES scan for the homolytic rupture of O-NO bond of nitrite (Step 2 of the 1st Cycle), on the singlet and triplet stages.

c. *Energy Profile*

As shown, XO can actively reduce nitrate into nitrite and nitrite into nitric oxide, when the cellular supply of oxygen is low. According to the computed energies (Figure 78), the nitrate reduction into nitrite is irreversible catalyzed by XO. Thus, an increase in the amount of nitrite does not cause an increase in the nitrate pool by the action of XO.

Contrary, XO seems to be able to revert the reduction of nitrite into nitric oxide. During the first stage, the reverse reaction appears to be very difficult to occur due to the high stabilization of the triplet state (-25.9 kcal/mol), but the second reduction cycle can be reversible. The second stage starts with Mo(V), and the reduction of nitrite is slower when compared with the first cycle, but it does not originate a very stabilized product, allowing the reverse reaction to occur. According to these results, it can be expected that XO mechanism can be stalled with Mo(V) species when oxygen is not available, and a huge amount of nitric oxide is supplied.

Although the second stage of nitrite reduction is reversible, it is not expected that XO can actively oxidize nitric oxide to produce nitrite. This reaction would occur until an early stage where all XO enzymes would have the Mo ion in its V oxidation state, and, therefore, not able to continue reacting.

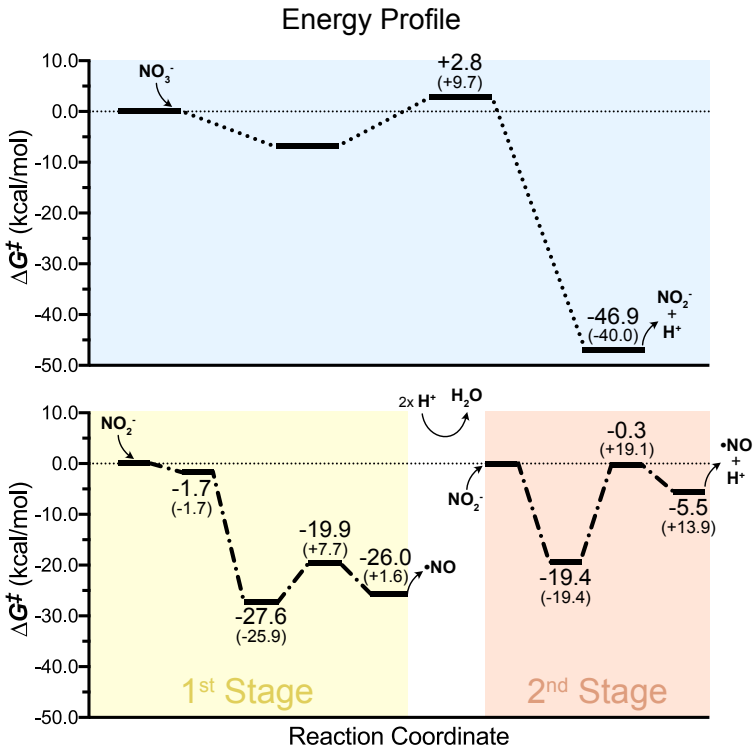


Figure 78. Complete energy profile (B3LYP/6-311++G(3df,3pd)//B3LYP/6-31G(d,p),LanL2DZ:ff99SB) for the reduction of (Top) nitrate into nitrite, and (Bottom) nitrite into nitric oxide by XO.

2.1.3. Conclusion

This work describes the complete catalytic mechanism of the nitrate/nitrite reductase activity of XO, studied using a computational approach. Although the main catalytic activity of XO is the oxidation of hypoxanthine and xanthine into xanthine and uric acid, respectively, experimental evidence show that XO is capable of reducing nitrate and nitrite [77, 84-92].

This experimental evidence is now corroborated, for the first time, by a computational study. These results provide additional insights about the molecular and electronic features of this reaction.

The nitrate/nitrite reductase activity of XO acts as a rescue pathway to allow the oxidation of the Moco cofactor (Mo(IV) to Mo(VI)). In some pathological conditions, cells can be deprived of oxygen, impairing the oxidation of $FADH_2$ and, consequently, oxidation of the Moco cofactor. If no alternative mechanism existed, XO would not be able to complete the enzymatic turnover and would be stuck at the Mo(IV) state.

The results show that XO can accomplish the alternative oxidation of its Moco cofactor through two independent mechanisms: i) XO can either reduce one nitrate molecule or ii) two nitrite molecules per catalytic cycle.

The reduction of nitrate into nitrite comprises a three parts mechanism. Initially, one nitrate molecule diffuses to the active site, and spontaneously binds to the Mo ion (**Figure 79** 1). Then, the O-N bond of nitrate is rapidly broken, originating a nitrite molecule (**Figure 79** 2), which is highly stabilized (-40.0 kcal/mol). Finally, the product of the reaction (nitrite) is released as well as the proton from the SH group of Moco cofactor, completing the enzymatic turnover (**Figure 79** 3).

Alternatively, a molecule of nitrite can enter the active site of XO and binds the Mo ion (**Figure 79** 4). This process is also spontaneous, and it is followed by the homolytic rupture of the O-N bond of the substrate. Consequently, Mo(IV) ion is oxidized to Mo(V), and a nitric oxide molecule is formed (**Figure 79** 5). Afterwards, the nitric oxide molecule is released, and the oxygen atom that is still bonded to Mo ion is protonated and released as a water molecule (**Figure 79** 6).

In this stage, the Moco cofactor is not fully oxidized yet, and the Mo ion assumes the V oxidation state. Consequently, a new nitrite molecule needs to enter the active site to be reduced into nitric oxide again (**Figure 79** 7). The reduction reaction follows the same mechanism, originating a new nitric oxide molecule (**Figure 79** 8). Finally, the catalytic cycle is closed by the of the SH group proton of the Moco cofactor and the resulting nitric oxide molecule (**Figure 79** 9).

Thus, XO can complete its standard enzymatic turnover using an alternative pathway. However, the role of this mechanism goes beyond the importance of achieving the enzymatic turnover. This mechanism is also a source of nitric oxide, which is a very relevant signaling molecule for the living organisms [79-83], including humans. In this context, the role of nitric oxide as a vasodilator is crucial to help in situations where the blood, and consequently, the oxygen, supplies are compromised. Since the oxygen deprivation triggers the alternative reductase activity of XO and XO can produce nitric oxide by this pathway, this enzyme assumes an important role in pathological situations as ischemic strokes [79-81].

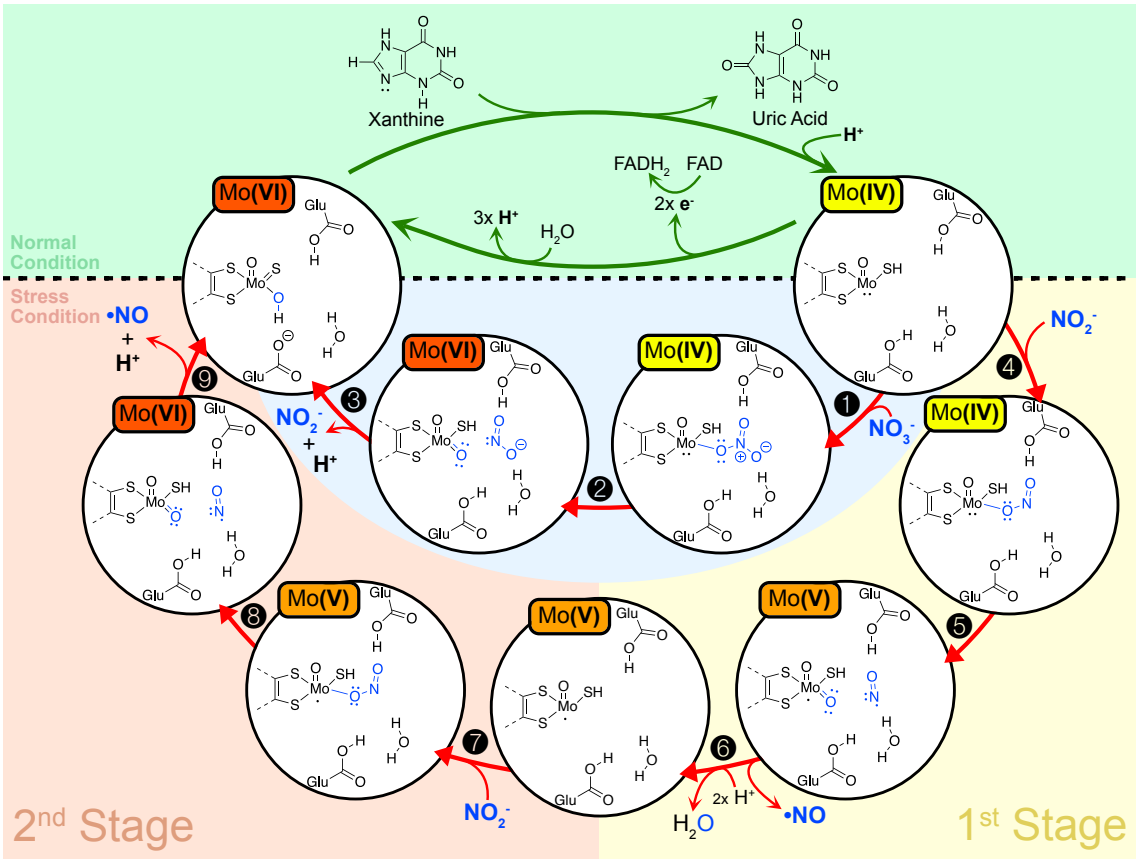


Figure 79. Overview of the (green arrows) natural catalytic mechanism of XO and the (red arrows) catalytic mechanism for the reduction of nitrate and nitrite into nitrite and nitric oxide, respectively.

The details provided by this work can contribute to a better understanding of the XO role in pathological events such as ischemic strokes. This can also be important to look again at the side-effects of XO's inhibitors that are usually used for treating gout. If XO plays, in fact, an indispensable role in the body-response to ischemic situations, those inhibitors could compromise that response.

3. Cofactor-independent Enzymes

This page was intentionally left blank.

3.1. The Catalytic Mechanism of the Main Protease of SARS-CoV-2

The SARS-CoV-2 virus was responsible for starting a pandemic with a deep impact on the daily life of millions of people worldwide. The governments of several countries, including Portugal, were obliged to set highly restricted measures to prevent the disease's spreading, including lockdown. Rapidly, the entire scientific community was mobilized to study and understand the structure and function of SARS-CoV-2, aiming intelligent manner to stop this fatal pandemic. This mobilization resulted in more than 23 thousand scientific publications[11], according to the Scopus repository. Some of that data was essential for developing different vaccines, which aim to immunize the population against the virus and are currently being administered to the population.

Even though the lockdown safety policies compromised, to some extension, the research work, additional efforts were required to carry out the study of the catalytic mechanism of the main protease of the SARS-CoV-2 virus. There are already several commercially available drugs targeting the main protease (M^{pro}) of several viruses. Therefore, the leading goal of this study was to provide a detailed description of the catalytic mechanism that could help in the development of antiviral drugs against the M^{pro} of the SARS-CoV-2 virus.

All the calculations were performed by Henrique S. Fernandes, as well as the writing of the article, which was revised considering the contributions of all the authors. The publication is currently in peer-review in the Molecular Diversity Journal.

[11] Publications with the term "sars-cov-2" in title, abstract, and/or keywords.

This page was intentionally left blank.

3.1.1. Introduction

The main protease of SARS-CoV-2, also known as SARS-CoV-2 M^{pro}, is an essential enzyme for the replication of the virus causing the COVID-19 pandemic. M^{pro} is the enzyme responsible for cleaving the translated polyprotein chain in 11 different regions (A-K) [397] (Figure 80). This means that effective inhibiting M^{pro} activity can stall the viral replication cycle [398]. Since no homology or sequence similarity is found between the virus and human proteases [399], M^{pro} is seen as an appealing drug target for the development of new antiviral drugs to fight COVID-19, one among other strategies currently in development [400-408].

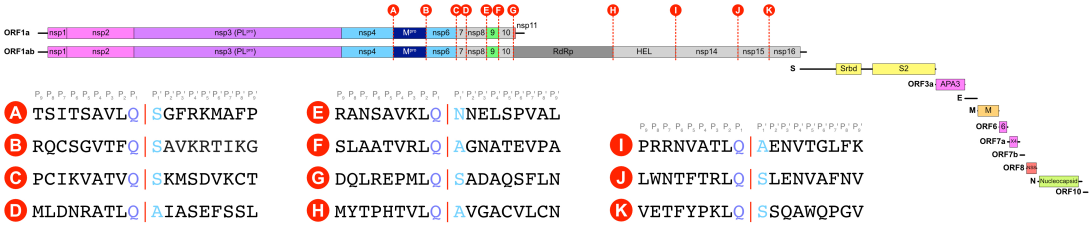


Figure 80. The complete genome sequence of SARS-CoV-2 virus. [397] Each red label (A-K) represents the respective protein cleavage regions catalyzed by M^{pro}. Each cleavage region is also shown with associated flanking amino acid sequences.

Similar therapeutic approaches were used with success, in the past, in the treatment of other viral infections, as for example, the now commercially available inhibitors Tipranavir, Darunavir, Amprenavir, Lopinavir, among others [409], which target the HIV-1 protease. For this reason and along with the development of vaccines, the search for antiviral compounds continues to be on the top list of the pharmaceutical companies and research laboratories that seek to discover a compound capable of treating and impairing the dissemination of COVID-19 worldwide [410-412].

In record time, a valuable amount of knowledge was obtained regarding SARS-CoV-2 virus and, in particular, about M^{pro}. M^{pro} was described to be located between nsp4 and nsp5 segments of the polypeptide, which means that Mpro is able to cleave itself (Figure 80 A and B). Several X-ray structures of M^{pro} have also been deposited on the Protein Databank (PDB), revealing its three-dimensional structure [400, 413-419].

M^{pro} is a homodimer whose subunits have independent active sites that are exposed to the solvent. Although the active sites are apparently independent, the monomer was shown to be inactive [399].

The solvent exposure of the active site favors the binding of the peptide substrates to the active site. Also, it ensures the availability of water molecules that are required for the completion of the catalytic mechanism (Figure 81).

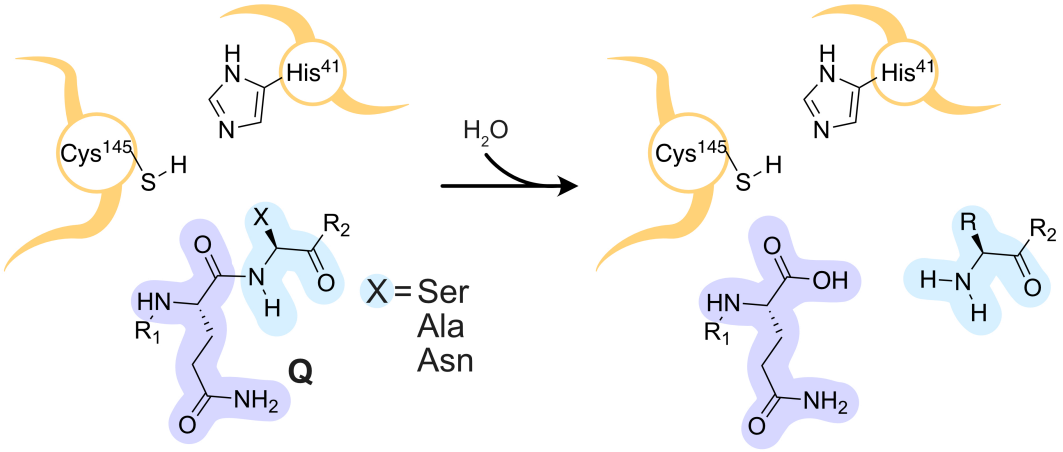


Figure 81. The general reaction catalyzed by M^{pro}. "R1" and "R2" represent the remaining part of the substrate (peptide). "X" represents the sidechain of the possible amino acid residues (Ser, Ala, or Asn).

From the sequence analysis of all the peptides that are cleaved by M^{pro}, a well-stablished pattern is observed. The peptide bond that is cleaved by M^{pro} always occurs between a Gln residue (P1) and a Ser, Ala, or Asn residues (P1'). Moreover, the key Gln residue of the substrate peptide is always preceded by a bulky non-polar residue: Val, Leu, or Phe (P2) (Figure 80).

In addition, several residues from the active site (Thr^{24}, Thr^{26}, Asn^{142}, Asn^{143}, His^{163}, His^{164}, Glu^{166}, Gln^{189}, and Thr^{190}) make hydrogen bonds with the peptide. These residues behave as anchors that allow the recognition of the peptide as a substrate and favor the correct alignment and orientation of the peptide inside the active site. Thus, close contact between the target Gln residue and the catalytic Cys^{145} residue is ensured. According to the literature, the P4-P1' range comprises the most important residues for the recognition and anchoring of the peptide [420].

The reaction is catalyzed by a Cys-His dyad. The His^{41} activates the nucleophilic character of Cys^{145} that covalently attaches the peptide substrate to the enzyme. This reaction intermediate is then hydrolyzed, and the product of the reaction released to the active site (Figure 81).

So far, there are a panoply of studies devoted to the study of the M^{pro} [415, 421-423], including the catalytic mechanism [424, 425]. In this work, the full catalytic mechanism of the SARS-CoV-2 M^{pro} was studied using an ONIOM QM/MM methodology. The obtained results provide valuable insights about the structures of the transition state and linked minima for each step of the mechanism. Moreover, a complete atomistic description of the mechanism was obtained.

3.1.2. Methodology

a. Model Preparation

The QM/MM model used in this work was prepared based on the X-ray structure of M^{pro} published by Z. Jin *et al.*, and deposited on PDB under the 6LU7 code [413]. The structure of M^{pro} was completely solved and all the 306 amino acids of each subunit of the homodimer are available. Moreover, a synthetic molecule was co-crystallized with the model as an analogue of the substrate. Since the main goal of this work was studying the mechanism of one possible natural substrate, the structure of the synthetic construct was used as template to model the following peptide sequence: ILMVF**Q**SGACN.

The protonation state of the amino acid residues of the entire system was assessed using the PROPKA software [275, 276]. Finally, the resulting structure was solvated using a TIP3P type of water molecules [274, 426] (orthogonal box with 12.0 Å of minimum distance), and 8 sodium ions were added to neutralize the system. Then, the solvated system was minimized through a set of four sequential minimization protocols (non-bonded interactions cutoff: 10 Å), using AMBER software [147, 148, 155, 277] and ff14SB force field [150]: 2500 steps for minimization of the water molecules; 2500 steps for minimization of all hydrogen atoms; 2500 steps for minimization of all system except backbone atoms; 10000 steps for minimization of all system. The resulting minimized structure was used to conduct a 60 ns molecular dynamics (MD) simulation: 50 + 50 ps NVT equilibration; 60 ns NPT production. The MD simulation was performed using a 2 fs integration step and the SHAKE algorithm [427] to constraint hydrogen atoms.

The minimized structure was used as starting point to build and conduct the QM/MM calculations. This strategy was employed to approximate the starting QM/MM model to the structure obtained experimentally.

b. QM/MM Calculations

The molUP plugin [316] for VMD [292] software, available through the VMD Store [369], was used to load the minimized structure obtained in the previous step and prepare the Gaussian09 [280] input files.

All the homodimer structure, the substrate peptide, and a 5 Å coat of water molecules were selected for the QM/MM model. The Gln and Ser residues of the substrate, the Cys^{145} residue, the sidechain of His^{41} residue, the backbone of Gly^{143} and Ser^{144} residues, and a water molecule (steps 3 and 4) were included in the HL layer. In total, 69/72 atoms were included in the layer treated using quantum mechanics. The remaining atoms of the systems were considered using MM. A shell of 10 Å from the HL layer were

kept free during the geometry optimizations. The relaxed coordinate scans and the geometry optimization of minima and TSs were performed using the following scheme: B3LYP[245-247]/6-31G(d,p)[182, 281-285]:ff14SB[150]. Hydrogen atoms were used as link-atoms to cap all the bond that span across the QM-MM boundary. The calculations were performed considering the electrostatic embedding approach as implemented in Gaussian09 software. All the minima and TS were confirmed and validated by vibrational frequency analysis. IRC calculations were conducted to obtain the reactant and product for each step. The ZPE, thermal, and entropic energy corrections were estimated at 310.15 K and 1.0 bar during the frequency calculation of TS and related minima structures.

The final energies were refined through SP energy calculations with the following scheme: DLPNO-CCSD(T) [232]/CBS[319]:ff14SB. ORCA software (v4.2.1) [250] was used to perform the SP energy calculations.

3.1.3. Results and Discussion

According to the calculations performed in this study, the catalytic mechanism of SARS-CoV-2 M^{pro} occurs through four sequential steps. After an initial recognition process, where specific amino acid sequences of the substrate are recognized by the protease binding pocket, a key Gln residue from the substrate becomes aligned with the Cys^{145}-His^{41} dyad, and the proteolytic reaction begins.

In the first step, the proton from Cys^{145} is transferred to His^{41}. This reaction is characterized by a TS with an imaginary frequency at 655.5i cm^{-1} and a $\Delta G^{\ddagger}$ of 2.4 kcal/mol, and it is exergonic in 0.7 kcal/mol. In the product of the reaction, His^{41} becomes positively charged, Cys^{145} becomes anionic and gets close to the C1 carbon of the peptide substrate (SX---C1: 3.06 Å at reactant versus 2.30 Å at product). (**Figure 82** Step 1; **Figure 83**). This step is very important to improve the nucleophilic nature of Cys^{145}, and therefore, to enhance its reactivity towards the peptide substrate. The nucleophilic nature of the SX of Cys^{145} is corroborated by its atomic charge variation from -0.10 a.u. to -0.25 a.u. during this first step. In addition, the approximation of SX to C1 causes a decrease in the atomic charge of the electrophilic carbon of the substrate that passes from 0.65 a.u. to 0.47 a.u.. This shift of the electron density triggers the next step of the mechanism. This first step originates an ion pair where both His^{41} and Cys^{145} become charged, as reported in the interesting publication of Ramos-Guzmán *et al.* [425]. However, the activation and reaction energies reported in that publication are slightly higher than the ones obtained in this work. The difference could be assigned to the different peptide substrates used in both studies. Particularly, the P2 position is occupied by a Phe residue in our study versus a Leu residue, which can lead to a greater stabilization of His^{41} through a π-stacking interaction. Consequently, the positively charged His^{41} residue is more stabilized, and the reactions become faster and marginally exergonic. Contrary to our results and the Ramos-Guzmán *et al.* [425] work, the pioneer mechanism proposal by Swiderek *et al.* [424] suggests that the deprotonation of Cys^{145} is concomitant with the nucleophilic attack to the substrate with much higher activation and reactions energies. However, the study was conducted with a substrate analogue with a bulky non-amino acid molecular species at the P1' position, which in our study is occupied by Ser residue. The residue that occupies P1' position has been shown to play an important role in the stabilization of the ion pair either in SARS-CoV orthologs [428, 429] and SARS-CoV-2 protease [425]. Moreover, it seems that the hydroxyl group of Ser residue can stabilize the ion pair in about 1.4 kcal/mol [425] when compared with Gly or Ala residues.

The second step of the catalytic process comprises the nucleophilic attack of the anionic Cys[145] to C1 carbon of the peptide substrate (Figure 82 Step 2). At the same time, one proton migrates from the cationic His[41] to the N1 atom of the substrate. This reaction is also fast, presenting a $\Delta G^{\ddagger}$ of 5.1 kcal/mol. However, this second step, where the peptide bond is broken, is considerably more exergonic than the previous one, with a ΔG_R of -4.6 kcal/mol. The TS of this step was characterized and confirmed by an imaginary frequency at 1064.3i cm^{-1}. In the product of this reaction, a thioester-enzyme adduct is obtained, which covalently attaches the substrate to the enzyme through the Cys[145] residue. In this process occurs a change in the hybridization of C1 carbon of the substrate from sp^2 to sp^3, which is corroborated by the improper torsion angle, Cα-C1-N1-O1, that changes from 170.5° to 131.0°.

The activation Gibbs free energy obtained in this work is considerably lower than the one predicted by Ramos-Guzmán *et al.* [425] (5.1 kcal/mol vs ~10 kcal/mol). Nonetheless, the reaction product, after these two initial steps, shares similar energy in both studies (-5.3 kcal/mol vs ~-2 kcal/mol). The considerably faster reactions reported by our calculations could be a consequence of the Phe residue at the position P1' as reported in step 1 and also the fact that, in our calculations, the backbone of critical Gly[143] and Ser[144] residues were included in the QM region since they are known to be important for stabilizing the oxo group of the substrate.

In the next steps of the catalytic process (Figure 82 Step 3 and 4), the hydrolysis of thioester-enzyme adduct takes place, requiring the presence of one water molecule, which is widely available in the active site due to its exposure to the solvent. In the optimized structure of the reactants, the water molecule becomes very close to the Cys[145] residue, and it is stabilized by three hydrogen bonds provide by His[41] (HW1: 1.79 Å), a carbonyl group of the substrate (HW2: 1.92 Å), as well as the amine group of the released peptide (OW: 2.14 Å).

After the appropriate accommodation of the water molecule within the active site, it makes a nucleophilic attack to C1 carbon of the thiol-enzyme adduct (OW---C1: 3.15 Å at reactant versus 1.59 Å at product). Simultaneously, the His[41] residue abstracts a proton from the water molecule (Nε---HW1: 1.79 Å at reactant versus 1.05 Å) (Figure 82 Step 3). At this point, the His41 residue becomes again positively charged, showing that it is also important for this second stage of the mechanism. The reaction is characterized by a TS with an imaginary frequency at 575.5i cm^{-1} and a $\Delta G^{\ddagger}$ of 13.5 kcal/mol. This reaction is also endergonic in 7.7 kcal/mol, which can be justified by the formation of two punctual charges in the carbonyl group of the substrate (O1) and in the His[41] residue. Although the product is less stable, this anionic tetrahedral intermediate (Figure 83) is

deeply stabilized by three hydrogen bonds provided by the NH groups of Gly^{143}, Ser^{144}, and Cys^{145} to the negatively charged O1 atom. These three residues of the active site form what is commonly known as the oxyanion hole and have an important role in the catalytic process of M^{pro}. This important interaction was not included in the QM part of the previous study of the mechanisms with the substrate, possibly leading to higher activation and reaction energies when compared with the results obtained in this work (13.5 kcal/mol vs 15.6 kcal/mol [425] and 7.7 kcal/mol vs ~12 kcal/mol [425]). When we repeated the calculations with a smaller QM region, excluding the backbone of Gly^{143}, Ser^{144}, and Cys^{145} residues, the estimated activation and reaction energies increase to 21.0 and 20.2 kcal/mol, respectively. This evidence confirms the importance of these residues in the QM region to properly stabilize the oxyanion formed in this step of the mechanism.

In the fourth and last step of the catalytic process, a reversion of the first step occurs, and the proton (HW1) is transferred from His^{41} to Cys^{145} (HW1---SX: 2.61 Å at reactant versus 1.38 Å at product). At the same time, the bond between the Cys^{145} residue and the C1 carbon is cleaved, which releases the reaction product to the active site (SX---C1: 1.59 Å at reactant versus 3.23 Å at product) (**Figure 82** Step 4). The TS was also verified and present an imaginary frequency at 617.0i cm^{-1}. This last step occurs when the activation Gibbs free energy barrier of 6.7 kcal/mol is overcome. This last step is also endergonic, this time, in 4.1 kcal/mol.

At the end of this step, the initial protonation state of all the active site residues is restored, and the enzyme is ready to catalyze a new proteolytic reaction.

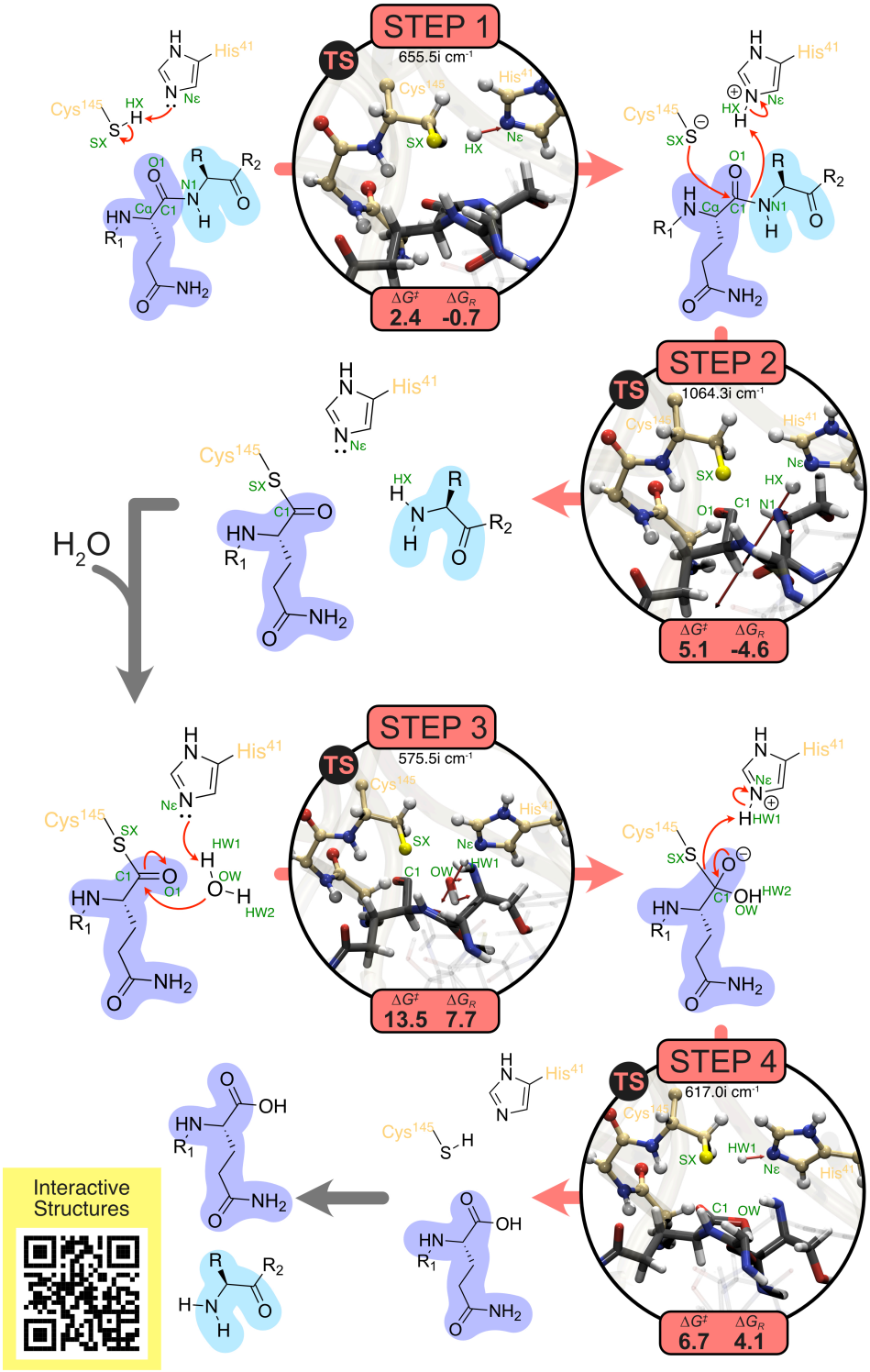


Figure 82. Representation of all the 4 steps of the catalytic mechanism of SARS-CoV-2 M^{pro}. The activation and reaction Gibbs free energies are expressed in kcal/mol units. The eigenvectors of each TS structures are shown as red arrows.

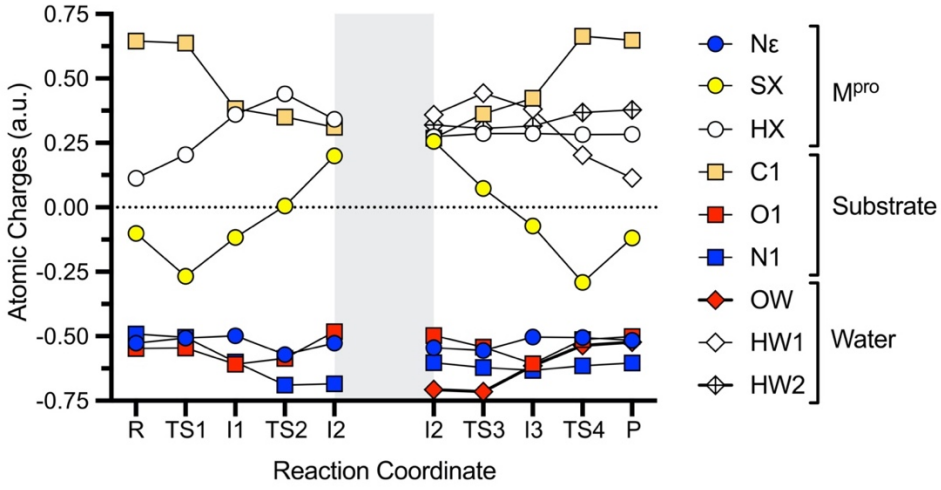


Figure 83. Charge fluctuation of the key atoms over the catalytic mechanism of the M^{pro}. Circles, squares, and lozenges identify atoms belonging to the catalytic dyad, the substrate, and the water molecule, respectively. Markers are colored by element: nitrogen (blue), sulfur (yellow), hydrogen (white), carbon (orange), and oxygen (red).

The Gibbs free energy profile of the proteolytic mechanism catalyzed by M^{pro} is shown in Figure 84. In this profile, two well-defined stages can be identified. In the first part of the mechanism, the peptide substrate becomes covalently attached to the enzyme, resulting in the thioester-enzyme adduct and the release of the N-terminal part of the substrate. In the second stage occurs the hydrolysis of the thioester-enzyme adduct through two sequential steps and from which results the release of the product of the reaction, i.e., the C-terminal part of the peptide. This reaction is the rate-limiting step of the full catalytic process and requires total activation Gibbs free energy of 14.4 kcal/mol (Figure 84).

The available kinetic data for Gln-Ser SARS-CoV protease shows turnover numbers between 0.847 and 12.2 s^{-1} [428, 429], corresponding to approximately 16.0 to 17.6 kcal/mol, considering the transition state theory. The maximum activation barrier calculated in this work agrees with the recent evidence that the catalytic efficiency of SARS-CoV-2 M^{pro} is slightly higher than the SARS-CoV ortholog [413], which suggests a faster reaction and, consequently, lower activation energy for the rate-limiting step.

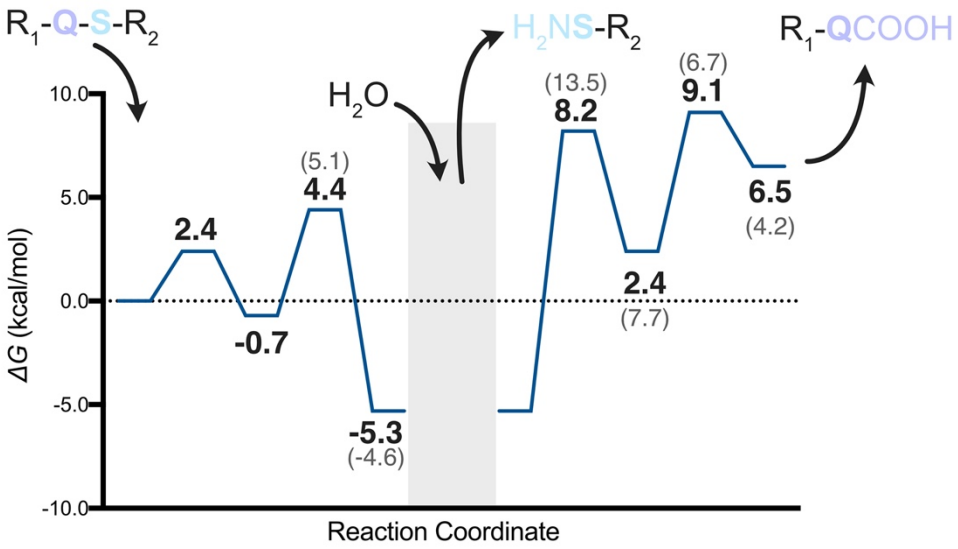


Figure 84. Complete energy profile for the catalytic mechanism of SARS-CoV-2 M^{pro}. The values placed between parenthesis correspond to the activation/reaction energies of each step.

3.1.4. Conclusion

In the past year, several studies have been published assessing different features of the SARS-CoV-2 virus and associated human disease. The M^{pro} was one of the most studied drug-targets due to the success in the treatment of other viral infections using inhibitors of similar proteases. Valuable computational studies reported important insights about the catalytic mechanism of the M^{pro} [424, 425, 430, 431]. In this work, we provide new relevant features of the catalytic mechanism of this protease. The calculations reported in this publication corroborate the hypothesis of an ion pair formation prior to the nucleophilic attack of Cys^{145} to the substrate as proposed in another study [425] and contrary to what has been suggested by Świderek *et al.* [424]. The non-natural and bulky molecular group at the P1' position of the substrate seems to be responsible for the concomitant nucleophilic attack described in that work, as well as the higher activation and reaction energies. Moreover, our results elucidate the key role played by Gly^{143}, Ser^{144}, and Cys^{145} residues in the stabilization of the oxyanion hole in lowering the activation barrier of the rate-limiting step. Their inclusion in the QM region of any QM/MM model seems essential for properly modeling their important role.

The insights provided by the calculations reported in this work, together with the important conclusions already described in previous publications, can be useful to rationally develop TS analogue inhibitors targeting M^{pro}. Usually, inhibitors should have the following characteristics: i) be recognized by the active site and allow the correct alignment of the Gln or a similar molecular scaffold in relation to the catalytic dyad, and ii) react with the active site residues and irreversibly inhibit the enzyme. To fulfill the first point, the inhibitor should follow a pattern of intermolecular interactions similar to what is observed with the original substrate peptides. Since the drug candidate will compete with the natural substrates of M^{pro} to bind into the active site, it is important to achieve a comparable or higher degree of affinity. In order to irreversibly inhibit the enzyme, the compound must react with the catalytic dyad and generate a similar thioester-enzyme adduct, as observed with the natural substrate. This reaction intermediate should be stable and prevent the hydrolysis or the reverse reaction. A good starting point is the three-dimensional structure of the TS from step 2 (available in supporting information), which can be modified to fulfill the requirements described above. We believe that this knowledge can be a step forward in the development of more specific and potent M^{pro} inhibitors.

This page was intentionally left blank.

4. Development of Scientific Software

The computer is the main, if not the only, device interface that a computational (bio)chemist has to deal with daily. Although it could seem simpler than to operate a panoply of different equipment, computational-related researchers have to handle a set of different types of computer architectures and software implementations.

The success of a computational (bio)chemistry research project relies on all the fundamental basis of any other research project. However, it is also highly dependent on the quality of the user-software and software-hardware interfaces.

This means that it is fundamental that scientific software is correctly implemented and well coded and compiled for the target hardware. The correct implementation of software to address a specific (bio)chemical problem ensures that the calculations do not introduce additional errors to the ones already accounted for a certain methodology. More important than that, an implementation should avoid introducing errors that the researchers are not aware of. Regarding adequate coding and compilation, a good approach is related with more efficient calculations and consequently faster simulations. So far, mainly the software-hardware interface was addressed, highlighting some of the important features that need to be taken into account when simulations are being performed. However, the user-software interface is also crucial for the success of the entire workflow.

In the computational (bio)chemistry field, the majority of software run on the shell with the exception of molecular visualization software as VMD and PyMOL, for example. Moreover, it is relatively simple and fast to compute a huge number of different simulations, leading to the production of a big amount of raw data. Here appears the hardest task for a computational (bio)chemist: extract valuable information from the tons of raw data that is generated by the calculations. At this specific point, the automation of processes and graphical-users-interfaces (GUIs) provide excellent support handling these tasks.

If a repetitive process could be automatized, it means that that task can be accomplished faster and with a considerably lower error rate since humans tend to make more mistakes than well-implemented automated scripts. Notwithstanding, the automation helps researchers to save time for a careful analysis of the results. Considering the automation scripts and the analysis of results, software emerges as excellent tools to make these tasks more productive and efficient.

Therefore, during the course of the project of this dissertation, some software was developed to facilitate some tasks and overcome some limitations of the already existing software packages. The following part of this subchapter is going to provide a quick overview of the software that was developed and implemented during this Ph.D. project.

4.1. VMD Store

The VMD Store plugin was developed by Henrique S. Fernandes with contributions from Nuno M.F.S.A. Cerqueira. Henrique S. Fernandes wrote the article, which was revised considering the contributions of all the authors. The work was published in the Journal of Chemical Information and Modeling, and the following content is almost a transcription of the published version with an asset of statistical analysis regarding the popularity of the VMD Store.

Herein we present the VMD Store, an open-source VMD plugin that simplifies the way that users browse, discover, install, update, and uninstall extensions for the VMD software. The VMD Store obtains data about all the indexed VMD extensions hosted on GitHub and presents a one-click mechanism to install and configure VMD extensions. This plugin arises in an attempt to aggregate all VMD extensions into a single platform. The VMD Store is available, free of charge, for Windows, macOS, and Linux at https://biosim.pt/software/ and requires VMD 1.9.3 (or later).

This page was intentionally left blank.

4.1.1. Introduction

VMD [292] is one of the most popular computational software, accounting for more than 16,000 citations since 1994 (according to Thomson Reuters Web of Science database). This popularity is a consequence of its powerful capabilities and free distribution. Notwithstanding, VMD permits an easy interaction with the Tcl scripting language, enabling the development of additional extensions programmed in the Tcl/Tk language. This means that users can develop their own scripts/extensions to complement VMD and fit their needs. This possibility gave rise to a flourishing community of VMD plugin developers and users all around the world.

Over the past years, a panoply of VMD extensions have been developed and published, namely WATCLUST [432], Epock [433], molUP [316], PyContact [434], vsLab [435], RIP-MD [436], ChemPathTracker [437], METAGUI 3 [438], vmdMagazine [439], CaFE [440], VolArea [441], TopoGromacs [442], vmdICE [443], SIRAH tools [444], and DelPhi Force [445].

Although VMD is equipped with a powerful and complete set of tools, there are some specific needs for all the research fields that resort to VMD. Third-party VMD extensions provide valuable new capabilities that strengthen VMD's native capabilities and improve its ability to handle other tasks, such as the preparation and analysis of calculations with quantum mechanics, ONIOM QM/MM, protein-ligand and protein-protein-docking, virtual screening, molecular mechanics, molecular dynamics, as well as structural analysis, with different external software packages.

The main problem is that these plugins are spread over the internet, lacking a general repository. Consequently, the process of finding a new plugin becomes very difficult, and sometimes the files become lost, either because the authors no longer support them, have changed university or research group, or the website was shut down.

Another important aspect is the update of the plugins, which is often challenging to make, and it is imperative to maintain the availability and compatibility of the plugins. Moreover, the installation procedure is usually distinct among different plugins.

In order to overcome all these issues regarding the handling of VMD plugins, the VMD Store was developed [369]. VMD Store is an independent open-source VMD plugin that allows VMD users to browse, discover, install, and manage other VMD extensions, connecting the individual plugin developers with the VMD plugin users.

4.1.2. Development

The VMD Store was developed to offer a free and intuitive GUI for search, discover, install, update, and remove VMD extensions (Figure 85). It was prepared to work in every computer capable of running VMD version 1.9.3 or higher.

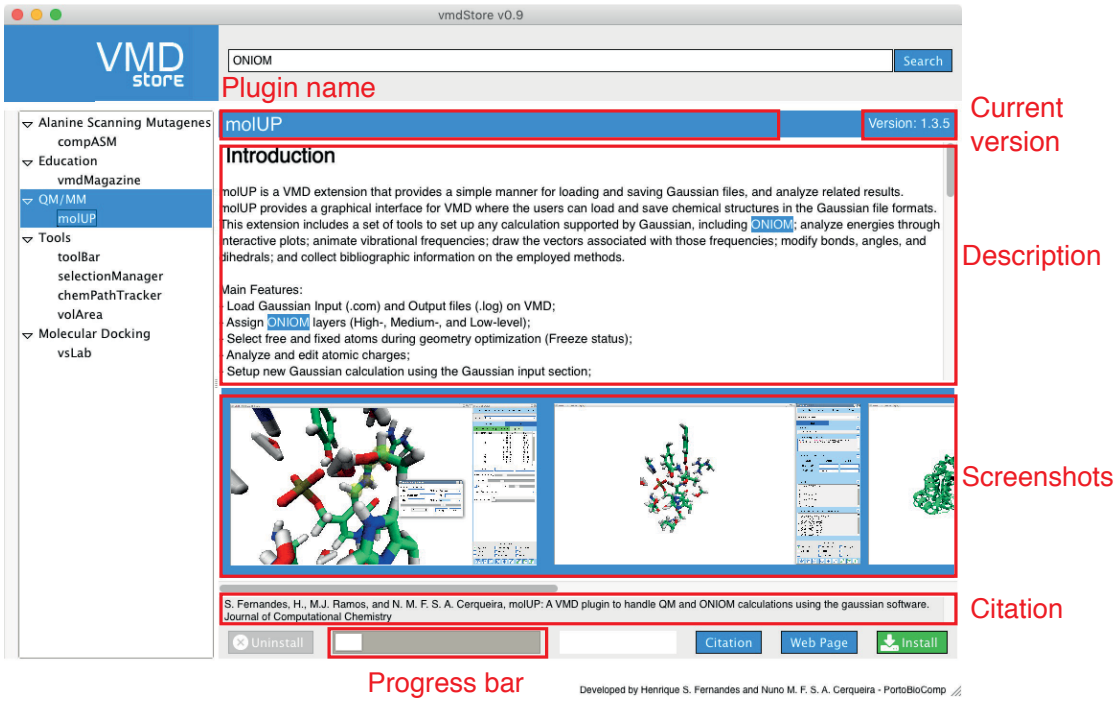


Figure 85. The main window of molUP extension on the VMD Store. On the left panel, a list of all available plugins is presented, organized by category. In the main section, the VMD Store provides information about the plugin's name, latest version, description, screenshots, and citation. Moreover, there are buttons to install or uninstall the plugin, as well as links for the citable paper and its repository on GitHub.

The VMD Store can be installed through a graphical installer that works on Windows, macOS, and Linux, as well as a script to install the VMD Store through the bash or the Tk console of VMD.

All the backend of the VMD Store is based on GitHub. GitHub is one of the biggest repositories of software in the world. This platform was chosen because it is already used by millions of developers around the world, and it is free.

Therefore, the VMD Store is hosted on GitHub, and it also collects all the data from GitHub repositories. The VMD Store will index all repositories that are listed on the "Available Plugins" section of the VMD Store repository. Each repository is associated solely with one plugin that is shown in the "List of VMD extensions" of the GUI. Once the user selects a certain plugin, the VMD Store collects all the available data from the respective repository on GitHub, such as the plugin name, the latest version available, the description, and the screenshots. At any time, the user can press the "Install" button

to start the installation process. This process involves the download of the latest version, unpack of the files, and modification of all required files to make the VMD plugin available for the user. Each plugin has its own “install.txt” file that includes all the instructions to install the plugin, including, for example, the loading of required external packages.

This approach ensures that different developers can have their own code on their own GitHub repositories. Consequently, developers can manage their own code, and the VMD Store is still able to index and install. This strategy makes the VMD Store indexing process seamless, and no delay is added to the release of new software versions.

Any developer of VMD extensions can make his/her plugin available on the VMD Store, following two simple steps:

1. Make the VMD plugin available on a public GitHub Repository following the VMD Store guidelines of indexing;
2. Contact VMD Store developers to include the new plugin in the VMD Store.

VMD Store also provides an excellent platform to promote VMD plugins, ensuring that the credits are attributed to the plugin’s authors. This is ensured by including the proper citation and the webpage where the article describing the plugin is available.

4.1.3. VMD Store Popularity and Statistics

So far, the VMD Store indexes eight VMD plugins developed in our research group, the DelPhiForce [445] plugin developed by Emil Alexov *et al.*, and the Pathways [446] plugin by Ilya Balabin *et al.*. The VMD Store is always prepared to include additional VMD extensions in its repertoire.

Since the initial release of the VMD Store installer in February 2018, the VMD Store was already downloaded more than 3300 times (**Figure 86**) worldwide by 78 different countries. Among all these countries, China, the United States of America (USA), India, Portugal, and Spain lead this list with almost half of the downloads.

Regarding the most common operating systems, the VMD Store was downloaded more than 50% for Windows computers, followed by Linux with more than 26%, and macOS with approximately 17%.

All the statistics were independently grabbed from the SourceForge repository (https://sourceforge.net/) that hosts and make the VMD Store installers available.

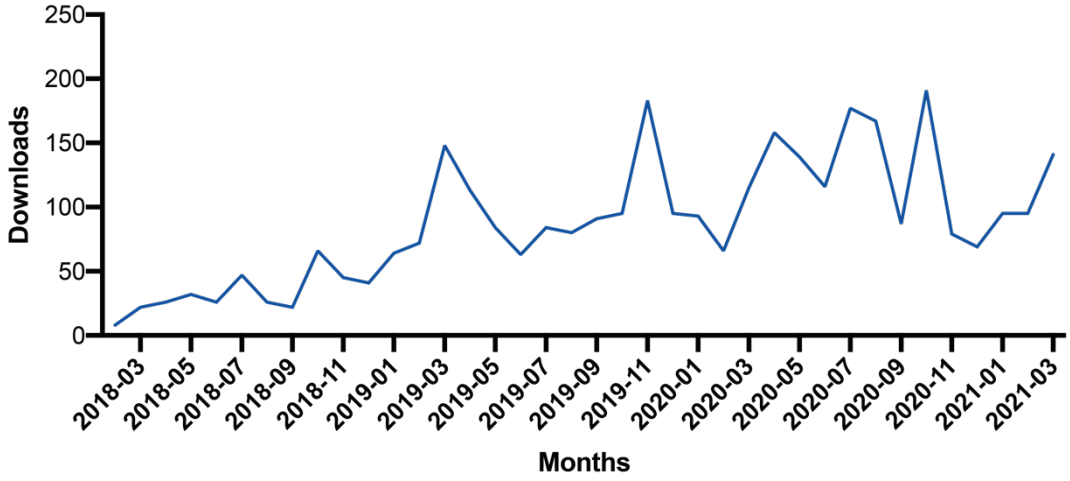


Figure 86. Statistical data regarding the number of downloads of the VMD Store during the time period from February 2018 to August 2020.

The VMD Store is available, for free, at https://biosim.pt/software/ and the tutorial of the installation process is available at https://biosim.pt/how-to-install-vmd-store/, including a video tutorial.

Download VMD Store

VMD Store Installation Tutorial

4.2. ToolBar

4.2.1. Introduction

The 21st century was a turning point in terms of access to technology for the general population. The development of the technology industry was able to provide more powerful devices at low prices. This reality crossed multiple fields, and the chemical and biochemical research areas are no exception. The recent advances in CPU and GPU processing power allow us today to compute, analyze, and generate high-resolution representations of complex biochemical structures, such as proteins and DNA. These features, together with the substantial advances in algorithms and numerical methods, made possible the simulation of molecular behavior and chemical reactivity through computers. These features forecast the simulation of experimental data and, at the same time, propose new ideas (some of them were later successfully validated by experimental means) [447-449].

There is no doubt that molecular modeling plays, today, a prominent role in biological and chemical research. However, while mathematical analysis is critical, it remains true that many insights come only from visual inspection. Nowadays, many programs are dedicated to macromolecular visualization and analysis. This allows rationalizing the chemical understanding of complex systems in a more detailed and atomistic manner. The computational study of enzymatic mechanisms [188, 196, 450-454] was one of the fields that developed most with the improvement in molecular visualization software. This kind of software also boosted molecular docking and virtual screening studies [455-457], providing essential insights in the rational drug design.

Some of the most popular software packages include the Visual Molecular Dynamics (VMD) [292], PyMOL [458], and UCSF Chimera [459]. VMD is one of the most popular ones, mentioned in more than 36,000 publications, according to Scopus (2020)[12]. In comparison, PyMOL and Chimera were mentioned in 24819[13] and 20660[14] publications, respectively. Although VMD has been updated frequently, including more features and tools, its graphical interface continues lacking several features that other programs already have, such as a toolbar to make a quick access to the most used commands, an

[12] Keywords “Visual Molecular Dynamics” or “VMD” searched in all fields. Access date: 2020, December 21st
[13] Keyword “PyMOL” searched in all fields. Access date: 2020, December 21st
[14] Keyword “UCSF Chimera” searched in all fields. Access date: 2020, December 21st

easy and quick way to render images, the availability of presets that can easily change the representation and colors of a visualization, among others. The lack of such functionalities makes VMD challenging to use by inexperienced users and time-demanding for advanced users due to the several steps that need to be accomplished to perform simple tasks.

Herein, we developed a ToolBar for VMD, an open-source extension developed to improve the VMD's user-friendliness and enhance its versatility.

4.2.2. The ToolBar

As aforementioned, the ToolBar is a free and open-source extension for VMD that grab frequently used tools in a sidebar that can be easily accessed by VMD users. Additionally, this plugin provides a set of new tools that are not natively available on VMD. The following sections are going to discuss some of the most important features of the ToolBar, starting with a summary of how it can be installed.

a. Installation

The installation of ToolBar is highly facilitated by the VMD Store [369] extension that was discussed in the previous chapter. Since the VMD Store is installed, the ToolBar can be installed simply by clicking on it in the left sidebar and then on the “Install” button.

Alternatively, the ToolBar can be manually installed using the information provided in the “install.txt” file. The installation is conducted simply by copying the content of the “install.txt” file into the .vmdrc/vmd.rc file, pointing the “lappend auto_path PATH” command to the path of the directory containing the ToolBar files.

After the successful installation of the ToolBar, the ToolBar could be assessed through the “Extensions” > “VMD Store” > “Tool Bar” menu of the VMD.

b. Graphical User Interface (GUI) of ToolBar

The ToolBar was designed to fit closer to the VMD OpenGL Display window, offering the shortest possible path to access the ToolBar’s icons (Figure 87). The minimalist design of the ToolBar, placing icons in two columns, maximizes the productivity without compromising the area of the screen available for molecular visualization.

The ToolBar gathers a set of icons for tasks that are commonly needed in the routine of VMD users. In particular, buttons to load/save chemical structures and visualization states, access the “Main” and “Representations” windows of VMD, and easily switch between mouse modes (rotate, scale, and translate).

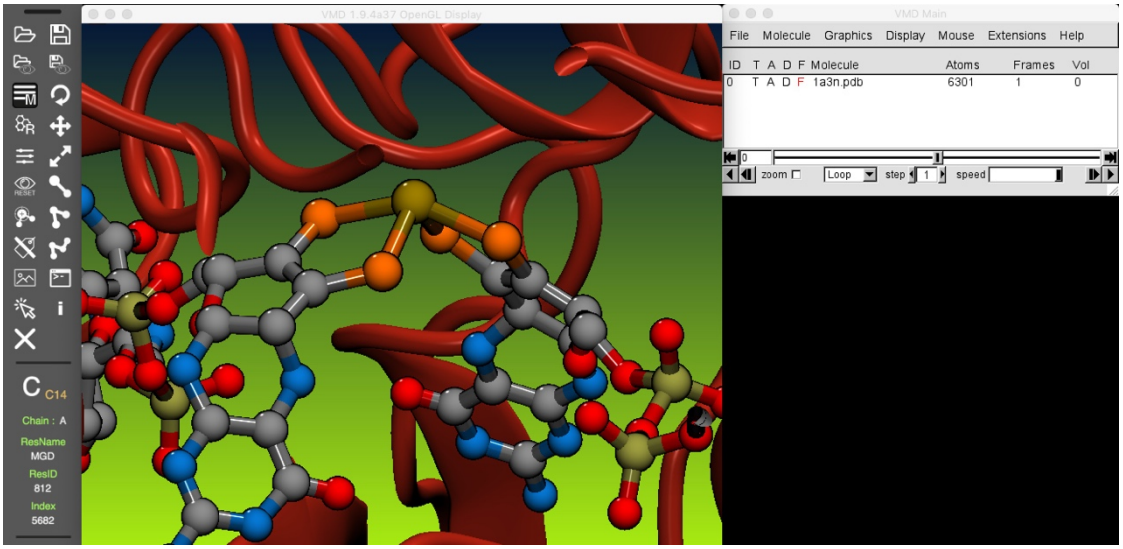


Figure 87. Screenshot showing the ToolBar (left), the VMD OpenGL Display window (middle), and the VMD Main window (right). The structure of Nitrate Reductase (PDB code: 2NAP) [460] was loaded as an example.

This extension not only provides a set of icons to assist the way users can reach some built-in VMD features but also includes a set of new tools to make VMD even more versatile (Figure 88). In particular, this version includes five new features: “Delete all labels”, “Image Render”, “Data about picked atoms”, “Presets”, and “Modify Tool”.

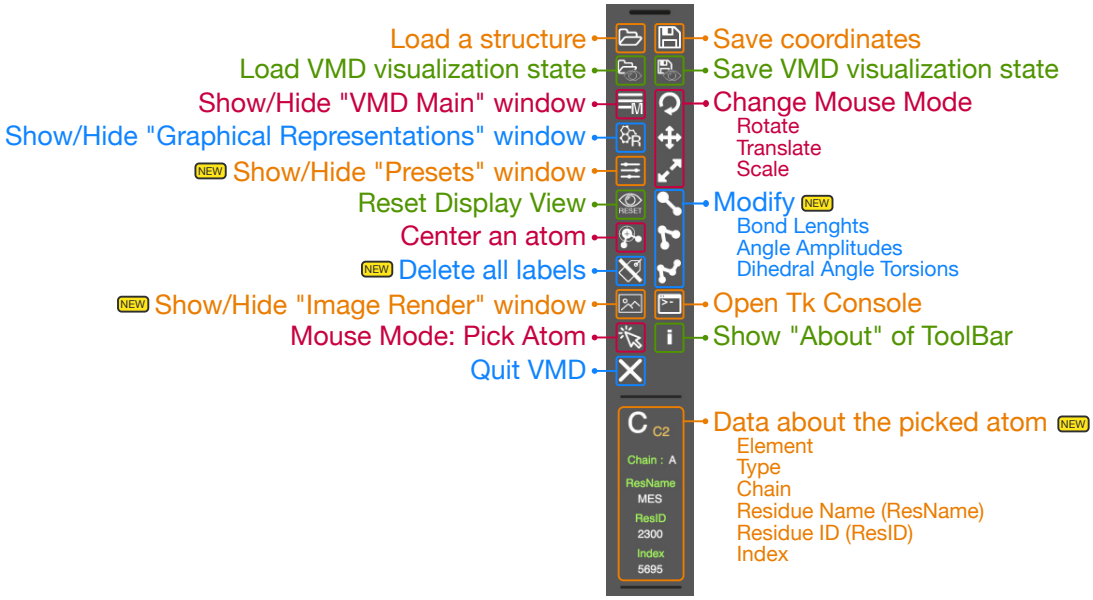


Figure 88. Description of the functions associated with all elements available in the ToolBar. The “NEW” tag identifies all the new features added by the ToolBar to the VMD’s experience. All other features were already available in VMD, but they were included in the Toolbar, in a pleasant and more accessible manner.

The “Delete all labels” button activates a simple script that automatically deletes all labels displayed in the OpenGL Display window. This feature is particularly useful when a certain chemical system is being inspected, and several atoms are picked to identify their

molecular attributes, such as *resid*, *chain*, *index*, among others. In the end, the display usually becomes chaotic with all those labels, and VMD does not provide any standard and easy way to clean all those labels through a single step.

The “Image Render” is a simple graphical interface that organizes the options available for the VMD built-in render software - Tachyon. In a simple way, render is a package that collects data about the tridimensional representation to produce high-quality images. The “Image Render” allows getting, through a single click, a high-resolution image from the OpenGL window of VMD. This feature resembles some options that are available in other visualization software but not in VMD. Within the “Image Render” menu, users can explore options to improve the quality of their image productions, namely modifying “Ambient Occlusion”, “Shadow”, “Anti-Aliasing”, and “Resolution”. Depending on the features that are activated or parameters that are changed, the time required to render the image can increase dramatically.

Standardly, when an atom is picked in the OpenGL window of VMD, data about the picked atom is printed in the shell, which is commonly placed behind all other VMD’s windows. In order to make this information more accessible, the ToolBar includes a section where this data is displayed in a more user-friendly approach (Figure 89). In addition, the atom that is picked is highlighted in the OpenGL window with a red sphere. This solves one of the popular problems in VMD, where there is no visual indicator about what is selected or clicked in the OpenGL window of VMD. This feature was also included, while atoms are selected to measure bonds, angles, and dihedral angles.

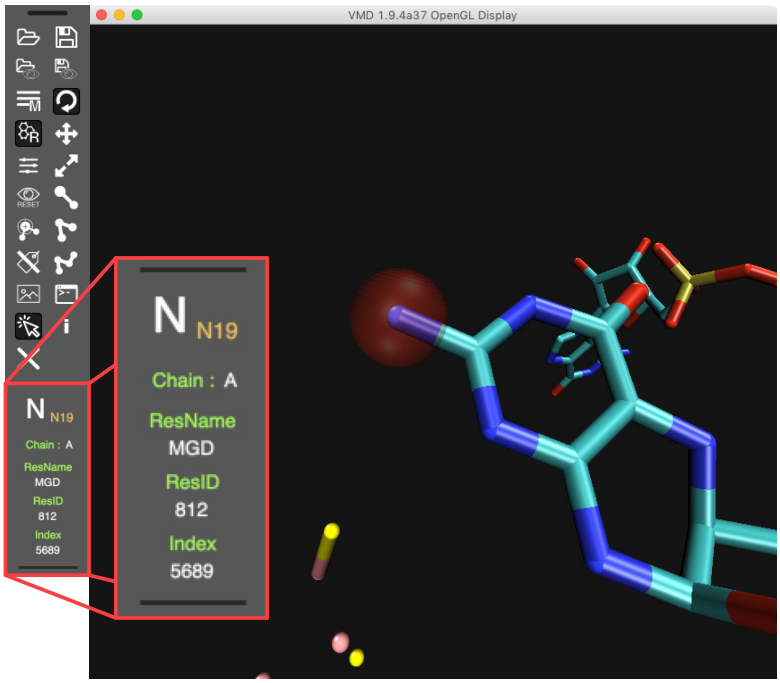


Figure 89. Screenshot showing a nitrogen atom picked using the ToolBar. The ToolBar displays information about the atom that was picked, namely the element, atom’s *name*, *chain*, *ResName*, *ResID*, and *index*.

Although VMD is capable of producing exceptional representations of molecular systems, the preparation of such representations is not an easy task, particularly for inexperienced users. The “Presets” menu was developed to provide an easy-to-use GUI where users can change the background’s color/gradient, adjust the view, or regulate the clip-plane (Figure 90). Notwithstanding, default and user-custom representation templates can be straightforwardly applied to any molecular system loaded in VMD. This feature lets users create high-quality molecular representations through a single click. A set of presets was already included.

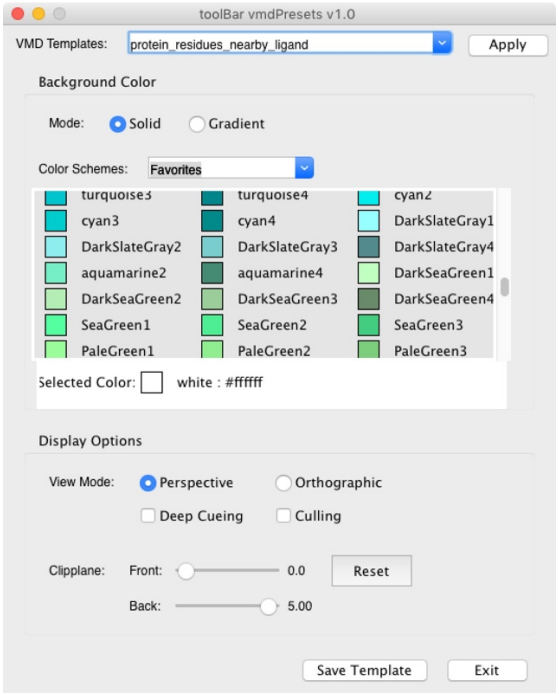


Figure 90. Screenshot of the “Presets” menu available on ToolBar. The “Presets” menu provides options to modify the appearance of the molecular representation in VMD.

VMD is an excellent software to visualize molecular systems, but it lacks tools for moving atoms or groups of atoms. As an attempt to overcome these limitations, ToolBar includes a set of three tools to move atoms and groups of atoms associated with the adjustment of chemical bond lengths, angles amplitudes, or dihedral angles torsions (Figure 91).

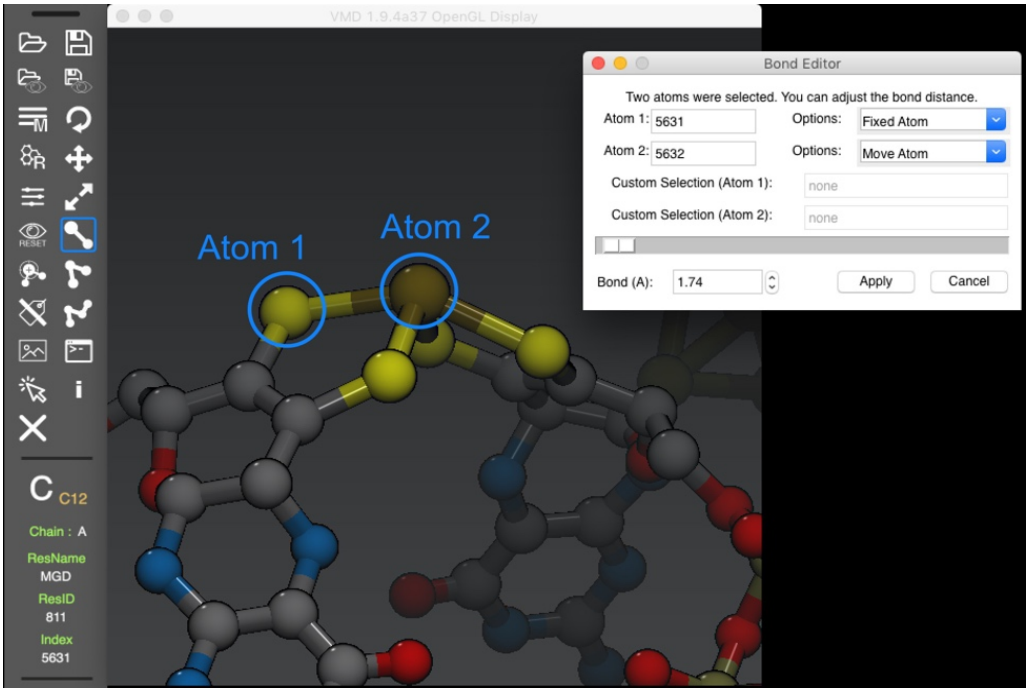


Figure 91. Screenshot showing the bond length editor provided by the ToolBar. Since the tool is activated, the user picks two atoms consecutively to enable the “Bond Editor” window, where the distance between those atoms can be adjusted.

4.2.3. Technical Implementation

As for VMD Store, the ToolBar was also developed in Tcl/Tk, the language natively available on VMD. No external library or packages are required, making the ToolBar widely compatible with all systems capable of running VMD.
The ToolBar was tested on the latest stable VMD release (1.9.3) and the current tester version (1.9.4). The results showed that the ToolBar works flawlessly in all tested setups. Since ToolBar is a useful extension for almost all situations, it is highly recommended to include the Toolbar as a plugin that automatically opens every time VMD is started. Therefore, it is advisable that the line “toolBar:startGui” be added to the .vmdrc/vmd.rc file.

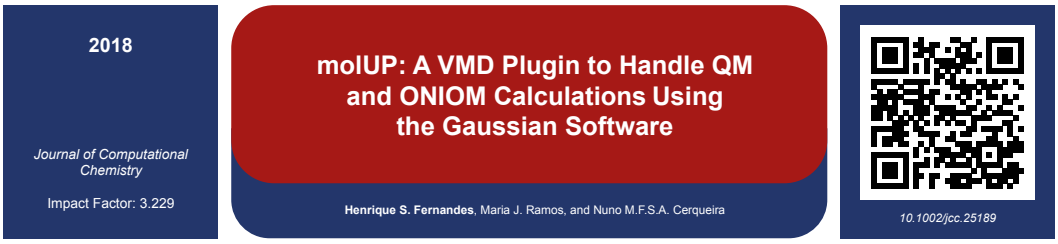
4.2.4. Conclusion

VMD is one of the most powerful and used software for molecular visualization. This position is massively guaranteed by the fact that it is free, supports the representation of millions of atoms, allows automation by script, and the development of external tools by the community. The ToolBar is precisely an example of an external tool developed to improve the natively available tools of VMD. In particular, the ToolBar seeks to improve the VMD’s user-friendliness and productivity of the users using it. Moreover, it can contribute to shortening the learning time for new users.

Based on our experience, VMD is an exceptional tool for molecular visualization in teaching activities, either in classes or workshops. The learning curve of VMD use has decreased substantially with the use of the ToolBar. This allowed inexperienced users to become more interested in using VMD, and the outcome of the visualizations and the complexity of the analysis have improved significantly. Advanced users can also take advantage of using the ToolBar in their research activities, as in our research lab. In this case, the ToolBar improved productivity, decreasing the number of steps to perform simple or even more complex tasks in VMD. For example, considering the button to delete all labels, this task can be accomplished with a single click using the ToolBar. If the users want to delete all labels without the ToolBar, they need to go to “Graphics” > “Labels…” menu, select all the labels and click on “Delete”. Since it was installed, it has become one of the most used VMD extensions.

4.3. molUP

The molUP plugin was developed by Henrique S. Fernandes with some contributions from Nuno M.F.S.A. Cerqueira. Henrique S. Fernandes wrote the article, which was revised considering the contributions of all the authors. The work was published in the Journal of Computational Chemistry, and the following content is almost a transcription of the published version with some modifications regarding the recent updates of the plugin. This manuscript was also highlighted as the cover of the issue where the paper was published (Appendix: page 391).



The notable advances obtained by computational (bio)chemistry provided its widespread use in many areas of science, in particular, in the study of reaction mechanisms. These studies involve a huge number of complex calculations, which are often carried out using the Gaussian suite of programs. The preparation of input files and the analysis of the output files are not easy tasks and often involve laborious and complex steps. Taking this into account, we developed molUP: a VMD plugin that offers a complete set of tools that enhance the preparation of QM and ONIOM (QM/MM, QM/QM, and QM/QM/MM) calculations. The starting structures for these calculations can be imported from different chemical formats. A set of tools is available to help the user to examine or modify any geometry parameter. This includes the definition of layers in ONIOM calculations, choosing fixed atoms during geometry optimizations, the recalculation or adjustment of the atomic charges, performing relaxed scans or IRC calculations, etc. molUP also extracts the geometries from the output files as well as the energies of each of them. All of these tasks are performed in an interactive GUI that is extremely helpful for the user. molUP was developed to be easy to handle by inexperienced users, but simultaneously to be a fast and flexible graphical interface to allow the advanced users to take full advantage of this plugin. The program is available, free of charges, for macOS, Linux, and Windows at the BioSIM page https://biosim.pt/molup/

This work was also presented in the following conferences:

- (Oral Communication) *Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, **2017**, Coimbra, Portugal
- (Oral Communication) *XXV Encontro Galego-Português de Química*, **2019**, Santiago de Compostela, Spain

4.3.1. Introduction

Most of the research works described in this thesis address the study of the catalytic mechanism of different enzymes using a hybrid approach called ONIOM. The ONIOM methods allows the treatment of huge biomolecular systems using two different computational methods with different accuracy and computational cost.

The ONIOM method, developed by Morokuma and co-workers, is well implemented in the Gaussian software, and it became a popular methodology to study chemical systems within a QM/MM approach. According to Scopus, in August 2020, the “ONIOM” term was mentioned by more than 1400 publications, corresponding to almost 25% of all the publications that mentioned the term “QM/MM” (more than 5700).

Although all the popularity of this method and Gaussian software, there was a lack of appropriate software to prepare and analyze the calculation results.

Gaussview is a graphical software, developed by Gaussian, Inc., that provides some tools to handle the preparation of Gaussian input files and the analysis of output files. Although Gaussview is well designed for small molecular systems (up to a few hundred atoms), it lacks tools and features that make it useful for the preparation and analysis of results from complex and big biomolecular systems. Therefore, the molUP software was developed [316], as a plugin for VMD, to address these issues and provide a powerful free alternative to Gaussview.

molUP equips VMD with a set of tools and features that makes VMD capable of creating, manipulating, and analyzing the input and output files from calculations performed with Gaussian software. The popularity of VMD and Gaussian software dictated developing molUP to interface these two pieces of software.

4.3.2. Development

molUP was coded in Tcl/Tk to ensure perfect integration as an extension of VMD. In addition, some external packages, such as GNU grep, sed, head, and tail, were also included to speed up the handling of huge Gaussian output files that can have more than a dozen gigabytes.

4.3.3. molUP Features

molUP GUI is composed of five main sections: a menubar, a molecule selector, a frame navigator, a central section, a quick representations chooser, and, at the bottom, a set of icons where frequently used tools were made more accessible (Figure 92).

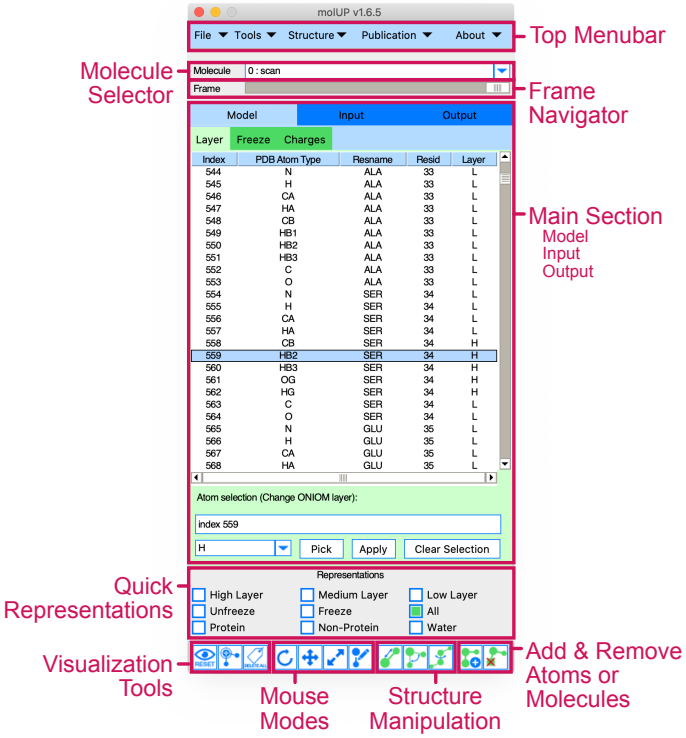


Figure 92. Screenshot of the main GUI of molUP. The main areas of the GUI are tagged.

The top menubar was designed to provide five dropdown menus ensuring easy access to some of the key features of molUP. For example, from the “File" menu, the user can load Gaussian output files to the VMD or simply save new Gaussian input files. Another interesting feature of molUP is available under the “Publication" menu. In this section, users have access to a methodological overview of the calculation that has been set up in molUP, as well as some references that need to be cited according to the type of methods used.

The molecular selector provides access to a dropdown list of all the molecules loaded on VMD. In this way, users can alternate between molecules with a simple couple of mouse clicks. This is particularly interesting to analyze more than one calculation at the same time, recurring to the VMD’s tools to overlap and compare structures. This kind of analyzes is not possible using other types of molecular visualizing software.

The frame navigator is a horizontal scrollbar designed to facilitate the navigation across multiple structures of a certain structure (if applicable). Although this feature is also available on the main window of VMD, several users asked for this implementation to facilitate their tasks.

The main section of molUP is where the main data is shown, and the main operations take place. Further details are going to be discussed in the following sections.

There is a set of molecular visualization commonly used by researchers that perform QM or QM/MM calculations. In particular, representations that show the protein in *NewCartoon*, the water molecules, the atoms belonging to each ONIOM layer, the atoms fixed during the optimization process, among others. In order to simplify the way atoms are shown, a set of custom molecular representations are available just a click away. Nonetheless, users are free to use the native “Representations” tool of VMD as well.

Finally, the bottom section includes shortcuts to some of the most frequently used tools available on VMD and also some molUP assets that allow users to modify bonds, angles, and dihedral angles, as well as the addition and elimination of atoms or molecules.

All the features integrated into molUP allow users to follow the typical workflow used to study a molecular system. This workflow usually starts with the (1) preparation of the input files, (2) running the calculations, in this case using Gaussian software, and (3) the analysis of the results. (**Figure 93**) Based on the overall protocol, each section of the molUP plugin is going to be presented in more detail in the following sections.

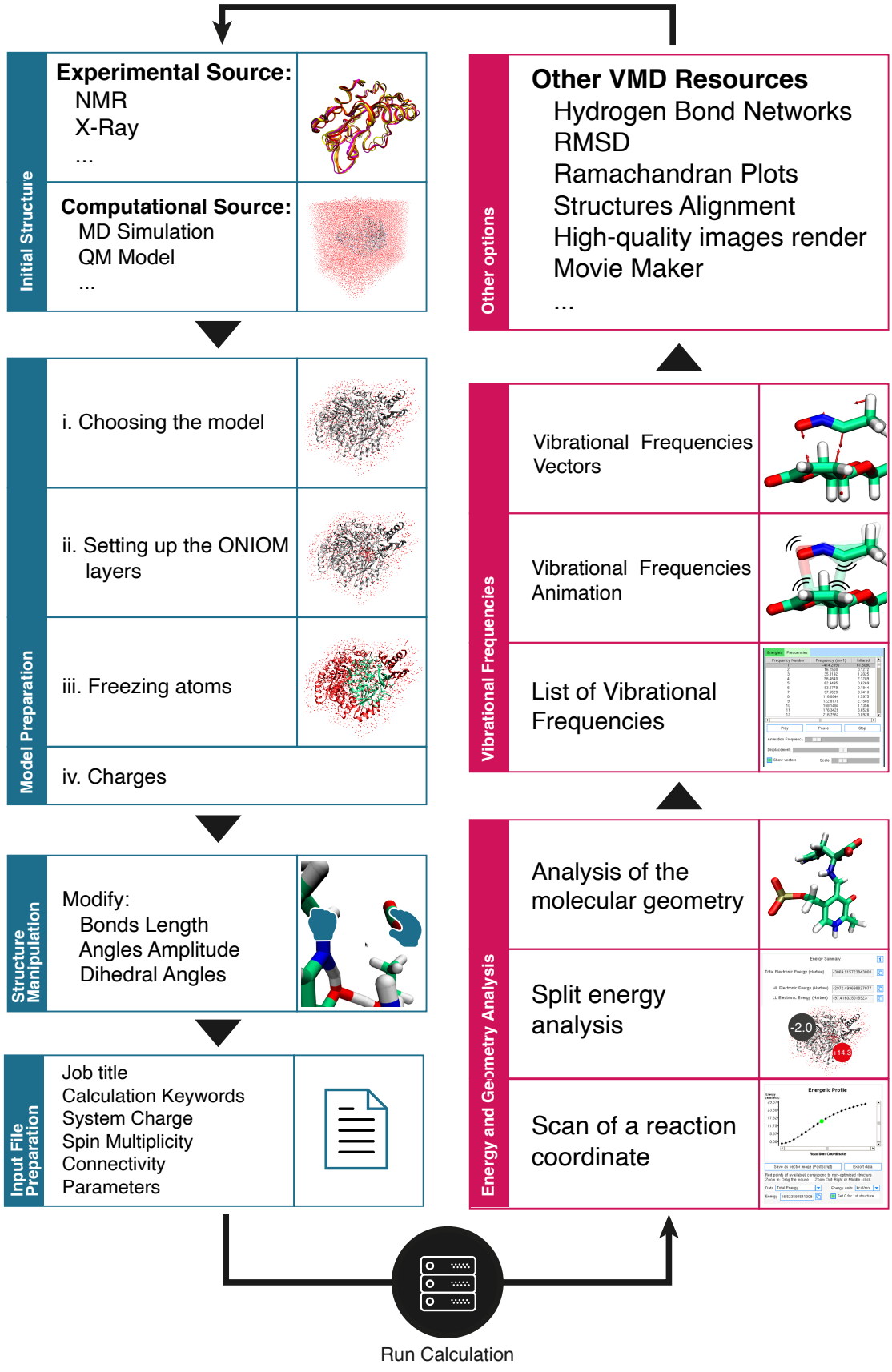


Figure 93. Flowchart of the common strategy used to carry on a QM/MM study using Gaussian software, VMD, and molUP.

a. *Model Section of molUP plugin*

The input file for a Gaussian calculation includes the keywords for the required type of calculation as well as the atomic coordinates. Moreover, this file can also include other auxiliary information needed in some types of calculations, such as the MM parameters.

molUP was built as a plugin of VMD. Therefore, it benefits from the VMD's capabilities to load structures from a wide range of different chemical formats as PDB, MOL2, and XYZ, or even specific file formats for software packages like AMBER, CHARMM, Autodock, Gromacs, and among others. Consequently, molUP can easily interpret structures from different file formats and prepare them for Gaussian calculations. Although VMD provides a panoply of supported file formats, VMD cannot load Gaussian input or output files. Consequently, molUP was prepared to load Gaussian-related files, supporting different types of calculations from simple QM to more complex QM/QM/MM methodologies. molUP can load Gaussian input and output files from the GUI (Figure 94) or directly from the *bash* through a built-in script delivered with molUP that need to be placed as a *bash* alias:

```
alias molUP='vmd -e "/molUP[15]/lib/bashScript.tcl" -args'
```

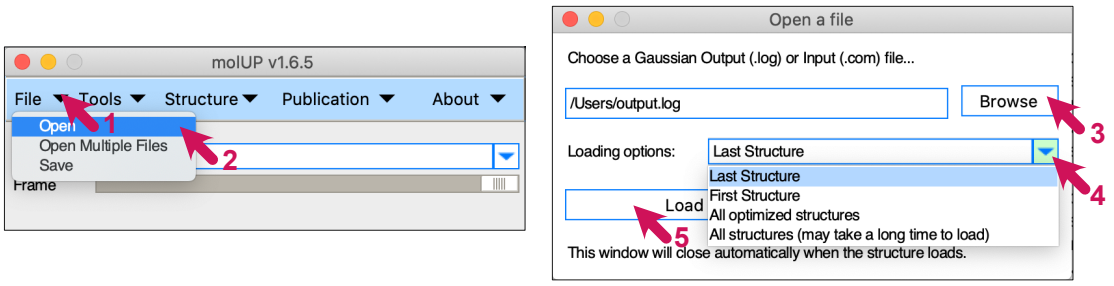


Figure 94. molUP guide to load a Gaussian-related file using the molUP GUI.

Regarding the Gaussian output files, users can choose which structure(s) should be loaded: first structure, last structure, all optimized structures (e.g., relaxed coordinated scans), or all structures (e.g., geometry optimization).

Since the molecular system is loaded in VMD, users can adjust the positions of atoms or even add or remove atoms or molecules. A set of tools were developed for molUP to overcome this limitation of VMD that natively does not allow atom insertion, elimination, or modification. molUP includes a set of three tools to adjust the bond length between

[15] In system paths, "molUP" represents the molUP installation directory.

two atoms, the angle bend between a trio, or the dihedral torsion of a set of four atoms (Figure 95). Additionally, VMD includes a tool to delete atoms simply by clicking on them or using VMD atom selections. The tool to add atoms or molecules creates a layer where the atoms are created. The user can add and adjust the position of all the added atoms/molecules before definitely merge them with the original system. All of these are available in the interactive and user-friendly interface of molUP.

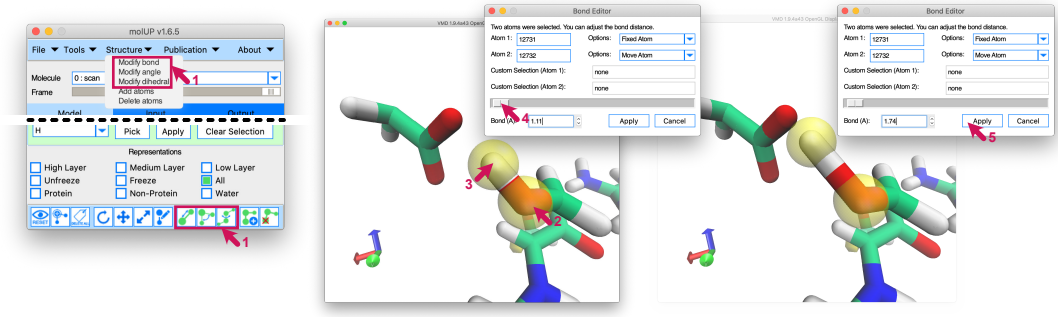


Figure 95. molUP guide to adjust the bond length between two atoms.

After all the modifications in the atomic positions, it is necessary to set up the model system for the computational calculations. In the study of an enzymatic mechanism by computational means, the choice of the model is one of the first requirements. A good model must represent the chemistry of the system that will be studied and must provide a balanced compromise between model size and computational cost. In particular, when a QM/MM study is conducted, particular attention needs to be given to which atoms are included in the HL layer that is treated using a more costly method. However, other adjustments could also be required, such as choose which atoms should be optimized and/or frozen during geometry optimizations or modify the charge of the atoms. These tasks can be accomplished through the “Model” tab of molUP that includes a subtab for each case: “Layer”, “Freeze”, and “Charges” (Figure 96). The “Layer” and “Freeze” subtabs share the same type of behavior, allowing interactive selections of the atoms in the table, picking the atoms from the representation, or using VMD atom selections. Once the atoms are selected, users can assign a layer for hybrid calculations or fix/free the position of those atoms during geometry optimizations. The “Charges” subtab shows the MM charges of all atoms of the system and provides a simple manner for users to edit them or updated them based on a Gaussian output file.

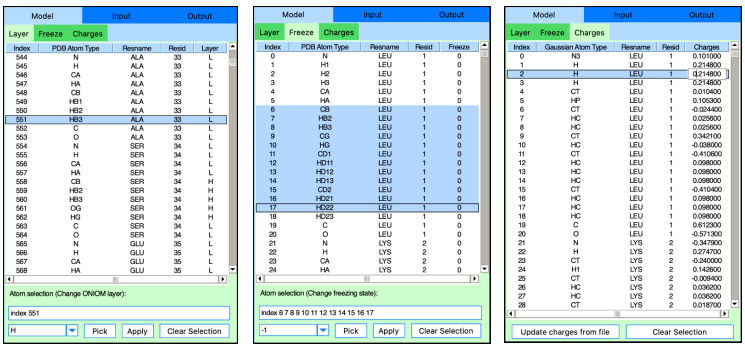


Figure 96. Screenshot of the “Layer”, “Freeze”, and “Charges” subtabs of the “Model” section of the molUP GUI.

b. Input Section

After all the structure modifications, defining layers, choosing fixed atoms, and adjusting atomic charges, users are ready to save a Gaussian input file. Though, the Gaussian calculation keywords need to be set beforehand, as well as other options such as the job’s name, system global charge and spin multiplicity, connectivity, and MM parameters. In molUP, this information can be customized in the "Input" section, following the same order of a typical Gaussian input file (Figure 97).

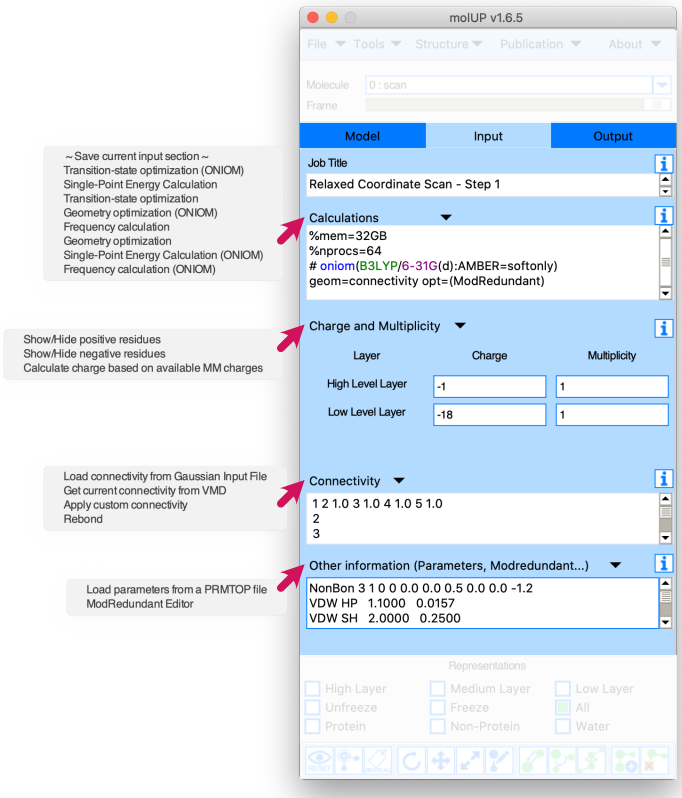


Figure 97. Screenshot of the “Input” section of molUP and the tools available on the dropdown menu of each block.

The job's title is a section where users can freely use to identify the job or place some notes regarding the calculation.

Immediately below, an assisted text area is used to place all the Gaussian-related keywords for the calculation ("Gaussian Route Section"), together with the "Route 0 section" that usually includes information about the required hardware. As mentioned, this field is assisted by a color scheme that identifies different types of keywords, *e.g.*, functionals, basis sets, methods, etc. Additionally, users can customize the colors of each keywords and also add other keywords to this mechanism through the "/molUP/user/references.txt" file.

In the course of a computational project, it is common that the same Gaussian input keywords are used for several different calculations. Therefore, molUP was coded to include the possibility to save and re-use templates. Through a click over the "Calculation" dropdown, users can save or use templates to speed up their work.

The next section of the "Input" section is where the total charge and spin multiplicity of the model system are defined. In the case of ONIOM multi-layered calculations, this section can also include the charge and spin multiplicity of the atoms included in different layers. In order to simplify the process of assignment of charges for each layer, molUP automatically estimates the charge of each layer based on the atomic charges, if available. Additionally, molUP allows users to highlight the molecules that have a negative or a positive charge, again based on the available atomic charges.

The "Connectivity" block includes information about how the atoms are bonded to each other in the system. Although this information is not required for QM calculations, when MM is involved, the connectivity is essential to apply the force field energy equation. The connectivity can be imported from another Gaussian input file, from VMD, or predicted from scratch based on TopoTools [461].

Lastly, the "Input" section shows the "Other information" field where additional input data can be provided, e.g., *ModRedundant* commands to freeze one or more coordinates, bond lengths, angles, or dihedral angles during geometry optimization or perform relaxed coordinate scans; include custom basis sets; and/or include MM parameters.

Coordinate relaxed scans are one of the most important tasks used in the study of mechanisms. Therefore, molUP includes an assisted interface ("ModRedundant editor") to pick the atom and adjust the parameters for a particular scan. This tool invokes a window where the size and the number of steps or the scan can be defined.

molUP was also prepared to facilitate the migration from an AMBER simulation to the QM/MM study, for example. So, users can find, in the "Other information" section, a tool

to automatically import the MM parameters from an AMBER topology file (*.prmtop*) and convert them to the suitable format for Gaussian.

The inclusion of additional AMBER parameters is not always mandatory. Gaussian includes the AMBER force field and the MM parameters for standard amino acid residues and nucleic acids natively. However, if the target molecular system includes non-standard molecules that were previously parameterized, the inclusion of these MM parameters becomes obligatory.

After all the adjustments made so far, the system is ready to be saved as a Gaussian input file. This can be accomplished through the “File” > “Save” command on the top menubar (Figure 98). For most of cases, this process is simple and straightforward, but in more advanced situations, some customization can be made. molUP was prepared to allow users to save only a part of the system or even choose a different file format as an ORCA [250] input file. As an example, it can be useful to save only the HL layer of a QM/MM calculation. In that case, molUP saves the HL layer as a new input file adding hydrogens atoms as link atoms for the bonds that span across the interface between two layers.

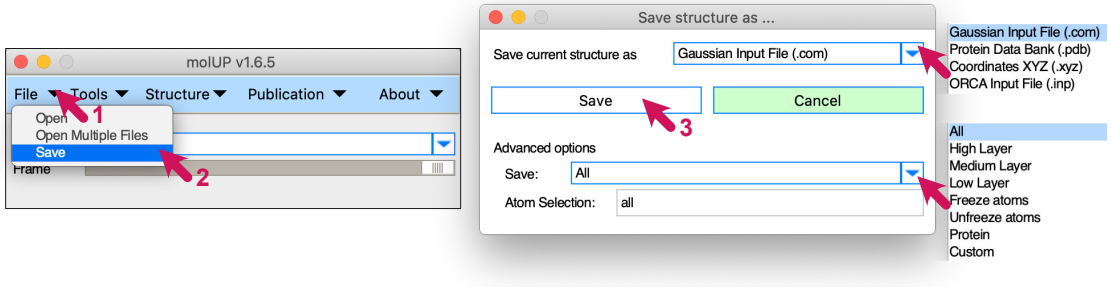


Figure 98. molUP guide to save Gaussian input file using molUP. Alternative file formats and advanced options are shown on the right.

c. *Output Section*

After running a Gaussian calculation, a detailed and extensive output file is generated, which can also be loaded and analyzed by molUP. Routinely performed tasks, such as the analysis of the molecular geometries, energy data, and vibrational frequencies can be easily performed using molUP.

A Gaussian output file can be loaded following the same steps as explained before (Figure 94). In the case of a still running calculation, incomplete output or error, warning feedback is displayed to alert users for that condition.

When solely one structure is loaded from the output file, molUP summarizes the total computed energy, as well as some energy decomposition for multilayered calculations or thermal corrections when available (Figure 99).

Furthermore, users can opt for loading all structures of a geometry optimization or all the optimized structures of a scan calculation. In these cases, molUP plots the energies for all the loaded structures in an interactive graph (Figure 99). The interactive graph can be adjusted in terms of energy units, layer energy contribution[16], and set the baseline for the first structure. Notwithstanding, users can also click on each dot on the plot to show the corresponding structure. In the case of some of the structures are not totally optimized, the plotted dots will be colored red to warn the user.

molUP also provides tools to export the graph as a *PostScript* image or even export the data as a text file to import in a spreadsheet, for example.

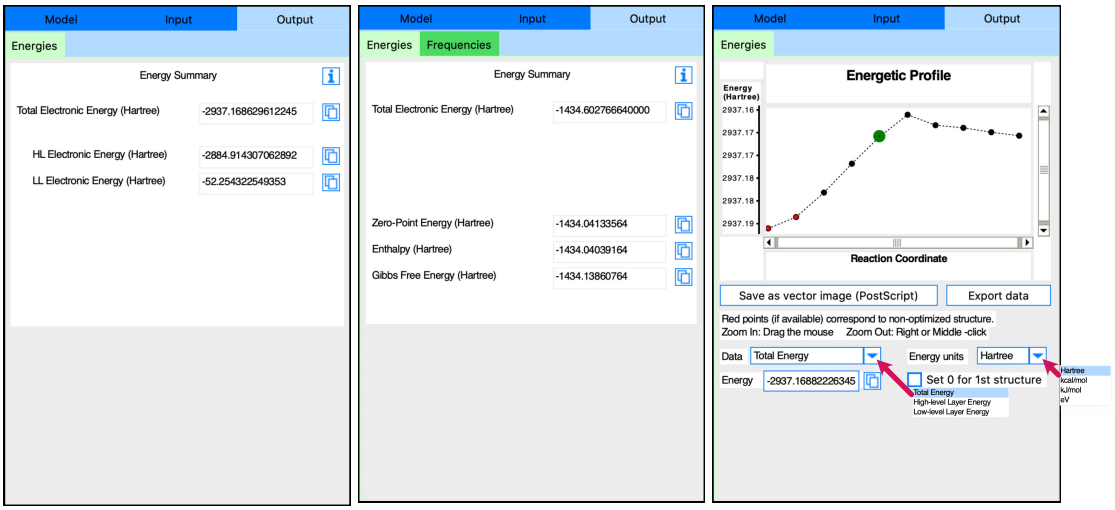


Figure 99. Screenshot of different layout of the “Energies” subtab of the “Output” section of molUP. In the case of multiple structures (right), the plot’s adjustments are shown.

Another subtab included in the “Output” section offers a set of tools to analyze vibrational frequencies (Figure 100). Usually, vibrational frequencies are computed to complete and validate the characterization of minima. For reactants and products, it is expected that no imaginary frequency is reported in the calculation. However, the TS should solely one imaginary frequency associated with the vibration of the atoms involved in that particular chemical reaction.

[16] Only available for multilayered calculations

In the case of this type of calculation has been requested, molUP automatically checks for it and imports the computed values in a table containing three columns: “Frequency number”, “Frequency value (cm^{-1})”, and “Wavelength” in the infrared. The table is interactive, allowing users to click on each line to select the frequency (Figure 100). Straightaway, molUP animates all the atoms related to the selected frequency. Then, users can adjust the spend and amplitude of the animation or control the animation using the play, pause, and stop buttons. Additionally, molUP allows representing all the computed vectors associated with the vibrational frequencies. The vectors are customizable in terms of color, size, and minimum threshold.

The visual confirmation of an imaginary frequency associated with a TS is essential to ensure that the correct TS was isolated. In some cases, small imaginary frequencies are computed for TSs, but they do not correspond to the atoms involved in the break and formation of chemical bonds.

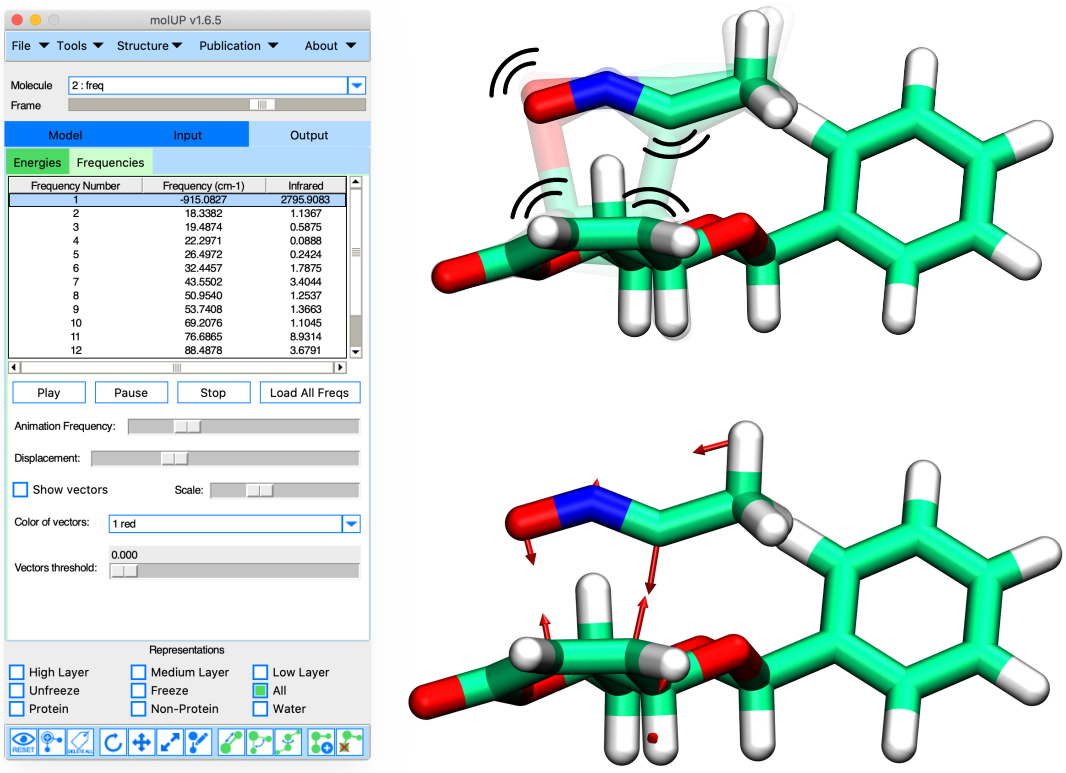


Figure 100. (Left) Screenshot of the “Frequencies” subtab of the “Output” section of molUP. (Right) Representation of the vibrational frequency animation and related vectors.

d. Other Features

Since molUP is integrated into VMD it can take advantage of all the built-in features present in VMD, including several plugin extensions. This means that all the VMD

extensions available for VMD can increase the potential of using molUP to analyze Gaussian calculations. Some analyses, including hydrogen-bond networks, RMSDs, Ramachandran plots, or the alignment of different amino acid sequences, are simplified by molUP and its integration with VMD.

molUP also includes a tool to measure bond lengths and the amplitude of angles and dihedrals. This tool is available under the name "Measure Bonds, Angles, and Dihedral" in the "Tools" menu. Once the tool is called, a new window pops up, and users have to select the desired atoms. At this time, the information regarding the index, type, and resid are filled in as well as the distance and angles that are measured.

Since VMD has a built-in plugin to create movies, the animation of vibrational frequencies or the pathway across a scan can be exported as a video. Moreover, since molUP is able to load the Gaussian-related files in VMD, high-resolution images can also be rendered using the VMD features.

4.3.4. Conclusion

The primary motivation to develop molUP was enhancing the handle of Gaussian-related input and output files. The development of this plugin took into account the vast experience of the research group in QM and QM/MM calculations. This is not a competitive application but rather to complement the existing ones, in particular, offering a free option.

Particular attention was given to the issues that often arise when analyzing and preparing Gaussian input files that occupy several gigabytes of disk space and contain molecules with thousands of atoms. With molUP, some of these issues were overcome, and all the previously cumbersome tasks are now easy and seamless.

A set of tutorials were also prepared and published online (https://biosim.pt/molup/) to teach and introduce the QM and QM/MM calculations using molUP to inexperienced users. The tutorials were assessed more than two thousand times[17] worldwide during the last two years.

[17] Data from Google Analytics reports

4.4. Energy Split

Energy Split is a software devoted to energy partitioning applied to the AMBER force field. This idea was discussed with Professor André Melo, who provided insights about the strategies to accomplish the MM energy partitioning. The main goal of the project was developing a software able to split a molecular system in different parts and compute their energies independently.

Henrique S. Fernandes developed the initial version of the software in Tcl, which includes a VMD extension to generate the input files from an MD simulation or minimization run with AMBER. The VMD extension also includes the capability to link molUP and generate input from a Gaussian file with AMBER parameters. The actual calculation software was prepared to parallelize the computation of the non-bonded interactions.

Although the software is able to compute the energies for most of the interesting systems in a reasonable amount of time, it is not capable of doing that fast enough to be applied in several hundred or thousands of structures. Therefore, this project was improved in collaboration with Sara Pereira, André Pereira, and Professor Alberto José Proença from the University of Minho. The implementation of the software in C++ language allowed a speedup of about 600 times.

The following section will be focused on the main principles behind this software and the initial Tcl development accomplished by Henrique S. Fernandes.

This page was intentionally left blank.

4.4.1. Introduction

As discussed in chapter B, the MM methods are widely used in a panoply of computational studies. In particular, they are usually applied in large molecular systems where more accurate methods cannot be employed. For the study of catalytic mechanisms, MM Hamiltonians are employed in MM minimizations, MD simulations, and QM/MM calculations.

During an MD simulation or between steps of a reaction studied with QM/MM, the energy of the system varies as a consequence of the rearrangement of the atoms. Those fluctuations can be more or less extensive and depend on the sum of all contributions. Unfortunately, in such big systems, it is difficult to know which molecules or groups of atoms are responsible for the major contribution to the energy variation. Therefore, a scheme where the energy of the system could be split into different parts would help to assess which atoms or molecules are contributing the most.

In the course of the research conducted during this Ph.D. program, two main situations were identified to potentially benefit from an energy partitioning approach. This could be helpful to understand the contribution of certain amino acid residues to the interaction with a ligand or to identify which residues are contributing to huge fluctuations in the MM energy between minima of the same reaction in a QM/MM study.

Considering an enzyme with an inhibitor placed at the active site, different active site residues have distinct roles in the interaction with the inhibitor. Therefore, it would be relevant to account for the energy contribution of each residue to that interaction, depicting the most important interactions and identify possible points of destabilization. This information could help in identifying better inhibitor derivatives.

The procedure of studying a catalytic mechanism using a QM/MM approach can also advantage from using a MM energy partitioning technique. In the study of a catalytic mechanism, three structures are identified for each step of the mechanism: reactant, TS, and product. The energy difference between TS and reactant, and product and reactant, offers the Gibbs free energy of activation and reaction, respectively. Ideally, the QM part must contribute the most to these two quantities since it is the region that includes the atoms involved in the reaction and treated with the highest accurate method. Unfortunately, in some models, the MM part assumes a huge contribution (more than 10 kcal/mol) for the activation or reaction Gibbs free energies, or both. In these cases, it is desirable that the molecular entities involved in that energy variation could be included in the QM region to be accurately accounted. Since the molecules responsible for that variation are not known, their inclusion in the QM region is accomplished by trial test.

However, the lack of reasoning on this approach leads to multiple unfruitful computer-demanding calculations. Once again, the application of a MM energy partitioning technique to the chemical structures of interest would help to identify the species that account the most for that energy difference. Consequently, it would facilitate the identification of the molecules that might be considered in the QM region.

In order to provide an accessible tool to address the aforementioned issues, Energy Split software was developed using Tcl language (version 8.6) and Tcl threading technology to parallelize the most time-demanding calculations.

4.4.2. Development

a. Implementation

In the course of this thesis, the MM calculations were always conducted using the AMBER force field. The MM parameters were applied as implemented in Gaussian software, where for each atom type (i.e., "HP"), the atomic radius (R in Å) (i.e., "1.1000") and the well-depth (ε in kcal/mol) (i.e., 0.0157) are included in the following format:

```
VDW HP    1.1000    0.0157
```

These atomic parameters are essential for the calculation of the non-bonded terms: Lennard-Jones and Coulomb potentials. In the case of the Coulomb potential, the atomic charge for each atom is also required.

Then, for each pair of atom types (i.e., "CT" and "HC"), which appear in the system covalently bonded to each other, the bond force constant (k_b in kcal/mol/Å^2) (i.e., "340.00") and equilibrium bond length (l_0 in Å) (i.e., "1.0900") are included:

```
HrmStr1 CT HC 340.00 1.0900
```

The approach is similar to the angular bend between sets of three atoms. The trio of atom types (i.e., "C", "N", and "H") are associated with the respective angular force constant (k_a in kcal/mol/rad^2) (i.e., "50.00") and equilibrium angle value (θ_0 in °) (i.e., "120.0001"):

```
HrmBnd1   C   N   H   50.00 120.0001
```

In the case of dihedral angles, each set of four atoms sequentially bonded (i.e., "NJ", "CJ", "ND", and "CK") is included in the parameters' section like:

```
AmbTrs NJ CJ ND CK   0 180 0   0    0.000   4.750   0.000   0.000 1.0
```

The following four values ("0", "180", "0", and "0") correspond to the phase offsets (A_i in °). For each phase offset, a magnitude factor (k_d in kcal/mol) is also available (i.e., "0.000", "4.750", "0.000", and "0.000"). The last value corresponds to the number of paths (N) (i.e., "1.0").

Finally, the improper torsions are calculated based on the following MM parameters:

```
ImpTrs CJ NJ CD NH    1.1   180.0 2.0
```

The improper torsion is calculated as the displacement of one atom in relation to a plane defined by three other atoms (Figure 9). Therefore, four different atoms are needed to define an improper torsion (i.e., "CJ", "NJ", "CD", and "NH"). The energy contribution for an improper torsion is calculated considering a magnitude factor (k_i in kcal/mol) (i.e., "1.1"), a phase offset (A in °) (i.e., "180.0"), and a period (P) (i.e., "2.0").

Based on these parameters that are available on every Gaussian input file for AMBER calculations, the Energy Split was developed to use that and apply the respective energy equation:

$$E_{System} = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_a(\theta - \theta_0)^2 + \sum_{dihedrals} \sum_{i=1}^{4} \frac{k_d(1 + \cos(i.\phi - A_i)}{N} + \sum_{impropers} k_i(1 - \cos(P(\sigma - A))) + \sum_{i<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) \times S_{LJ} + \sum_{i<j} \frac{1}{4\pi\epsilon_0} \left(\frac{q_i q_j}{r_{ij}}\right) \times S_{Coulomb}$$

Equation 72. Energy of the system according to AMBER force field.

Based on the connectivity data, it is possible to identify which atoms are bonded to each other. This information is essential to calculate the bonded terms: bonds, angles, dihedrals, and improper torsions. This is also crucial to apply the appropriate scaling factors to the Lennard-Jones (S_{LJ}) and Coulomb ($S_{Coulomb}$) potential.

For each bond, the Energy Split calculates the arithmetic distance (l) between the respective ij pair of atoms:

$$l = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}$$

Equation 73. Distance between two atoms ij.

Then, this distance (l), the equilibrium bond length (l_0), and the bond force constant (k_b) are used to calculate the energy contribution of each bond in the system.

The angles are calculated following a similar approach, where the first step is to calculate the angle between three atoms ijk. Firstly, the vectors $\vec{ji}$ and $\vec{jk}$ are determined, and then the angle between them is calculated following the dot product expression:

$$\cos(\theta) = \frac{\vec{ji} \cdot \vec{jk}}{\|\vec{ji}\| \|\vec{jk}\|}$$

Equation 74. Angle between two vectors $\vec{ji}$ and $\vec{jk}$.

The calculated angle (θ) enters in the energy expression with the equilibrium value (θ_0) and respective equilibrium value (k_a).

Dihedral angles are considered every time four atoms ($ijkl$) are covalently bonded in a sequence. The dihedral angle is, by definition, the angle established between the ijk and

jkl planes. The calculation was accomplished through the determination of the $\overrightarrow{ij}$ and $\overrightarrow{jk}$ vectors for the ijk plane, and the $\overrightarrow{kl}$ and $\overrightarrow{kj}$ vectors for the jkl plane. The cross product was used to compute the norm vectors for each plane. Finally, the angle between the two norm vectors is calculated using the same approach mentioned before. Starting with that value, the energy contribution for the dihedral angle is calculated as presented in the respective term in Equation 72.

For the improper torsion, all the atoms that are exactly bonded to other three atoms are considered for this calculation. An improper torsion occurs when three atoms ijk are sequentially connected to each other, and l atom is connected to atom j. The improper torsion was developed to provide an energy penalty when atom l goes out of the ijk plane. The penalization is proportional to that displacement, according to the respective term in Equation 72. Similar to the dihedral angle calculation, two planes are defined: ijk and ikl planes. Then, the norm vectors are calculated and the angle between them are calculated as mentioned before.

So far, it was described how the bonded terms are calculated in the Energy Split implementation. Most of the operations were implemented using the "math::linearalgebra" module of tcllib (version 1.19) available at https://core.tcl-lang.org/tcllib/doc/trunk/embedded/index.md.

Finally, the non-bond interactions are calculated considering all atom pairs of the system since no cut-off was applied. However, this part of the calculation is highly time-demanding and was parallelized through the Thread module for Tcl (https://www.tcl.tk/man/tcl8.6/ThreadCmd/thread.htm). The calculation of non-bond interactions was split into the number of available threads of the running machine.

The Coulomb term for each pair of atoms is simply calculated by the product of the atomic charges of those atoms (q_i and q_j), divided by the distance between them, and multiplied by two constants. In a MM calculation, the atomic charges are known and kept fixed through the calculations, so only the distance between the atoms need to be calculated according to Equation 73. Since ϵ_0 is the vacuum permittivity, the term $\frac{1}{4\pi\epsilon_0}$ is constant and equal to 332.063712827427 kcal.mol^{-1}.Å. The Coulomb contribution is also affected by a scaling factor ($S_{Coulomb}$) that depends on how the atoms are connected to each other. In the case of atoms spaced by one or two covalent bonds, the scaling factor is zero because the interaction between those atoms is already accounted by the bond and angle terms (described before). Regarding the atoms spaced by three covalent bonds, the scaling factor is $1/1.2$, following the implementation in Gaussian 09. All other case scenarios are entirely accounted, so the scaling factor is 1. These values are purely

empirical and were set to reproduce more accurate calculations and/or experimental data.

The scaling factors are also employed for the calculation of Lennard-Jones potential. In this case, the atoms spaced by up to two covalent bonds are nulled, while the atoms spaced by three covalent bonds are scaled by 0.5. The scaling for the remaining pair of atoms (S_{LJ}) is 1. Similar to the Coulomb expression, the Lennard-Jones potential depends on the distance between the two atoms involved (ij). Therefore, the distance calculated for the Coulomb term is used in the determination of the Van der Waals contribution. However, the expression for the Lennard-Jones potential required two additional parameters for each pair of type of atoms: A_{ij} and B_{ij}. These two terms are calculated according to the following equations:

$$A_{ij} = \varepsilon_{ij}.D_{ij}^{12}$$

Equation 75. Calculation of the A_{ij} parameter.

$$B_{ij} = 2\varepsilon_{ij}.D_{ij}^{6}$$

Equation 76. Calculation of the B_{ij} parameter.

From the above equations, two variables are required for each pair of atoms. Those variables are calculated based on the atomic radius (R_i and R_j) and well-depth (ε_i and ε_j) for each atom type. This data is also included in the MM parameters, as mentioned at the beginning of this section. The e_{ij} and D_{ij} terms are calculated according to the following expressions:

$$\varepsilon_{ij} = \sqrt{\varepsilon_i.\varepsilon_j}$$

Equation 77. Calculation of the ε_{ij} parameter.

$$D_{ij} = R_i + R_j$$

Equation 78. Calculation of the D_{ij} parameter.

So far, the MM energy function of AMBER was detailly described, including all the mathematic expression required for the implementation. It is also important to notice that the MM paraments stored by Gaussian software are not consistent in terms of units. Therefore, the actual code implementation in Energy Split involved some conversion of units.

The implementation of the MM energy function in Energy Split was tested and compared with the results obtained with Gaussian 09 for the exact same systems. A set of different

systems were used to compare the computed energies of each individual term: bond, angles, dihedral angles, improper torsions, Van der Waals and Coulomb interactions. After this validation, the Energy Split was prepared to split the systems into as many fragments as the user requested.

In order to clarify the implementation of the energy partitioning in the Energy Split, a N atoms system is split into two fragments (Figure 101).

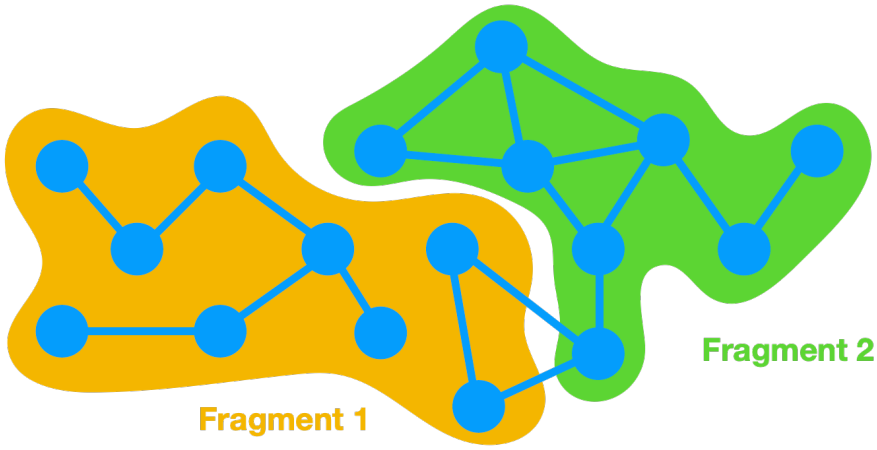


Figure 101. Example of a hypothetical molecular system split into two different fragments. The circles and lines represent atoms and covalent bonds, respectively.

The energy partitioning implicates the accounting of the different energy terms for the different fragments as separate sums. Therefore, in the case of a two-fragments system, two energy sums are contemplated: E_{frag1} and E_{frag2}. The energy of fragment 1 (E_{frag1}) accounts for the energy of all terms (Equation 72) involving only atoms belonging to fragment 1. The same is applied for fragment 2 and all other possible fragments of the system. In addition, a third term is also required to account for the energy of the interactions between atoms that belong to more than one fragment. A covalent bond can span the boundary of two adjacent fragments and, therefore, it is not considered in any of the cases. For those situations, the energy is accounted in that third term, the interaction energy between fragments 1 and 2: $E_{int(1-2)}$. The same is applicable for all other terms where the involved atoms belong to different fragments, be them angles, dihedral angles, improper torsions, or Van der Waals and Coulomb interactions.

Even in more complicated systems, with multiple fragments, the number of interaction terms depends on the interactions between all the atoms involved. Consequently, it is totally common to find $E_{int(1-2)}$, $E_{int(2-3)}$, $E_{int(1-3)}$, $E_{int(2-4)}$… The dual fragment interaction term can accommodate all the interactions between terms that only involve two atoms (bond term, Van der Waals and Coulomb interaction). However, angles, for example, can span across three different fragments, and dihedral angles and improper

torsions can involve atoms from four different fragments. In such cases, trio (e.g., $E_{int(1-2-3)}$) and quadruple (e.g., $E_{int(1-2-3-4)}$) interaction terms are considered.

The total energy of the system is given by the sum of all these contributions:

$$E_{system}^{N-frag.} = \sum_{i=1}^{N} E_{frag(i)} + \sum_{\substack{i=1 \\ i \neq j}}^{N} E_{int(i-j)} + \sum_{\substack{i=1 \\ i \neq j \neq k}}^{N} E_{int(i-j-k)} + \sum_{\substack{i=1 \\ i \neq j \neq k \neq l}}^{N} E_{int(i-j-k-l)}$$

Equation 79. Total energy of the system according to Energy Split.

The Energy Split calculates the MM energy of a molecular system in the same manner as the standard MM software but offers a straightforward way to split the system into different fragments. As result, the energy is calculated in terms of fragments and the interaction between them.

b. GUI

A simple GUI was also developed to help users defining the fragments for the energy partitioning calculation. The GUI was developed using Tcl/Tk language and acts as a VMD extension (**Figure 102**). The VMD extension was prepared to automatically convert the AMBER parameters, from a PRMTOP file to the format employed by Gaussian 09 and required for further calculations. Consequently, the user only needs to load the molecular system of interest in VMD as an AMBER system. It can be either an MD simulation or an energy minimization, for example. Then, the VMD extension is used to set up the fragments that need to be considered in the energy partitioning calculation. The fragments are defined using VMD atom selections, allowing the user to inspect the fragments visually. The “X” fragment is always available by default, and it corresponds to the remaining molecular system that does not belong to any fragment (**Figure 103**).

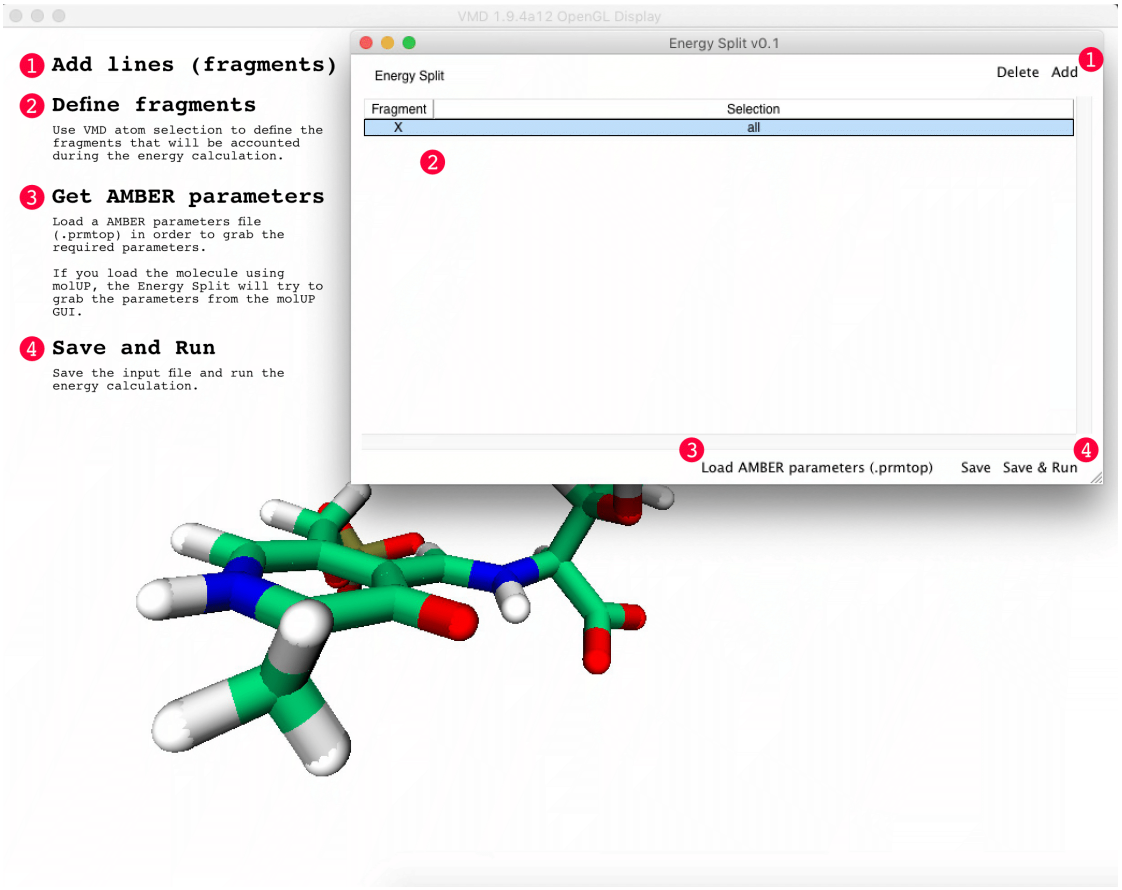


Figure 102. Screenshot of the Energy Split extension for VMD. The labeled numbers and associated text provide quick instruction on how to use the extension.

For example, if a system has three amino acid residues (residue 0, 1, and 2), and the user chooses a fragment containing residue 1, the fragment “X” will contain the residues 0 and 2.

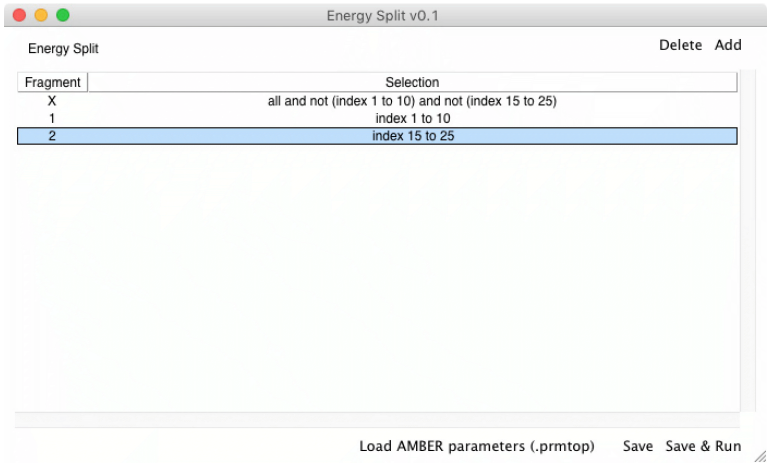


Figure 103. Screenshot of the Energy Split extension for VMD showing a system split into three fragments.

Afterward, the AMBER parameters file (PRMTOP file) needs to be loaded within the Energy Split extension window. This process ensures that the MM parameters are

evaluated and prepared for the subsequent calculation by Energy Split. The PRMTOP file used in this step should be the same used to load the molecular system in VMD.

Finally, the user can choose to save the Energy Split input file to start the calculation out of VMD or on another machine, or simply click “Save & Run” to save the input file and immediately start the calculation.

The Energy Split extension for VMD was also developed to work seamlessly with molUP plugin. Therefore, users can load Gaussian input/output files of QM/MM or MM calculation using molUP, and then define fragments using Energy Split extension. In this scenario, there is no need to load AMBER parameters from a PRMTOP file since the MM parameters are automatically grabbed from molUP.

c. Calculation’s Output

An Energy Split calculation can be started directly from the Energy Split extension for VMD, clicking on “Save & Run”. Otherwise, users can save the Energy Split input file and start the calculation in the terminal running the following command:

```
tclsh energySplitPath/energySplitCalculation.tcl <inputfile>.tcl
```

Energy Split will automatically detect the number of available CPU cores to parallelize the most time-demanding tasks. Once the calculation is finished, the output of the calculation has the following structure:

```
################################################################################
### Energy Split
################################################################################
### Output
################################################################################

### Energy Split detected 8 cpu(s) available.

####################
### Fragments
####################
        Fragment 0: 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395
        Fragment 1: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
91 92 93 94 95 96 97 98 99 100
####################
## Bonds
####################
        Energy Decomposition:
                Fragment 0 + Fragment X                                 1.1175408337120441e-5   Hartree
                Fragment 0                                              0.001192631648569113    Hartree
                Fragment 1 + Fragment X                                 3.1963997239640406e-5   Hartree
                Fragment 1                                              0.005877925197235168    Hartree
                Fragment X                                              1.3487114819893853      Hartree
        Total Energy: 1.3558251782407664 Hartree

####################
## Angles
####################
        Energy Decomposition:
                Fragment 1 + Fragment X                                 0.003909277948827497    Hartree
                Fragment 0                                              0.005346621078139691    Hartree
                Fragment 0 + Fragment X                                 0.0009422968362060424   Hartree
                Fragment 1                                              0.01587722526327147     Hartree
                Fragment X                                              2.840380163313124       Hartree
        Total Energy: 2.866455584439569 Hartree

####################
## Torsions
####################
        Energy Decomposition:
                Fragment 0 + Fragment X                                 0.027413345672736608    Hartree
                Fragment 1 + Fragment X                                 0.019303997109712725    Hartree
                Fragment 0                                              0.000960452740606765    Hartree
                Fragment 1                                              0.07027831060095976     Hartree
                Fragment X                                              11.779855376889085      Hartree
        Total Energy: 11.8978114830131 Hartree

####################
## Out-of-plane
####################
/!\ WARNING /!\ Missing parameter for improper CA-CB-CN-NA (indexes 160 168 159 157)
/!\ WARNING /!\ Missing parameter for improper CA-CN-CB-C* (indexes 166 159 168 154)
/!\ WARNING /!\ Missing parameter for improper H-H-N3-CT (indexes 3365 3366 3364 3361)
/!\ WARNING /!\ Missing parameter for improper CA-CB-CN-NA (indexes 5713 5721 5712 5710)
/!\ WARNING /!\ Missing parameter for improper CA-CN-CB-C* (indexes 5719 5712 5721 5707)
```

```
/!\ WARNING /!\ Missing parameter for improper CA-CB-CN-NA (indexes 6390 6398 6389 6387)
/!\ WARNING /!\ Missing parameter for improper CA-CN-CB-C* (indexes 6396 6389 6398 6384)
/!\ WARNING /!\ Missing parameter for improper H-H-N3-CT (indexes 9595 9596 9594 9591)
/!\ WARNING /!\ Missing parameter for improper CA-CB-CN-NA (indexes 11943 11951 11942 11940)
/!\ WARNING /!\ Missing parameter for improper CA-CN-CB-C* (indexes 11949 11942 11951 11937)
/!\ WARNING /!\ Missing parameter for improper H5-CY-NG-CF (indexes 12484 12485 12483 12481)
/!\ WARNING /!\ Missing parameter for improper H5-CY-NG-CF (indexes 12522 12523 12521 12519)
        Energy Decomposition:
                Fragment 0 + Fragment X                                 7.384437533533025e-5    Hartree
                Fragment 1 + Fragment X                                 1.843132775039692e-7    Hartree
                Fragment 0                                              0.0002228953732006277   Hartree
                Fragment 1                                              0.000930847728797576    Hartree
                Fragment X                                              0.14035946356321405     Hartree
        Total Energy: 0.14158723535382509 Hartree

####################
## VDW
####################
        Energy Decomposition:
                Fragment 0 + Fragment X                                 -0.018663926510680447   Hartree
                Fragment 0                                              0.0005457981739828601   Hartree
                Fragment 1 + Fragment X                                 -0.05217163842479787    Hartree
                Fragment 1 + Fragment 0                                 -1.8613620853836156e-6  Hartree
                Fragment 1                                              0.0001458846864959111   Hartree
                Fragment X                                              -7.088191728714288      Hartree
        Total Energy: -7.1583374721513735 Hartree

####################
## Coulomb
####################
        Energy Decomposition:
                Fragment 0 + Fragment X                                 0.08737279750499699     Hartree
                Fragment 0                                              0.05086567722001398     Hartree
                Fragment 1 + Fragment X                                 -0.37673283820791265    Hartree
                Fragment 1 + Fragment 0                                 0.0015886339333808883   Hartree
                Fragment 1                                              -0.14002814587392498    Hartree
                Fragment X                                              -44.06932366374086      Hartree
        Total Energy: -44.44625753916431 Hartree

####################
## Total Energy
####################
        Energy Decomposition:
                Fragment 0 + Fragment X                                 0.09714953328693164     Hartree
                Fragment 1 + Fragment X                                 -0.40565905326365315    Hartree
                Fragment 0                                              0.059134076234513035    Hartree
                Fragment 0 + Fragment 1                                 0.0015867725712955048   Hartree
                Fragment 1                                              -0.04691795239716509    Hartree
                Fragment X                                              -35.04820890670034      Hartree
        Total Energy: -35.342915530268414 Hartree

##############################################################################
### The calculation finished successfully in 2019_Oct_28 at 17:03:50.
##############################################################################
### Developed by Henrique S. Fernandes (henriqueferl1@gmail.com)
##############################################################################
```

The output provides information about how many CPU cores were detected and used during the calculation. Then, a “Fragments” section lists all the fragments considered in the calculation, where each fragment is followed by a list of atoms included in that particular fragment (atom indexes). In the above example, two fragments were defined by the user (Fragments 0 and 1). Therefore, the calculation was accomplished considering three fragments, fragments 0, 1, and X.

Then, four sections are presented corresponding to the energy calculation for all bond terms: bonds (Bonds), angles (Angles), dihedral angles (Torsions), and improper torsions (Out-of-Plane). The improper torsions appeared with some warnings about missing parameters, which are a consequence of how improper torsions are evaluated in Energy Split. Since Energy Split considers all atoms bonded to exactly three other ones as potential improper torsions, some of those calculations are not accomplished due to lack of MM parameters. The warning is displayed only to advise users of such situation, and those atoms might be carefully checked.

After these four sections, the energy for Van der Wall (VDW) and Coulomb (Coulomb) interactions are displayed. Finally, the total energy of the system (Total Energy) is presented.

For all sections, the energy is presented following the energy partitioning described by **Equation 79**. The energy of each fragment is presented as “Fragment *N*”, where *N* can

be “0”, “1”, or “X”. In addition, the interaction energy, which corresponds to energy terms that involve atoms from different fragments, are represented as “Fragment *N* + Fragment *M*”, where *N* and *M* can be “0”, “1”, or “X”, and $N \neq M$.

4.4.3. Conclusion and Future Perspectives

The Energy Split is a software for energy partitioning using the AMBER energy function. The main goal is to provide a quick and easy to use tool to evaluate the contribution of different parts of a molecular system for its total energy.

Currently, the software is in the late stage of development of the C++ version. This optimized version of Energy Split will speed up the calculation and make it even better for quick inspection of multiple structures.

As mentioned, we have identified two potential applications of such type of analysis. In particular, the identification of molecular species that have a strong contribution to the activation and/or reaction Gibbs free energies in QM/MM calculation, or the identification of the most important interaction between two molecules during an MD simulation. Thus, *per si*, the energy partitioning is not so useful, but the energy differences between two geometries of the same system could provide valuable insights. Consequently, the next step involves the development of a GUI that can organize and plot the differences in energies for all the terms of the energy partitioning. In this way, Energy Split could be applied for long simulations of drugs and respective targets, for example. The results would indicate the amino acid residues that account the most for the interaction during the simulation. Moreover, it could also identify possible repulsive interactions that compromise the efficacy of the drug.

In sum, the core of the Energy Split is already developed, and the optimized version (C++) is in its final test phase. The scheduled task involves the development of a GUI for easily comparing multiple calculations for a given system.

The current version of Energy Split and respective VMD extension are available for free at https://github.com/henriquefer/energySplit.

Download
Energy Split

This page was intentionally left blank.

4.5. chemTool

chemTool is a VMD extension, developed by Henrique S. Fernandes, to handle ORCA-related input and output files. In fact, chemTool can be seen as a molUP version for ORCA calculations.

This work has been done in collaboration with Christoph Riplinger and Bernardo Souza from the FAccTs GmbH company. Riplinger and Souza are involved in the development of ORCA software and respective commercial version. Till Rudack, from the Ruhr University Bochum, in Germany, is also helping the chemTool development.

Although all the code has been written by Henrique S. Fernandes, the ORCA developers have been provided feedback and assistance in understanding how ORCA input and output files are organized.

chemTool is still a plugin in active development.

This page was intentionally left blank.

4.5.1. Introduction

ORCA [250] is a powerful software to perform QM calculations, with a panoply of tools for spectroscopic calculations. In 2019, ORCA introduced native support for QM/MM calculations, which was something only possible using coupling schemes where a different software was required to perform the MM calculations. Since then, ORCA is capable of running QM/MM natively, greatly increasing its range of applicability, in particular for the study of enzymatic mechanisms. The new version 5.0 was announced in the last quarter of 2020, and it will include a set of improvements in terms of the QM/MM implementation. Moreover, the promisor DLPNO-CCSD(T) method [232] is only available on ORCA package (Chapter B – 2.3.5. page 63).

Although ORCA is a popular software for QM calculations, with more than 5000 citations according to Scopus[18], it lacks an appropriate GUI for the preparation of input files and analysis of results. So far, ORCA depends on some molecular visualization software, such as Avogadro [462], Gabedit [463], Molden [464], and ChemCraft. However, there is no standard software that can handle all types of ORCA input and output files. Therefore, we started the development of a VMD extension capable of handling ORCA-related files, following the same fashion of molUP for Gaussian-related calculations.

[18] Access date: 2021, Janurary 4th

4.5.2. Development

chemTool was coded in Tcl/Tk to ensure perfect integration as a VMD extension. The implementation of chemTool relies only on a single external dependency, the PlotChart module for Tcl/Tk (version 2.4.1), developed by Arjen Markus. However, this module was included in the distribution of chemTool to make its use seamless across all operating systems capable of running VMD.

4.5.3. chemTool Features

chemTool was inspired by molUP extension (Chapter C - 4.3. page 245), and the main interface and features are common to both plugins (Figure 104). The similarities between both plugins improve the useability when users interchange between both software.

The main GUI of chemTool comprises a top menu bar, a molecule selector, a frame navigator, a main section, a set of quick VMD representations, a log console, and a set of shortcuts and new tools.

The top menu bar consists of some dropdown menus with the core functions of chemTool. In the case of “File” menu, it is possible to load and save ORCA input/output files.

Immediately below the top menu bar, a dropdown menu allows users to quickly exchange between loaded molecules on chemTool. This tool is extremely powerful when analyzing multiple molecular systems, and it is commonly cumbersome in some molecular visualization software. Due to the implementation made in chemTool, users can load huge molecular systems and retrieve maximum performance from VMD.

The frame navigator is a slider that can be dragged to navigate across a system with multiple structures. For example, during a relaxed scan or a molecular optimization, several intermediate structures are generated. They can be loaded on chemTool and visually inspected in the main window VMD.

More than half of the GUIs area is occupied by the main section that is composed of four tabs: Model, Input, Output, and Help. Each tab will be carefully discussed in the following sections.

As implemented in molUP, the chemTool also includes quick VMD representations. The quick representations are simple toggle switches to show and hide different VMD representations that are commonly required in the study of molecular systems. In the study of enzymatic mechanisms by QM/MM methodologies, it is commonly used to

visualize the QM or MM region, or simply the water molecules or the atoms that are free to move during geometry optimizations.

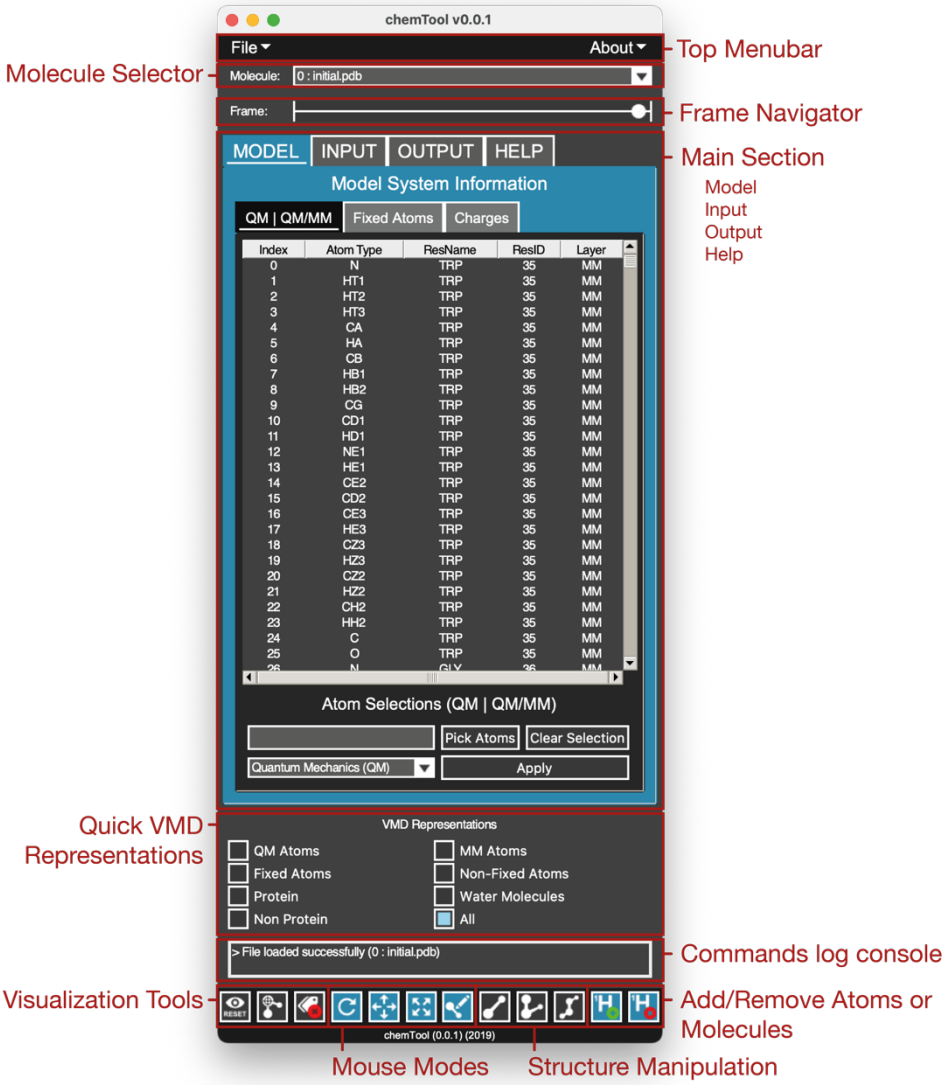


Figure 104. Screenshot of the main GUI of chemTool. The main areas of the GUI are tagged.

The log console is a new implementation that is not currently available in molUP. This is a console that keeps track of all the commands that were executed by chemTool or is being executed. Such information helps users to know if the plugin is busy executing a task or idle and ready to be used.

Finally, chemTool includes a set of new tools and shortcuts to provide quick access to commonly used native VMD commands or activate new tools developed by us.

The visualization tools include two shortcuts for reset display view and center the view on a picked atom. In addition, it also has an icon to execute a built-in script to delete all

the labels. Such type of feature cannot be quickly executed in VMD, and it was included by us to provide such an asset.

The four most used mouse modes were also included in this list of shortcuts to provide quick access to them.

As introduced in molUP, chemTool also includes a set of tools to adjust bond lengths, angle amplitudes, and dihedral torsions. The VMD is a molecular visualization tool and does not include tools to modify the structure of the loaded molecules. Therefore, this set of tools were developed to add an important asset for VMD. In the same line, VMD does not provide native support for deleting or add atoms and molecules. Thus, this feature was also included in chemTool, following the same implementation as in molUP.

a. Model Section

A basic ORCA input file includes the calculation keywords, the spatial coordinates for all the atoms of the system, and the respective charge and spin multiplicity of the system. The calculation keywords and the charge/spin multiplicity information are stored in an input file (".inp").

Regarding the coordinates, there are three manners to save this information for the calculation. The first one is to include a list of four variables for each atom in the ORCA input file (Atomic element, X coordinate, Y coordinate, and Z coordinate). However, ORCA can also read the atomic coordinates from an external XYZ or PDB file. Therefore, chemTool was developed to automatically detect which format is being used when an input or output is loaded.

Thus, users can load ORCA input and output files clicking on "File" > "Open". In the pop-up window, they need to browse the ".inp" file of the calculation of interest and select if they want to load the input file or the output files. In the case of selecting the "Output file(s)", additional options are unlocked. In particular, users could choose to load the last structure (default), all structures available, or just the optimized ones in the case of a relaxed scan (**Figure 105**).

Moreover, chemTool can analyze all types of molecular systems loaded in VMD through its native support. This means that every file format that can be loaded using VMD is compatible with chemTool. Users can load any compatible system using VMD, and chemTool will automatically analyze and grab all the required information to populate the respective GUI. chemTool is also able to save an ORCA input file from those systems previously loaded by VMD.

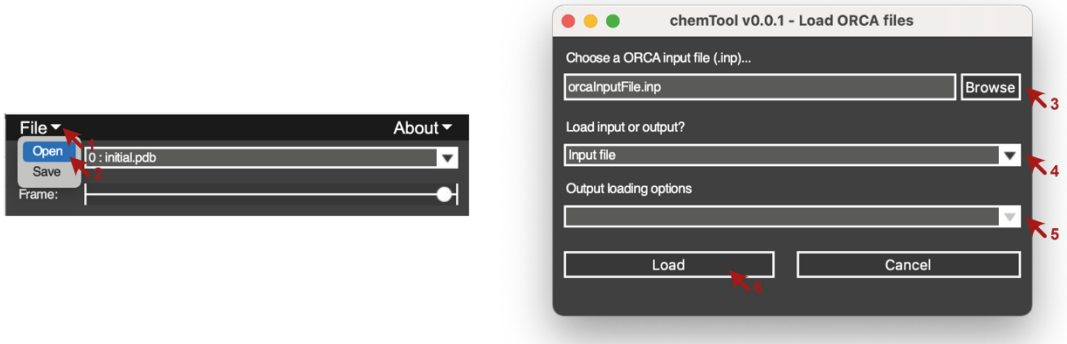


Figure 105. chemTool guide to load ORCA-related files.

Since the molecular system is loaded in VMD, either via VMD or chemTool, all the embedded tools and features available on chemTool can be used. In particular, users gain the possibility to adjust structural parameters like bond lengths, angle amplitude, and dihedral torsions (Figure 106).

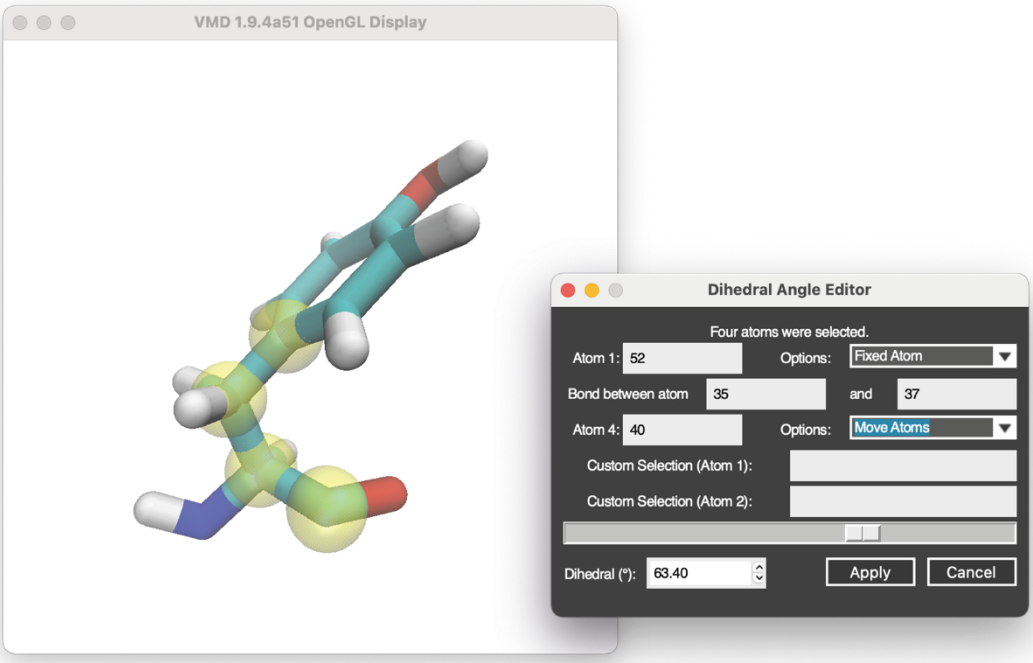


Figure 106. Adjusting a dihedral torsion using the “Dihedral Angle Editor” tool available on chemTool.

Additionally, chemTool will also include the possibility to delete and add atoms/molecules using the same implementation as molUP.

As mentioned, chemTool provides a set of tools to adjust and prepare the geometry of the molecules for the subsequent calculations. However, in every QM or QM/MM

calculations, there are also additional settings that need to be defined prior to the calculation itself.

The “QM | QM/MM” tab shows a list of all the atoms of the system, showing information about index, atom type, residue name, residue ID, and layer (**Figure 107** A). The layer is an essential field for QM/MM calculations because it allows users to assign different atoms to the QM or MM layers. In chemTool, users can select atoms in the table, pick atoms in the main window of VMD, or use VMD selections to assign a certain layer to an atom or group of atoms. These three different manners of assigning QM or MM layers to the atoms of the system make chemTool extremely versatile when users need to prepare QM/MM calculations.

A second tab called “Fixed Atoms” follows the same layout of the “QM | QM/MM” tab but the last column was replaced by another one where the atoms can assume a “Free” or “Fixed” state (**Figure 107** B). This option is important for both types of calculations (QM and QM/MM), since it instructs the ORCA software about which atoms are allowed to move during geometry optimizations. When this option is employed, users can choose atoms that must be fixed or free to move during geometry optimizations. The procedure to select the atoms is exactly the same described before for the “QM | QM/MM” tab.

Finally, the last tab of the main “Model” section is related to the MM charges of the atoms (**Figure 107** C). The atomic charges are important for MM calculations, and they are shown in this tab, if available. Here, users are able to check their values and modify them accordingly. Moreover, it will be implemented a procedure to automatically import new charges from the output of another ORCA calculation. This feature could help users to improve the accuracy of subsequent MM calculations through the update of their atomic charges.

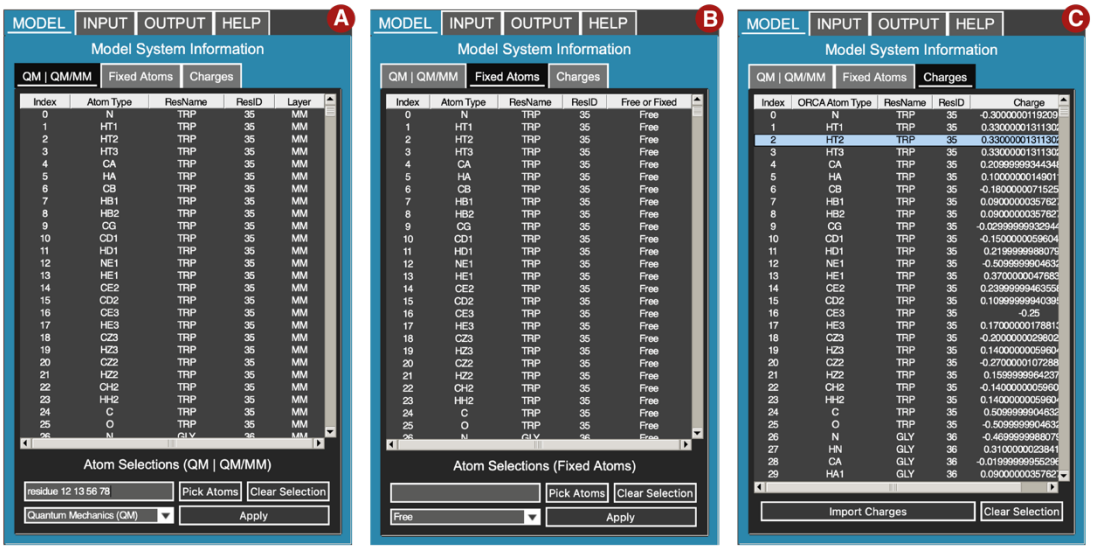


Figure 107. Screenshot of the "QM | QM/MM", "Fixed Atoms", and "Charges" tabs of the main "Model" section of the chemTool.

b. Input Section

Until now, it has been described all the steps required in the preparation of the molecular system and respective computational model. An additional step is required before saving the ORCA input file and start the calculation – the preparation of the input file. In chemTool, users can navigate to the "Input" main section to make the final steps in the preparation of the input file (Figure 108).

This section presents an editable text area where users can include all the required keywords to perform the desired calculation. This section includes a set of calculation templates and allows users to save their own templates. In this way, users can easily re-use the same calculation keywords for different structures, avoiding repetitive typing of all instructions and possible mistakes. The calculation templates are exceptionally important in the study of catalytic mechanisms with several steps, where multiple identical calculations are performed for different minima.

Immediately below, chemTool shows a "Charge and Spin Multiplicity" section where users need to introduce the charge and spin multiplicity of the QM region in a QM/MM calculation or the entire system in purely QM calculations. There is no need to provide this kind of information for MM calculations since the charge is automatically retrieved from the atomic charges. This section also includes an "Options" menu to automatically calculate the charge of the system (if the atomic charges are available) and represent positively and negatively charged molecules in the display window of VMD.

Finally, chemTool shows tools to automatically import MM parameters from different force fields, as AMBER, and convert them to the internal format of ORCA. This feature

will be implemented as soon as ORCA 5.0 is released since it will depend on the implementation of ORCA for the different MM force fields.

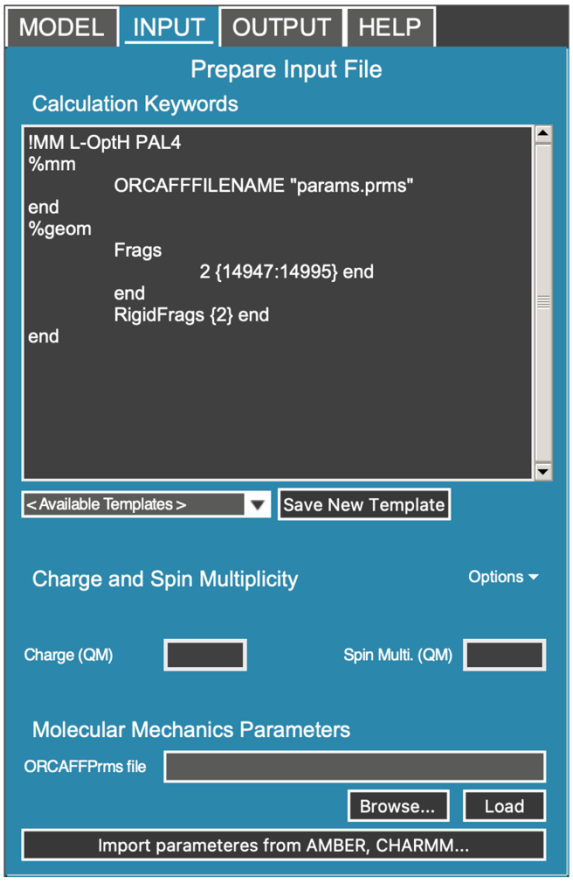


Figure 108. Screenshot of the “Input” main section of the chemTool.

The “Help” main section will provide assistance during the entire process, where guidelines will be presented for each feature and tools available on chemTool. Moreover, the “Help” section will also include sections dedicated to explaining how to use some ORCA keywords and how to overcome frequent issues that can arise from the calculations.

After the assigning of all keywords for the calculation, users only have to save the ORCA input file. To do so, users need to click in “File” > “Save” and then select how they want to store the structural data: PDB or XYZ file (Figure 109). The PDB file is recommended for QM/MM calculation of huge molecular systems like enzymes since it includes information about chains, residue names, and IDs. For small systems and QM calculation, where no additional data is required or available, the XYZ file is the best option. After selecting the file format for the atom coordinates, users are asked to choose a name and the location to save the ORCA input file. Afterward, all required files are stored in that location, and the calculation can be launched.

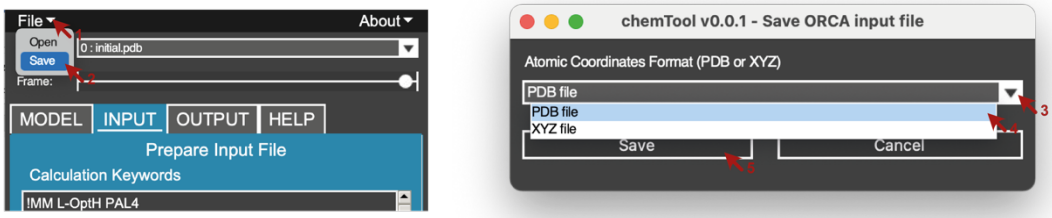


Figure 109. chemTool guide to save an ORCA input file storing the atomic coordinated in a PDB file.

c. *Output Section*

chemTool was also prepared to load and handle the results of an ORCA calculation. The process to load the output of a calculation is the same explained for the input file. Users need to click in “File” > “Open” and select the ORCA input file used in that calculation. Then, the “Output file(s)” option should be selected, and chemTool will automatically analyze all the available output files within the same directory.

When loading the output of an ORCA calculation, users can choose between loading only the last geometry or multiples structure, either all structures available or only the optimized ones in the case of a relaxed scan.

If a single geometry is loaded on chemTool, the “Output” main section will include an “Energy” tab with the total energy computed for that structure (Figure 110).

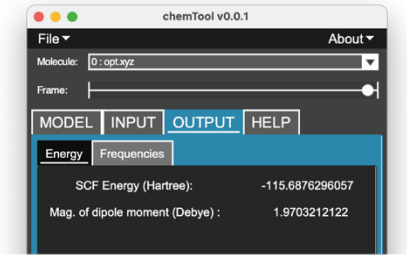


Figure 110. Screenshot of the “Energy” tab of the chemTool.

In the case of frequencies calculations, chemTool also includes the magnitude of the dipole moment in the “Energy” tab. Additionally, a new tab will be available regarding the frequencies calculations (Figure 111). In that tab, users can select any of the computed vibrational frequencies within the shown list. The list presents the value of the frequency in cm^{-1} and the respective infrared value. When a frequency is selected, the chemTool will automatically animate the atoms to vibrate according to the computed frequency. Moreover, users can adjust the displacement and speed of the animation to create

movies according to their needs. In addition, chemTool is capable of representing the vectors associated with the vibration of all atoms involved in the selected frequency. The vectors can also be customized in terms of size, color, and threshold of minimum magnitude to be represented (Figure 111).

The analysis of vibrational frequencies is extremely important in the characterization of all minima and TS of a chemical reaction. A single imaginary frequency should be identified only for the TS of a certain reaction, and it must correspond to the vibration of the atoms involved in the actual reaction.

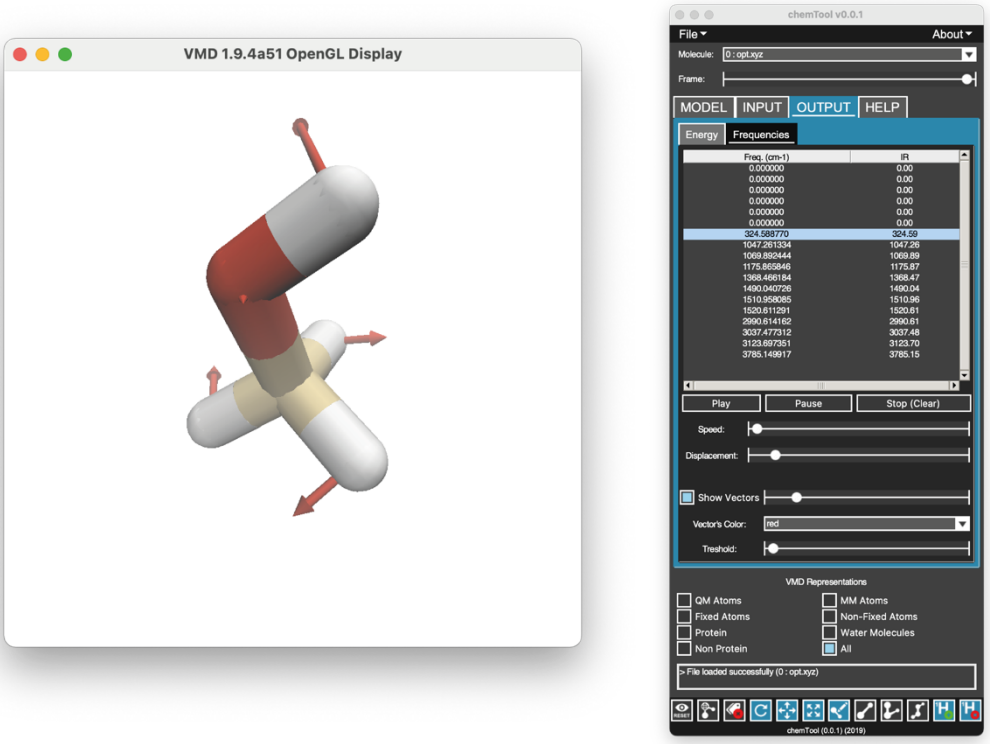


Figure 111. Screenshot of the analysis of vibrational frequencies using chemTool.

As mentioned before, chemTool was developed to handle outputs with multiple structures as well. This is the case of a relaxed scan or a geometry optimization. In these cases, chemTool will load all the geometries available and plot the respective energies in an interactive plot (Figure 112). The interactive plot can be adjusted in terms of energy units (Hartree, kcal/mol, or kJ/mol) and if the first structure should be set zero to make all the remaining points a ΔE in relation to the first structure. Users can also click on the points of the plot to see the respective structure on the display window of VMD. This feature is extremely important to compare different structures according to their relative energies. Moreover, it allows select different structures to proceed with further calculations. In the case of the study of a catalytic mechanism, the relaxed scan can be loaded in chemTool, and the highest energy point can easily be selected to generate a

new ORCA input. Then, the new ORCA input file is used to conduct an optimization to search for a TS.

The importance of the analysis of multiple structures calculations goes behind the isolation of TS guess structures. Sometimes, problematic optimizations can be inspected in terms of energy fluctuations where the optimization algorithm struggles to find the minimum. In those cases, the calculation can be restarted from a different point of the energy profile using or not a different optimization algorithm.

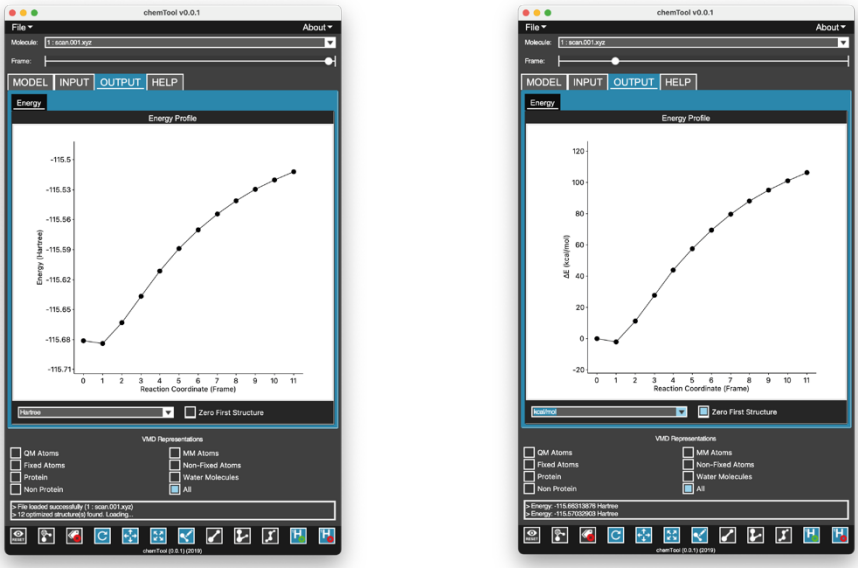


Figure 112. Screenshot of the energy profile analysis conducted using chemTool.

4.5.4. Conclusion and Future Perspectives

ORCA is a powerful software for QM calculations, including several features regarding spectroscopic calculations. Since a couple of years ago, ORCA supports QM/MM calculation natively with no need for external software or scripts to couple the MM part. Although ORCA is free for academic use and provides a panoply of features, it does not include any GUI or standard software for the preparation of input files and subsequent analysis of results. Therefore, chemTool is being developed to fulfill that absence.

chemTool is a VMD extension that takes benefit from all the powerful features of VMD and includes a set of tools that allows the loading of ORCA input and output files. chemTool was also equipped with a panoply of tools dedicated to the analysis and preparation of ORCA-related files.

The ORCA 5.0 will be released soon, allowing the continuation of the chemTool development. The new ORCA version will include some new features, some of them related to the QM/MM implementation, that will allow the integration with the chemTool plugin. The first public release of chemTool is scheduled for 2021, and further developments are already in the roadmap for this VMD extension. Two of the most requested features are the plot of spectroscopic data and the representation of molecular orbitals. These features are being studied for implementation in a later stage of the development. chemTool will be available for free through the VMD Store.

Although chemTool is not publicly available yet, the current version of chemTool can be download at http://bit.ly/chemTool and tested for the purpose of this thesis.

Download
chemTool
(just for evaluation purpose)

5. Pedagogic Applications

The role of a scientific researcher is to apply the scientific method to answer a specific question or provide a solution to a certain problem. Although the fundamental principle seems simple, the process can be hard, requiring perseverance, critical thinking, and much hard work. In addition, researchers are also called nowadays to demisting science before the eyes of society.

In the past, knowledge in general and science were considered elitist subjects that were only accessible to a few highly privileged people. Fortunately, this paradigm changed, and most of the information regarding the scientific advances are available for almost everybody. This accomplishment was possible largely because of the popularization of the internet. The internet's capacity of transforming the huge Earth planet into a small village promotes the information share around the globe.

Besides all the recognized benefits assigned to the internet, there are also some concerns that need to be considered. Not going into detail about privacy and security threats, the internet gave voice to uninformed people that can easily share misinformation with other cybernauts. The phenomena of misinformation and fake news are spread all over the internet and have repercussions regarding society's literacy and the advances of humanity.

Hereupon, scientists also have an important role in sharing quality information with the general public to suppress this misinformation plague. However, the biggest challenge

is how to present highly complex information to non-expert people. Sometimes, scientific discoveries are only truly understood by experts with decades of expertise in the field. How can those new discoveries be explained to non-expert people?

One possible approach is to create interactive and captivating strategies to share more simple concepts and insights about a certain scientific topic, for example. This method can break the first barrier established between the complex knowledge and the non-expert public. Afterward, people become engaged with the content and also develop more curiosity about a specific topic. Through this process, society can be stimulated to acquire basic concepts and interest to gradually pursue more complex information.

During the course of this Ph.D. thesis, some activities were conducted to reach the non-expert society providing a clear vision of how computational (bio)chemistry is being used for the good of humanity. Those activities involved the development of interactive applications aiming public engagement and explaining complex concepts in a simple overview. The main projects include BioSIM^{AR}, Protein Wars, and Criminal Investigation. All of these projects were developed and used in workshops, science demonstrations, and science fairs. The following sections of the thesis will summarize the purpose, development, and application of these projects.

5.1. BioSIM^{AR}

The BioSIM^{AR} is a web-based software that shows tridimensional representations of molecules in an augmented reality (AR) environment. The software was developed by Henrique S. Fernandes. Henrique S. Fernandes wrote the article, which was revised considering the contributions of all the authors. The work was submitted for publication in the Journal of Chemical Education, and the following content is almost a transcription of the submitted manuscript.

This work was also presented in the following conferences:

- (Oral Communication) *VII Encontro da Divisão de Ensino e Divulgação da Química*, **2018**, Escola Almeida Garrett, Vila Nova de Gaia, Portugal
- (Invited Oral Communication) *Biochemistry class*, **2019**, Faculdade de Ciências da Universidade do Porto, Porto, Portugal
- (Oral Communication – Speaker: Sérgio F. Sousa) *Encontro do Ensino e Divulgação da Química*, **2019**, Escola Secundária Avelar Brotero, Coimbra, Portugal
- (Invited Oral Communication) *Evento da Semana Acadêmica de Ciência e Tecnologia – IFRO – Ji-Paraná*, **2020**, Instituto Federal de Rondônia - Campus Vilhena, Paraná , Brazil

This page was intentionally left blank.

5.1.1. Introduction

Learning chemistry involves, to a very significant extent, understanding the relationship between the properties of molecules and their tridimensional (3D) structure [465, 466]. The molecular structure is dependent on the atoms that are present in the molecule, their spatial rearrangement, and how the atoms are connected to each other.

A large set of molecular properties are a consequence or are related to the 3D arrangement of their atoms, including the melting and boiling temperatures, density, solubility, stability, and type of reaction in which it can participate. Concepts such as stereoisomerism, chirality, chemical bonding, hybridization, molecular symmetry, intra and intermolecular interactions, and conformational stability also require a good perception of the 3D shape of molecules [467-469]. For these reasons, many chemistry related subjects courses involve the development of the student the ability to see, think, and relate the structures of molecules with their properties[465, 466]. Usually, this is not an easy task.

Through the years, a variety of alternatives to ease this process has been developed, implemented, and used with success. Notable examples include the Fisher projection [470] for general organic chemistry [471], the Haworth projection [472, 473] for monosaccharides and cyclic molecules, the Newman projections for alkane stereochemistry [470, 472, 474], and the popular wedge-and-dash notation [470, 474] and Natta projection. These two-dimensional representation schemes try to capture part of the 3D structural features of the molecules. They are quite effective in introducing the key concepts of molecular but are not always intuitive to the mind of the student [475-481]. Furthermore, these representations are normally limited to specific sets of molecules of limited size.

Physical molecular models have traditionally played an important role in developing the intuitive understanding of students about the molecular structure and associated concepts. However, the need for acquisition and/or assembly of the models or building blocks makes them of limited applicability also. [482, 483]

A wide range of software applications has been developed to turn the visualization of these molecular models easier, such as Avogadro [462, 484], Molekel [485], Molden [464], Rasmol [486, 487], VMD [292], PyMOL [488], Chimera [459], among many others. Today, with the current computational power behind any personal computer, these molecular models can be visually inspected in an intuitive manner. What needs to be improved is how the users interact with them and how they can extract useful information from it. This is important for students and teachers in high school, universities, or in more

advanced programs, but also for scientists that require this sort of information to perform their research.

The interaction with the molecular models is essential to overcome difficulties associated with the terms and concepts related to the 3D structure of molecules. Several studies have already concluded the advantages of using molecular models when teaching molecular structure [489-491]. For example, 3D printing models can be used to generate real molecular models that can be handled by hand. However, the models are not perfect, are difficult to build, are expensive, and static [482, 483]. Even though 3D printing is becoming cheaper and accessible to the general consumer, it is not so ubiquitous as a smartphone. In 2018, 1.42 millions of 3D printers were sold worldwide [492], about 1000 times less than the number of smartphones (1556.27 million) [493].

Virtual reality devices use glasses that provide an immersive 3D experience, allowing close interaction between the users and the 3D models. Some devices, e.g., Oculus Rift, PlayStation VR, HTC Vive, can cost hundreds of dollars and require complex adjustments to work properly on each setup. Notwithstanding, some related approaches have been described in the literature [494-498]. In spite of the number of features available through this kind of approach, they are considerably more expensive and difficult to operate by non-expert users. Additionally, they are definitely far from reaching the classrooms of most students and teachers.

AR has also been used to provide 3D models of molecules [476, 491, 499-501]. This approach is considerably cheaper, but usually, it relies on the installation of specific software that sometimes has compatibility issues or may require special graphical resources that are often not available. BiochemAR [502] is an example of a recent AR software for molecular representation. However, so far it is limited to compatible iOS and Android devices, making it incompatible with computers and non-iOS and non-Android devices. In addition, BiochemAR presently only holds a potassium channel structure available, providing no way to compare multiple structures or ask for more structures to be included. Similarly, MoleculAR [476] was also developed and published as a framework to create and use AR applications for chemistry teaching. However, the application is presently only available for Android devices, and it involves a time-demanding process to prepare the application and make it suitable for use with students in classes. Augmented Chemical Reactions [503] is another AR software for molecular visualization using AR technology, but currently it lacks support for different devices, since it only works on computers running Windows. The main advantage of BioSIM^{AR}, over these similar applications, is the simplicity in start using it, since no software installation is required, and even low-budget devices can handle it.

The BioSIM^{AR} is an open-source application that was developed, aiming to provide an interactive platform to explore molecular models. The main goal was to develop a versatile and powerful tool that was also easy to set up (Figure 113).

Figure 113. 3D representation of the structure of a DNA sequence in augmented reality using BioSIM^{AR}.

This technology just requires a laptop, a tablet, a smartphone, or a similar device with an integrated camera and an internet connection. No software installation is required. Users only need to download, free of charge, a set of 2D custom cards (Figure 114) that include a specific pattern. When this pattern is recognized by the device's camera, it will provide a 3D image of the molecular model instantaneously. The rotation of the molecular model is possible at any angle, as long as the camera can detect the code. This allows the users to overcome one of the major difficulties in seeing and analyzing molecular models in computers: the users can interact with it as a real object and not in a less intuitive way that is commonly used through the use of a mouse and a computer screen. The BioSIM^{AR} is available free of charge at ar.biosim.pt.

Figure 114. Example of a printed card showing the pattern code to observe the 3D structure of an RNA molecule using the BioSIM^{AR}.

5.1.2. Development

The development of BioSIM^{AR} took in mind its application in the classroom of science-related subjects. Therefore, BioSIM^{AR} was coded to run directly on every device with a camera, an internet connection, and a web browser. The restrictions regarding the type of device or operating system are waived by the approach assumed in the development of this software.

Regarding the internet connection requirement, BioSIM^{AR} was developed to avoid possible bad performance in poor-quality connections. When the web page is accessed, all the required data is downloaded (about 50 megabytes) and stored in the browser's cache. Then, the camera becomes active, and the users can interact with the cards. From this moment, all the visualization and interaction occur locally in the user's device. This strategy ensures the best possible performance independently of internet connection quality. Therefore, the internet connection only affects the loading time with no impact on subsequent usage. The structure files are also kept in the browser's cache for a month, avoiding repetitive downloads for each time that page is reloaded during that time range.

From our teaching experience in classes and workshops, the ease of use is essential to make a particular technology be useful and applied in the classroom context. If a specific educational asset is too difficult or time-demanding to be prepared, teachers usually avoid their use due to the scarcity of time [504, 505]. Consequently, the compatibility with most of the devices is also an important aspect to avoid issues and delays on using

BioSIM^{AR}. Therefore, BioSIM^{AR} was tested in the top-6 most used web browsers (Google Chrome, Safari, Mozilla Firefox, Samsung Internet, Edge, and Opera), which represent more than 95% of the total market share [506]. The tests were conducted using Android, iOS, Microsoft Windows, and macOS operating systems. According to those tests, Samsung Browser for Android showed poor performance when used in portrait mode, since the camera image did not fit properly on the screen. Table 2 summarizes all the compatible tests conducted, showing that BioSIM^{AR} is fully compatible with all main operating systems and web browsers.

As shown, BioSIM^{AR} is compatible with a panoply of different combinations of operating systems and web browsers. This compatibility with so many different platforms reduces the possibility of issues during its usage.

Table 2. Compatibility Analysis for BioSIM^{AR} running in the different web browsers and operating systems.

Operating System (OS)	OS Version	Web Browser	Web Browser Version	Compatibility
Android	8.0.0	Edge	45.11.2.5118	Full
		Google Chrome	87.0.4280.101	Full
		Mozilla Firefox	82.1.3	Full
		Opera	61.1.3076.56625	Full
		Samsung Browser	13.0.2.9	Bad performance (portrait)
iOS/iPadOS	14.3	Edge	45.11.11	Full
		Google Chrome	87.0.4280.77	Full
		Mozilla Firefox	29.2	Full
		Opera	2.5.3	Full
		Safari	14	Full
Microsoft Windows	10	Edge	87.0664.60	Full
		Google Chrome	87.04280.88	Full
		Mozilla Firefox	84.0	Full
		Opera	73.0.3850.284	Full
macOS	11.1	Edge	87.0.664.60	Full
		Google Chrome	87.0.4280.88	Full
		Mozilla Firefox	84.0	Full
		Opera	73.0.3856.284	Full
		Safari	14.0.2	Full

BioSIM^{AR} was coded based on two already available open-source projects: A-FRAME (https://aframe.io/) and AR.js (https://github.com/AR-js-org/AR.js). Based on those, BioSIM^{AR} was prepared to show 3D representations of molecules in an augmented reality interface (Figure 115). The interactivity was enhanced by using cards with the

molecule's name and a bidimensional code (QR code) (Figure 114). The QR code on the card is recognized by the device's camera and integrated by the software to present the corresponding molecule. In addition, the pattern code provides spatial information about where the molecule must be placed in order to provide a real augmented reality experience. This particular feature allows users to rotate the card and/or the device's camera to inspect the shown molecule from different angles.

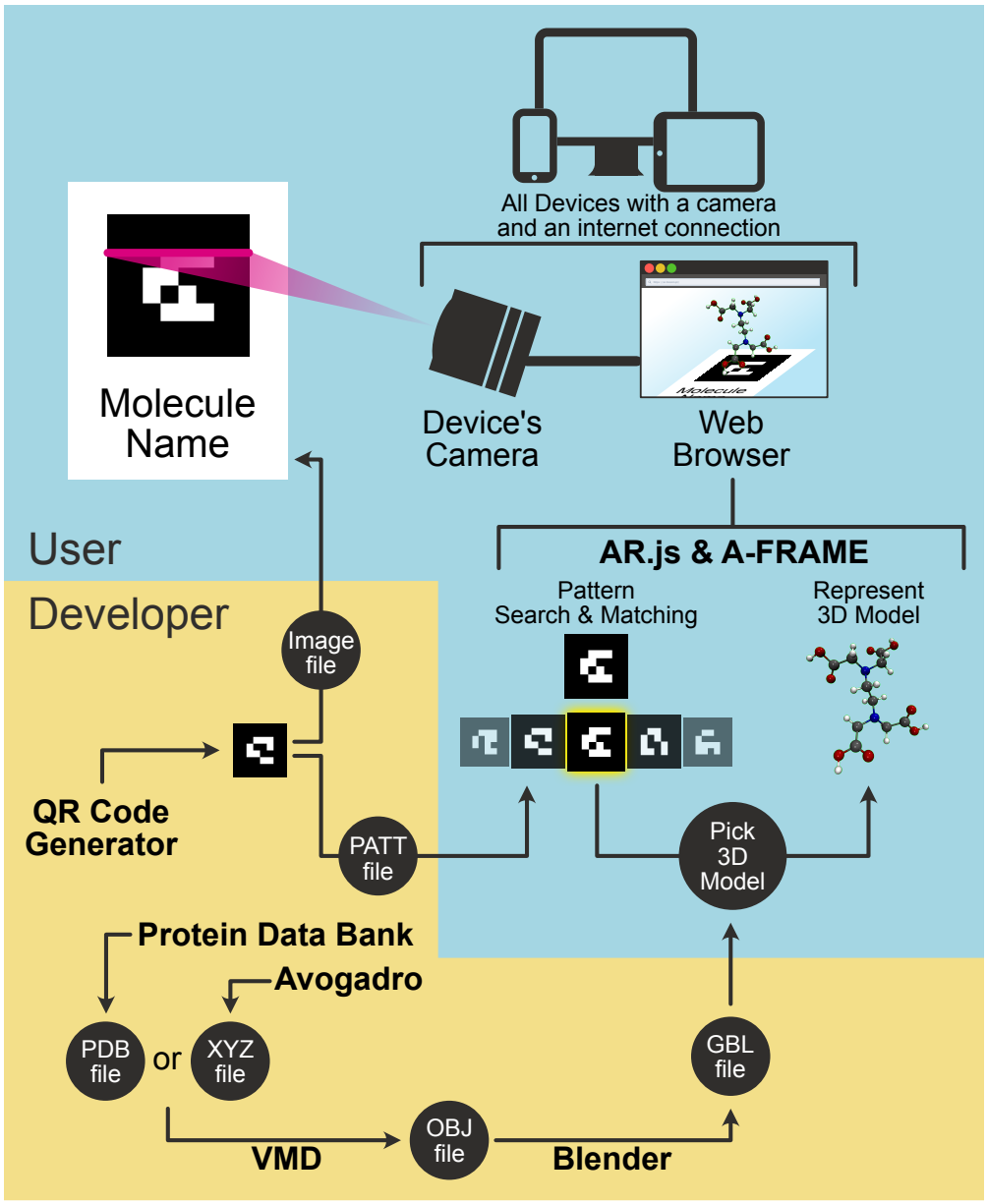


Figure 115. Overview of the workflow involved in the (yellow) development of BioSIM^{AR}, and the processes taken when users point their devices' cameras to a 2D pattern (blue).

The molecules used in this project were obtained from the PDB repository [507] or manually drawn and optimized using Avogadro 1.2.0 [462, 484] (Figure 115). The resulting structures, saved as a PDB or XYZ file, were loaded on VMD [292] in order to prepare their molecular representations. Then, the 3D model of each molecule was

exported as a Wavefront OBJ file using the render option of VMD. Subsequently, each Wavefront OBJ file was imported on the 3D model software Blender 2.83.1 [508]. Using Blender, the files were adjusted, and, in some cases, additional tags were included, such as information about bond length, angle amplitudes, dihedral torsions, etc. Finally, a GBL file of the 3D model was exported using Blender. The GBL file contains all the data required by BioSIM^{AR} to show each of the 3D molecular representations.

For each molecule that is generated for BioSIM^{AR}, it is also required to design a 2D pattern that is used to identify the molecule. The QR codes were chosen because they provide a unique form to identify different types of data, and they can be easily generated online with tools like “QR Code Generator” (https://www.the-qrcode-generator.com/). Hence, a QR code was generated for each molecule based on its name or internal identification code. Afterward, all the QR codes were converted in a PATT file according to the AR.js requirements. In the end, three different files are available for each molecule: a GBL file containing the 3D structure of the molecule, a picture of the associated QR code, and a PATT file with the pattern associated with the molecule. In this way, when the device’s camera identifies a pattern code, the code is compared with all the PATT patterns available on BioSIM^{AR}, and if it matches, the corresponding molecular representation is loaded from the GBL file and showed (**Figure 115**).

BioSIM^{AR} can be easily assessed from the following hyperlink (biosim.pt/AR/) that opens a webpage and requires access to the device’s camera. At this point, users are capable of pointing the camera to one or more cards containing the pattern code. Automatically, the software will show the 3D representation of the molecule, and the users can interact with it. Regarding privacy, BioSIM does not receive or handle the images captured with the camera. The models and patterns are delivered to the browser, and all the processing and analyses occur on the device with no communication with the server. No data is collected apart from the page views statistics collected by Google Analytics.

At this moment, 70 different cards are available for free download at https://ar.biosim.pt/. All the cards are available in English and Portuguese, but users can easily adapt them to other languages since the recognition patterns it what is required to identify them. Those cards include simple molecules as water, ammonia, or ethane, and more complex ones like RNA and DNA fragments. These cards have been developed progressively based on the requests of Portuguese and Brazilian high school teachers. BioSIM^{AR} is prepared to scale up with more structures to shelter different chemical aspects. In addition, due to the open-source nature of this project and of the software involved, users can develop their own sets of molecules and make them available to the community. Currently, there is no automatic process for users to add more molecules to BioSIM^{AR}.

The process involves a contact (e-mail), and then the suggestions are analyzed, prepared, and added by us. In the long-term, BioSIM^{AR} could be improved to support the creation of custom subsets of molecules, where users upload their own molecules.

5.1.3. Implementation

This project has been used in computational chemistry-related workshops and classes, as well as by high school teachers in science-related courses. In particular, the BioSIM^{AR} was used in a chemistry class about the geometry of molecules. A total of 26 high school students (10th grade of the Portuguese School System) attended the class and were asked to answer a survey about their opinion on BioSIM^{AR} anonymously. The students were invited to classify from 1 to 5 (strongly disagree to strongly agree), seven statements about BioSIM^{AR} (Figure 116). All 26 students answered all mandatory questions of the survey. The optional comment/suggestion field was answered by ten students.

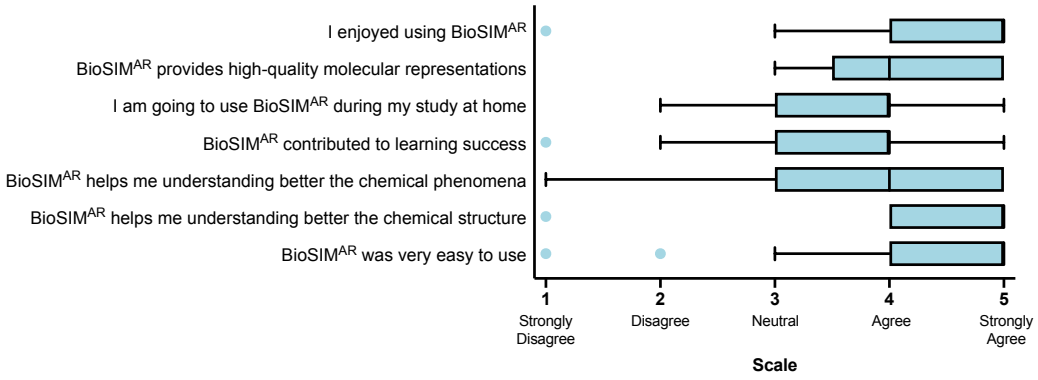


Figure 116. Tukey test results for students regarding the use of BioSIM^{AR} (n=26). Scores on the scale have a range of 1-5. The boxes correspond to the first quartile (left), median (center), and third quartile (right). The whiskers correspond to the lowest (left) and maximum (right) values observed. The blue dots represent outliers.

The first question ("I enjoyed using BioSIM^{AR}") intends to get an overall opinion of the students regarding the experience of using this technology. Although no information about previous experience with other molecular representation technologies was collected, the students considered BioSIM^{AR} easy to use. They also agreed that BioSIM^{AR} provides high-quality molecular representations. When asked about the intention of using BioSIM^{AR} during their study at home, most of them manifested a strong interest. Although most of the students thought that BioSIM^{AR} helps in their learning process, BioSIM^{AR} seems to be more useful to assist students in understanding chemical structure rather than chemical phenomena. This was expected since the molecules available on BioSIM^{AR} are static and were prepared to show structural differences between molecules. In the future, we intend to include dynamics simulations to show chemical reactions and other dynamics phenomena. The helpfulness of BioSIM^{AR} in terms of learning success was only assessed through the opinion of the involved

students. In a future workshop or class, it would be valuable to measure the actual impact of BioSIM^{AR} in terms of learning about chemical structure. In this case, students need to be split into groups, and only one group would use BioSIM^{AR} to answer a set of questions about chemical structure. The other group will only have access to the textbook.

In addition to these seven statements to classify, students had the opportunity to share custom feedback or suggestions. One student strongly disagreed with most of the statements because he/she was unable to use BioSIM^{AR} on his/her device ("I was unable to load the site, so I cannot evaluate its quality"). Although a bug report section is available on the BioSIM^{AR} website, this issue was not reported. Therefore, we do not have further information about what causes such incompatibility. Three other students suggest that the quality of the molecules could be improved, and another student suggests the inclusion of a button to show/hide further information about the shown molecule. Improving the quality of the models makes the related files considerably bigger, which deeply affects the loading time. However, we are working on strategies to dynamically adjust the quality of the models based on the internet connection speed. The suggestion about adding buttons to the interface was already on our roadmap for future development.

This practical case supports BioSIM^{AR} as a potentially valuable asset in understanding chemical structure using a cheaper and accessible technology. Moreover, the results of this survey agree with the comments we have received from other teachers and students. However, further real case scenarios need to conduct to assess the real impact of BioSIM^{AR} on the learning process. Different students with and without access to BioSIM^{AR} could be asked to answer the same test about chemical structure. Alternatively, the same group of students could be asked to solve a test before and after using BioSIM^{AR}.

The popularity of BioSIM^{AR} is corroborated by the number of users that access the online platform. Since the release of BioSIM^{AR} in November 2018, our platform was accessed more than 12,000 times, according to the Google Analytics Report.

5.1.4. Conclusion

BioSIM^{AR} offers a free and interactive alternative to explain a panoply of different chemical concepts. For example, it can be useful to provide a spatial comparison between L and D isomers of alanine (Figure 117). BioSIM^{AR} delivers an interactive tool to represent a difference between the two possible conformers of a chiral carbon.

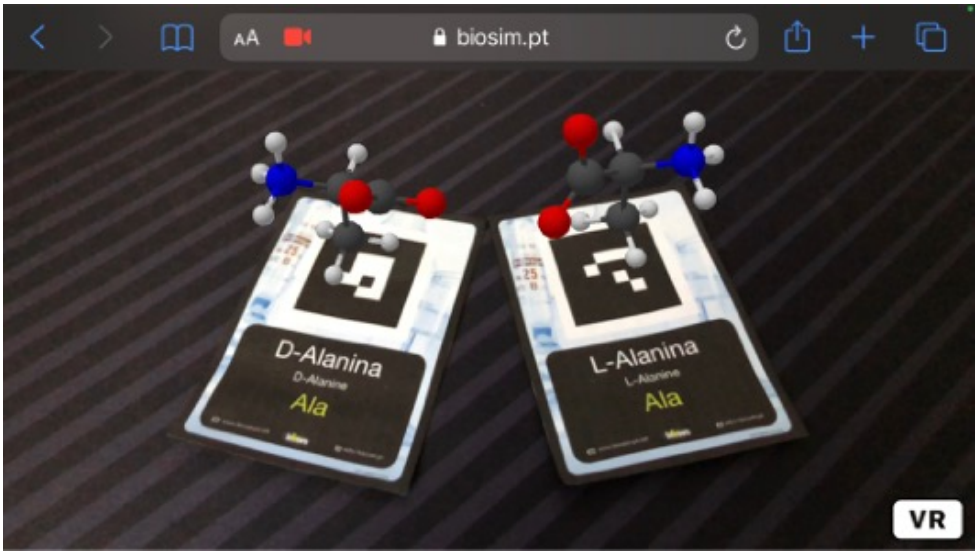


Figure 117. Comparing the structure of (left) D-Alanine and (right) L-Alanine, highlighting the chiral α-carbon.

In sum, BioSIM^{AR} ensures that more students, teachers, and schools have access to high-end technology without the requirement of expensive hardware or time-consuming and complex setups. Since students can easily use BioSIM^{AR} at home, it also offers an excellent complement to other 3D representation technologies only available at school. Therefore, they can continue to interact with 3D molecular representations during their study sessions at home.

This page was intentionally left blank.

5.2. Protein Wars

Protein Wars is an educative videogame developed by Henrique S. Fernandes and Nuno M.F.S.A. Cerqueira. Protein Wars was developed as a VMD extension, and it aims to elucidate the general public, particularly children, about how drugs target some specific proteins in the human body.

This project evolved to build a retro arcade game machine where Protein Wars runs on top of VMD, and the “gamers” can use the buttons and the joystick to shoot a laser targeting the drug-binding region of a target protein (Figure 118).

Figure 118. Photography of the Protein Wars arcade machine at “Mostra da Universidade do Porto”.

This page was intentionally left blank.

5.2.1. Introduction

The advances in science and technology play a very important role in the development of human society. Almost all aspects of our daily life were hugely improved due to the continued research carried on by several scientists and institutions around the world and from different fields. Although scientific knowledge has become more specific and complex throughout the years, it is important to keep in mind that an educated society is a better society. Therefore, it is essential to keep creating events and technologies that can bring complex science to society in a clear and informative manner.

In this project, Protein Wars was developed to create a catching device to captivate the general public attention, particularly the children. Protein Wars was developed as an educative game where players are challenged to identify and target the drug binding-site of different proteins associated with different diseases. The strategy was calling the attention of the players to the importance of developing highly specific drugs able to target key proteins. While gaming, the players acquire some important scientific knowledge, such as the association of certain proteins to certain diseases, how the drugs can inhibit the enzymatic activity, and the importance of highly specific drugs to avoid harsh side-effects.

5.2.2. Development

a. Software development

Protein Wars is a videogame developed as a VMD extension using the Tcl/Tk programming language. The game runs on top of a widely used scientific software, the VMD, benefiting from its accuracy and powerful features for representing molecular systems.

The game was developed following a level scheme, where players are confronted with a different level after successfully completing the previous one. In each level, players are virtually placed in a nanorobot that is traveling inside the human body and intends to solve problems associated with proteins. These problems can be related to the malfunction of proteins. Therefore, the gamers are asked for destroying enzymes involved in the pathology of cancer, AIDS, and other diseases. The game's main goal is to destroy these targets. However, this is not an easy task, and the player must discover which parts of each protein can lead to the complete destruction of the protein. The game has seven different levels with increasing levels of difficulty, and the player has only 20 bullets and limited fuel to finish all the levels successfully (Figure 119). The limited number of bullets intends to elucidate the importance of rationally design drugs specific to the target region of a protein.

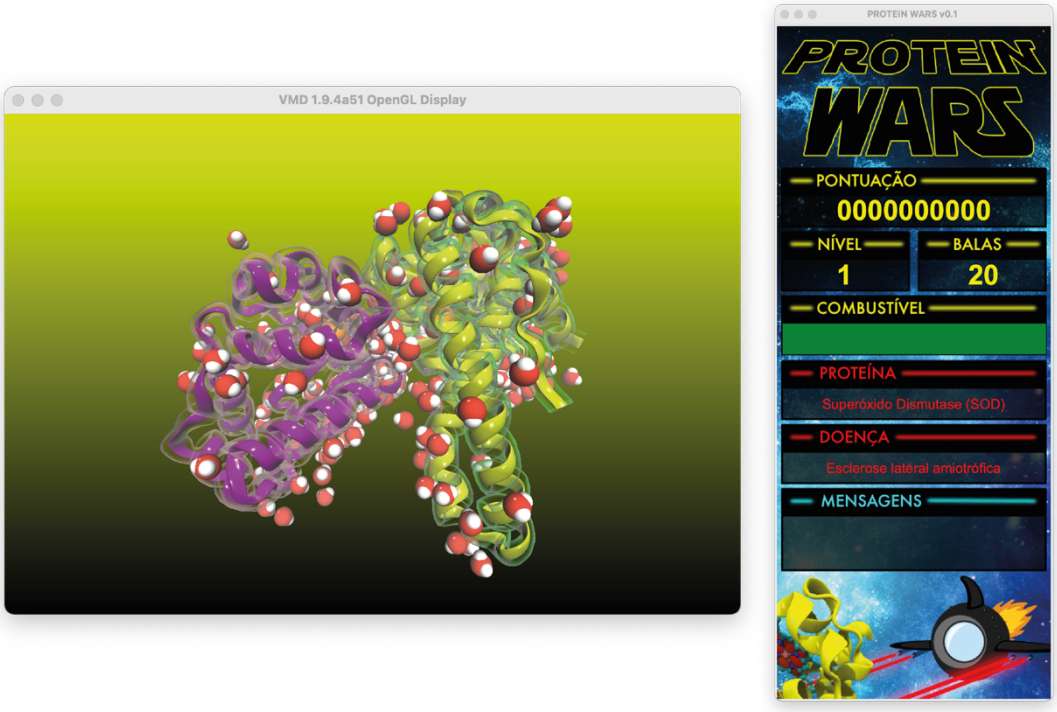


Figure 119. Screenshot of Protein Wars at level 1.

Before each level starts, an informative message is displayed (Figure 120), identifying the molecular structure that will be targeted, one associated disease, and a brief description of how that molecular species is involved in that particular disease. To instigate players to read the actual text, some tips are available in that message, helping players complete the level.

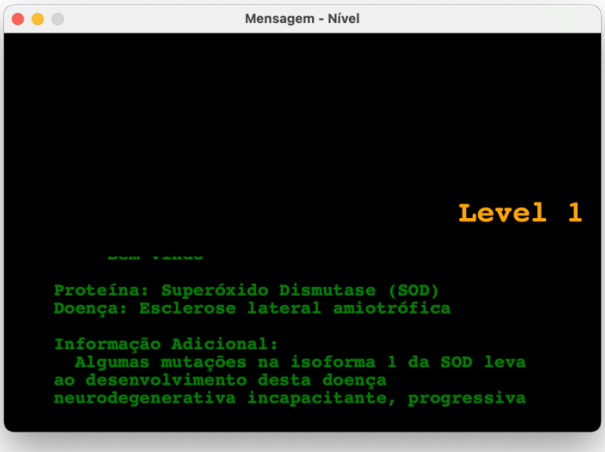


Figure 120. Screenshot of part of the initial message displayed for level 1 of Protein Wars.

The development of this game requires some external software to overcome some limitations of using VMD. In particular, the game includes music and sound effects that are not natively supported by VMD. Therefore, the *SoX* external and free software (http://sox.sourceforge.net) is used to play those sounds directly from the terminal. Moreover, additional software was required to emulate the left mouse button click when a specific key is pressed. This was necessary to perform the shoot when a key is pressed, triggering the mouse click at its current position. In this case, the *Cliclick* software (https://www.bluem.net/en/projects/cliclick/), developed by Carsten Blüm, was used for emulating mouse clicks using keyboard events. For Linux operating systems, the *xdotool* software (https://github.com/jordansissel/xdotool) was used instead. The x*dotool* was developed by Jordan Sissel, and it is freely available.

Because of these specific requirements, the Protein Wars is only supported on macOS and Linux-based machines.

b. Building the Retro Arcade Game Machine

This project went beyond the development of the software, and an arcade-style game machine was also built. The process involved the cut and assembling of multiple parts of medium-density fiberboard (MDF), painting, and the entire setup of the electric button and joystick (Figure 121).

All game controllers were wired to a controller card that was connected to a computer where the Protein Wars was running.

Figure 121. Photos of the process of building the arcade-style game machine for Protein Wars.

5.2.3. Conclusion

Protein Wars is an educative game developed as an extension for VMD. The main goal of such a game is to attract the general public, in particular children, to transfer scientific knowledge in a clear and funny manner. Therefore, Protein Wars has been used in several science fairs and social events of science divulgation. Since its development, Protein Wars was already presented in workshops, science fairs at the Faculty of Sciences of the University of Porto and “Mostra da Universidade do Porto”, and in the event of science divulgation during a stage of the “Volta a Portugal em Bicicleta”. In that particular session, Protein Wars was also presented in a report of the Portuguese national channel RTP 1 (report can be watched at https://bit.ly/393yV8Q).

Although high-quality scientific research is the main driven force of all scientists, they also need to assume this important role in bringing science closer to general society. Protein Wars is just another example of how science could be presented in a funny and educative approach.

Protein Wars is available for free download at https://github.com/henriquefer/ProteinWars.

Download
Protein Wars

This page was intentionally left blank.

5.3. Criminal Investigation

The Criminal Investigation is an educational experiment developed by Henrique S. Fernandes, Carla S.S. Teixeira, and Nuno M.F.S.A. Cerqueira. In the sequence of the development of Protein Wars, Criminal Investigation was designed to approximate the scientific research to the general society, mainly children.

This experiment was developed, taking into consideration a web-based platform where users read about the problem that they are facing and wet-lab experiments that they can also perform at home.

Criminal Investigation was presented at the “I DQBgotTalent” contest organized by the Chemistry and Biochemistry department of the Faculty of Sciences of the University of Porto in 2017. The contest asked for ideas of experiments that could be performed with young students to teach some scientific knowledge in an interactive manner. Criminal Investigation was awarded the first-place *ex aequo* with another participation.

This page was intentionally left blank.

5.3.1. Development

As mentioned, Criminal Investigation was developed considering a software interface and actual wet-lab experiments that are conducted in parallel.

The software part was developed as a web platform based on HTML, CSS, and Javascript languages. The embedded animation was also designed and coded using HTML5 and Javascript. The map available on the web page 3 (Figure 122 3) was created using the Google Maps API.

The web platform provides information about the crime scenario and animated tutorials about how forensic tests should be performed to identify the culprit. Here, users assume the role of a criminal detective and try to solve the crime, identifying the felon among a list of three suspects (Figure 122).

Figure 122. Screenshots of the first three web pages that are shown on Criminal Investigation. The third page provides details about the actual crime, including an interactive map with key locations with valuable information.

After the “detectives” have been read all the preliminary information about the “crime scene”, they have access to the crime evidence identified with bar codes.

In the first experiment, the “detectives” have three pieces of fabric from suspects’ clothes and one-piece found on the “crime scene”. The labeled pieces of evidence have codes that can be matched with the actual suspects from the web platform’s information (Figure 123). This first task of assigning the correct piece of evidence to the respective suspect

is relevant to show the importance of carefully analyzing the available data and correct labeling of all the samples and material used during a lab experiment.

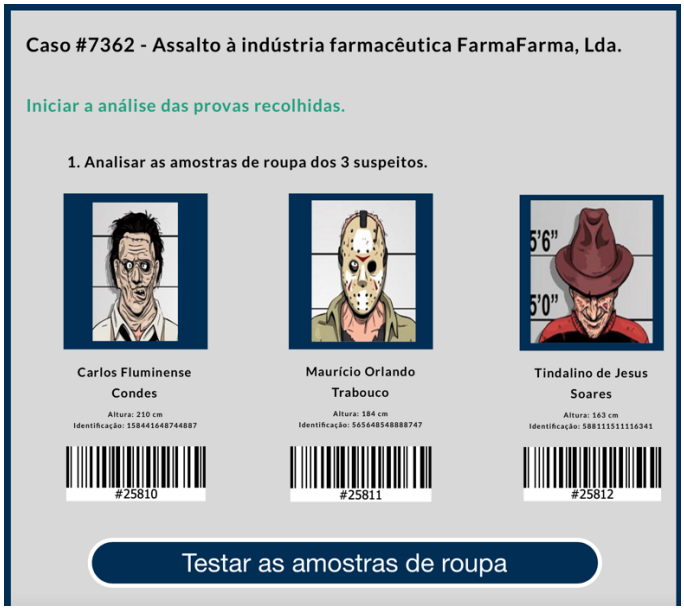


Figure 123. Screenshot of the web platform showing the evidence bar code associated with each evidence.

Then, the “detectives” are invited to follow a wet-lab experiment where the fabric pieces are added to an alcoholic solution. In this way, the color pigments are transferred to that solution. For the purpose of this experiment, the pieces of fabric are, in fact, small portions of ink marker loads. Afterward, they should perform a paper chromatography according to the instructions displayed in the animations available on the web platform (Figure 124). In this stage, the “detectives” compare the obtained results with the chromatographic profile of the evidence found in the “crime scene”. This first test allows “detectives” to exclude one of the suspects.

Figure 124. Screenshots of all the pages showed on the web platform with instructions about how to perform a chromatography test to acquit one of the suspects.

Finally, the “detectives” are instructed to perform a second experiment where the acidity of the soil founded on the shoes of each of the two remaining suspects are analyzed together with the sample of soil collected from the “crime scene” (Figure 125). Firstly, they need to create a “homemade” pH indicator based on an aqueous solution from boiling red cabbage. Then, the soil samples are mixed with the pH indicator in order to compare the samples from suspects’ shoes with the solid from the crime scene. At the end of this experiment, the “detectives” might be able to identify the actual culprit of this crime.

Figure 125. Screenshots of all the pages showed on the web platform with instructions about how to test the pH of different soil samples to find the felon of this “crime”.

a. Requirements

The web-based platform behind this experiment can be freely downloaded at https://github.com/henriquefer/CriminalInvestigation. Once the download is completed and the package is unzipped, the “index.html” file can be opened on a web browser.

As aforementioned, this educative experiment is composed of two sequential wet-lab tests: a paper chromatography and a pH indicator preparation and test.

The paper chromatography requires the following material:

- 6 cups
- 3 zip-bags
- 1 marker

- 2 ink marks loads of different colors
- 1 scissors
- 100 mL toilet water (or other alcoholic solution)
- 1 pipette
- 1 glass rod
- 1 hairdryer (optional)

In the case of the pH indicator experiment, the following material is required:

- 3 zip-bags
- 1 marker
- 100 g sand or soil
- 10 mL of vinegar (to acidify two of the soil samples)
- 10 mL of glass cleaner solution (to alkalize one of the soil samples)
- 1 red cabbage
- 200 mL water
- 1 kettle
- 1 knife
- 2 cups
- 1 strainer
- 1 pipette
- 3 test tubes
- 1 test tube support
- 1 spatula
- 1 glass rod

5.3.2. Conclusion

Criminal Investigation is a simple set of experiments developed to engage children and teenagers in science-related topics. This experiment involves a computational component to provide instructions and information that the target public needs to relate to and understand. Moreover, those instructions and information are essential for successfully carry on the forensic wet-lab experiments that will disclose which suspect was responsible for the simulated crime.

This experiment was designed to improve the involved people's understanding of technical information, organization during a lab experiment, critical thinking, and dexterity in carrying out the experiments. All of these aspects are allied to a captivating story able to captivate the target public's interest.

Unfortunately, this educational tool was not used in a real scenario yet. However, everything is ready to be implemented in a future suitable event, such as science fairs or other scientific divulgation events for general society.

Download
Criminal Investigation

6. Databases

The internet was an invention that becomes very popular at the very end of the last millennium. Since then, several services and fields were deeply improved by the using internet, particularly the access and share of information and knowledge. The journal publications are the main source of scientific divulgation, and they truly benefit from the popularization of the internet. Nowadays, all scientific journals provide access to their content through the internet. The access can be totally free for all internauts or based on paid subscription depending on the journal and publication type.

All these advances in sharing knowledge worldwide help researchers to reach information faster and easier. Consequently, the amount of knowledge available is now significantly higher than before. Unavoidably, the analysis of such an amount of data becomes more challenging, and more advanced strategies need to be employed to retrieve valuable information from all available raw data. The strategies for a quality review of the available data require grabbing the results, organizing them, excluding possible poor-quality data, cross validating the results from different independent sources, and using different data representation techniques to disclose relevant conclusions.

The review article is a type of scientific publication where experts take the current state-of-art about a certain topic and proceed to a careful analysis following the steps mentioned above. In some situations, the results are compared using extensive tables that summarize and compare different features of the available results. Even though this

strategy works quite well when a few results are available, static tables can become less useful when there is much information. In those cases, dynamic and interactive tables can be the better option to establish relationships between the features of different entries of the table. Here, we are talking about simple spreadsheets or even web-based databases that can be easily accessed and searched.

In this sense, a bibliographic review can originate a database where the available results are related to each other in a user-friendly and powerful tool. Depending on the extension of the database development, different levels of analysis tools can be made available to provide a different option for correlate data.

Additionally, if databases become incrementally updated to include the recently published results, they can also work as an important source of new data for researchers interested in that specific field. Nonetheless, the new data will also be automatically integrated and compared with all the previously published results.

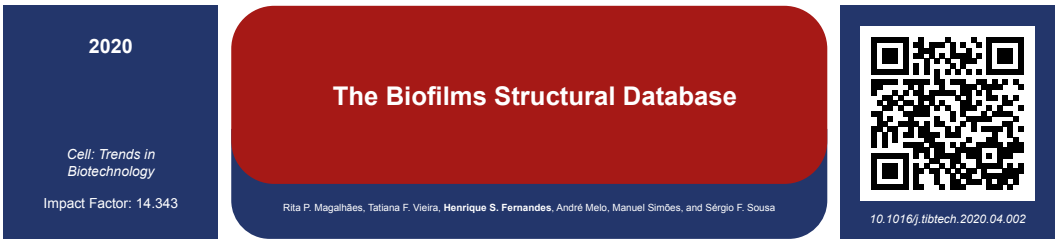
In sum, web-based databases can help researchers to organize the huge amount of data that is already published and correlate them in a smart and efficient way.

6.1. The Biofilms Structural Database

The Biofilms Structural Database (BSD) [509] collects structural, kinetics, mutagenesis, and inhibition data related to the process of biofilm formation. More than 400 protein structures from more than 40 different bacteria were manually curated and included in a free web repository available at biofilms.biosim.pt.

Rita P. Magalhães and Tatiana F. Vieira were responsible for all the curation process and organization of the database's data. Henrique S. Fernandes coded all the web design, database programming, and automatization.

The work was published in the Trends in Biotechnology journal, and the following content is almost a transcription of the published version.



The Biofilms Structural Database (BSD) is a collection of structural, mutagenesis, kinetics, and inhibition data to understand the processes involved in biofilm formation. Presently, it includes curated information on 425 structures of proteins and enzymes involved in biofilm formation and development for 42 different bacteria. It is available at biofilms.biosim.pt.

This work was also presented in the following conferences:

- (Poster) *6º Encontro De Jovens Investigadores Em Biologia Computacional Estrutural*, 2018, Porto, Portugal (Appendix: page 379)
- (Poster) *7º Encontro De Jovens Investigadores Em Biologia Computacional Estrutural*, 2019, Lisbon, Portugal (Appendix: page 385)

This page was intentionally left blank.

6.1.1. Introduction

Biofilms are heterogeneous and complex structures of microorganisms, typically adhered to a surface and presenting sophisticated singular and collective behaviors. Bacteria in biofilms are embedded in a self-produced matrix of extracellular polymeric substances (EPS) composed of (glyco)proteins, (glyco)lipids, mono or polysaccharides, extracellular DNA, minerals, and water [510]. This protected mode of growth allows bacteria to survive in hostile environments. In the health context, some diseases and adverse medical conditions are now recognized to be the result of a biofilm infection, and it has been estimated that more than 65% of all microbial infections in humans are caused by biofilms [511]. Accordingly, there has been a great amount of research to understand better biofilm development and to identify improved control strategies [512]. However, the structure, composition, and physiology of microbial biofilms have become inexorably linked with our failure to control them by antimicrobial treatments that are effective against planktonic bacteria. In fact, once set, biofilm-associated infections are very difficult to treat because the colonizing bacteria are ten to thousand times less susceptible to antimicrobials than their planktonic counterparts. Consequently, the therapeutic options for the treatment of many infections are compromised [513].

Antibacterial strategies focusing on inhibiting cellular growth often induce antimicrobial resistance by imposing a strong selective pressure on bacteria. By contrast, alternatives that target the processes associated with biofilm formation and development, which include bacterial motility, cell adhesion, biofilm dispersion, the synthesis of EPS, and also cell-to-cell communication or quorum sensing (QS), can be more efficient [514, 515], as these processes are less amenable to induce antimicrobial resistance. Throughout the years, many signals, regulatory pathways, and autoinducer molecules have been identified, but their intricate network remains unclear at the atomic and molecular level [516]. Understanding the mechanisms involved in these processes is essential for the development of new therapies against biofilms.

Over the past few years, there has been an immense increase in the number of 3D structures (particularly X-ray) of proteins and enzymes associated with biofilm formation and development. The data associated with these structures, the kinetic and mutagenesis data, amino acid sequences, and inhibitory activity of known molecules offer new opportunities to understand biofilms at a structural and atomic level. This recent development has allowed researchers to move from the cellular to the molecular level, inducing a whole new dimension in the development of anti-biofilm therapies. In particular, the new structures available offer an appealing alternative for the rational design of new drugs and for the application of techniques, such as molecular docking,

virtual screening, quantitative structure-activity relationship (QSAR) models, and molecular dynamics simulations.

The BSD [509] was designed to organize all the available structural information on these promising targets, presenting up-to-date atomistic information on the proteins and enzymes involved in biofilm formation and development. It is the first database dedicated to the biochemistry of biofilms from a structural and molecular perspective of the proteins associated. It is a platform that can be used not only by those who wish to study the proteins responsible for the mechanisms of biofilm formation but also by structural and computational biologists who want to study the molecular details of substrate recognition. It helps to elucidate the mechanisms of action of several key proteins and enzymes and opens the door to the rational development of new compounds with anti-biofilm activity.

6.1.2. Data Selection

The main focus of the BSD is at the molecular level. However, biofilm research integrates several different aspects. Therefore, the database was designed to create a comprehensive repository of structural data on biofilm research that could integrate with the information available in other databases, beyond the molecular level, to link different fields. An intensive literature search was performed to extract the relevant structural information about the proteins and enzymes involved in biofilm formation. The Institute for Scientific Information (ISI) Web of Science, Scopus, and PubMed databases were searched using combinations of biofilm-related keywords, such as “biofilm”, “quorum sensing”, “motility”, “dispersion”, “matrix protein”, and other relevant terms, with the keywords “structure”, “X-ray”, “NMR”, and “cryo-EM” (cryogenic electron microscopy). In addition, queries on the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) [507] were also performed. Additional queries on specific proteins known to be involved in biofilm formation and development were also conducted. The overall information was manually checked and cross-validated with the literature available for each protein, and the structures were visualized inspected.

Other databases, such as ChEMBL [517], BindingDB [518], ExPASy [519], KEGG [520], and UniProt [521], were cross-referenced to provide additional information for each protein. The BSD contains direct links for other drug-like molecules that have already been tested against that biofilm protein, information on kinetic and binding affinity data on mutant variants, and genomic and proteomic data regarding each specific protein. As investigations into biofilm progress and new structures are made available on the PDB database, the BSD will continue to be expanded, updated, and maintained to ensure that it remains a useful and scientifically accurate tool to study biofilms from a molecular

perspective. The database will be maintained and regularly updated by the BioSIM research group at UCIBIO/REQUIMTE, Faculty of Medicine, University of Porto, and any researcher can contribute with new entries by filling the available form at biofilms.biosim.pt.

6.1.3. Organization of the Database

The BSD is organized in a simple and intuitive manner. Each entry in the database corresponds to one atomic structure with a unique PDB code. A total of 425 PDB entries are currently available, corresponding to a total of 133 unique proteins (Figure 126 A).

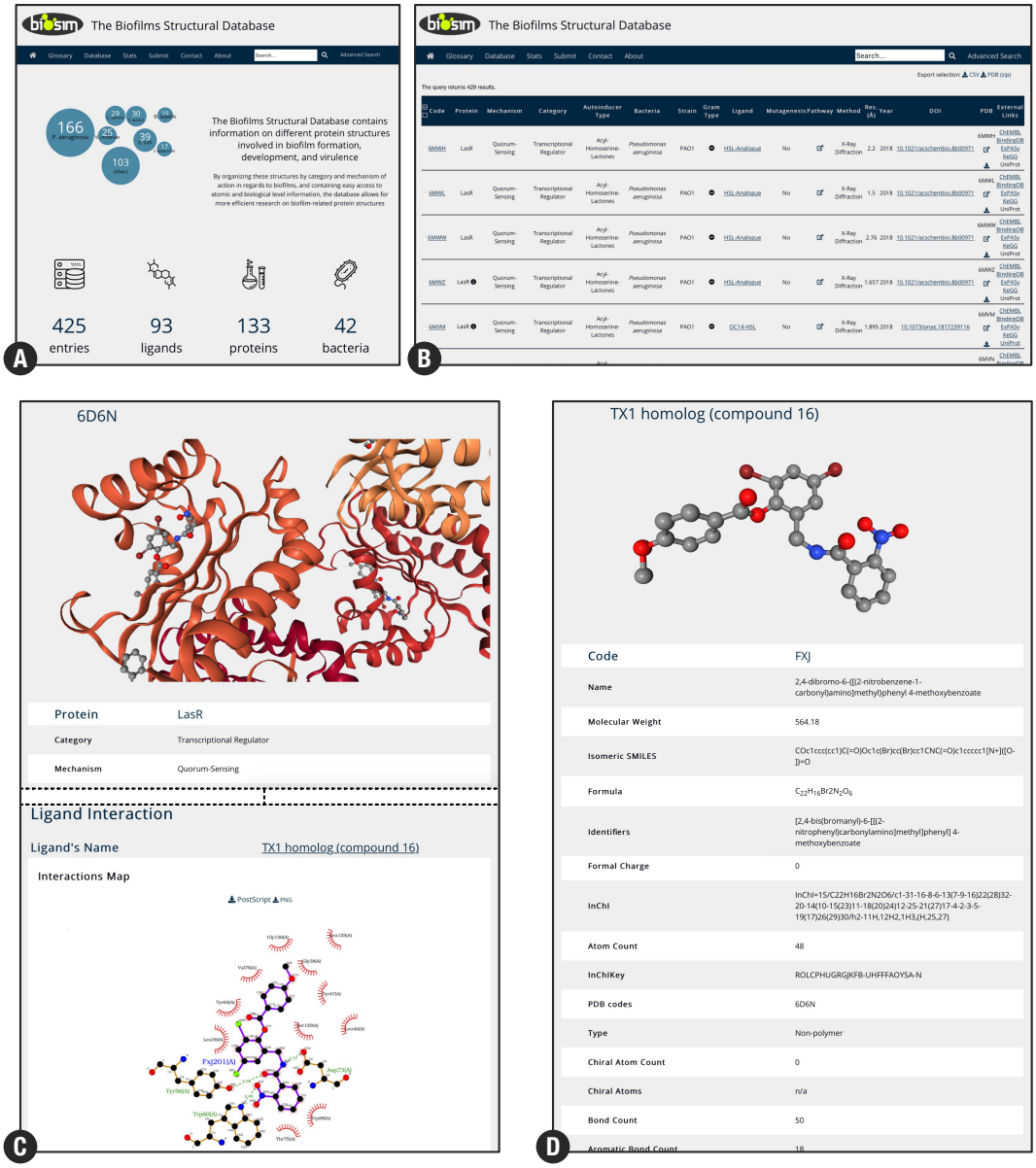


Figure 126. The Biofilms Structural Database (BSD). (A) BSD main page. (B) List of database entries in BSD. (C) Entries by structure for each Protein Data Bank (PDB) code with an indication of the main information available for each protein/structure and interactive 3D representation of the protein. (D) Entries by ligand with main information and interactive 3D representation.

Each entry is identified through the PDB code associated with the corresponding structure in the PDB and by the name of the protein. Entries are classified by protein category and mechanism. The term “category” refers to the general and main function of the protein. The term “mechanism” refers to the mechanism by which the protein is associated with biofilm formation and development. These include QS, quorum quenching, motility, EPS production, dispersion, and signaling. Database entries are also listed by bacteria, specific strain, and Gram type. Finally, all entries are also classified by the existence of a ligand in the structure and by its designation.

Mutagenesis data is important to identify amino acid residues with a special role in protein function, including ligand binding, enzymatic activity, among others. For each protein, the availability of mutagenesis information is indicated. Links to the mutagenesis section in BRENDA (Braunschweig Enzyme Database) [13] were included, providing information on the effect on kinetics (k_{cat}) and/or k_M data of the corresponding mutations.

Information about the biological and metabolic pathways in which each protein is involved is also presented (KEGG). This allows the user to quickly visualize the whole mechanism of action and consider including other intervening proteins in the study. The BSD also lists information about the method by which the atomic structure of the protein was obtained (X-ray crystallography, solution NMR, or cryo-EM), resolution, year of deposition on the PDB, and DOI of the associated publication.

Each entry contains direct links for structure retrieval from the PDB and for direct access to the entries associated with each protein in the ChEMBL, BindingDB, ExPASy, KEGG, and UniProt databases, when available (**Figure 126** B). The ChEMBL and BindingDB databases contain information on collections of molecules previously shown to exhibit experimentally confirmed activity against the corresponding protein. These links enable the direct download of all known inhibitors or activators for a specific protein, with the corresponding activity and properties for drug development efforts.

Each BSD entry contains a link to its dedicated page containing a summary of all the information indicated previously, including an interactive 3D representation of its structure (**Figure 126** C). A map of interactions of the ligand, obtained with LigPlot+ [522], is also presented. This is complemented with an interactive 3D representation of the ligand and surrounding amino acid residues and main interactions, which allows users to quickly visualize the 3D structure of the protein-ligand complex at the binding pocket without having to use molecular visualization software.

Entries are also organized by ligand (Figure 126 D). A total of 93 unique ligands are presently available, including substrates, autoinducers, and inhibitors. For each ligand, the corresponding PDB code, systematic name, molecular weight, isomeric SMILES (simplified molecular-input line-entry system), molecular formula, and several other molecular properties are included. An interactive image of the ligand's structure is also displayed.

The database enables queries through a variety of options, including queries by protein, ligand, category, mechanism, autoinducer type, bacteria, strain, and Gram type. Selections can be easily downloaded as a CSV file or as a set of PDB structures.

6.1.4. Conclusion

The BSD is a free-access catalog containing the structures of all known proteins and enzymes involved in biofilm formation. This database is a tool that helps to visualize, explore, and understand biofilm targets to design and develop new and effective antibiofilm drugs and to understand the structure and activity of proteins and enzymes involved in biofilm development. The interface is easy to use and accessible to anyone wishing to start their work in this field.

Since the publication of the article about the BSD, the BSD was accessed more than 500 times by more than 300 different users worldwide. Most of these visits came from China and the United States of America.[19]

[19] Data from Google Analytics reports

This page was intentionally left blank.

D. Final Remarks

Science never solves a problem
without creating ten more.
George Shaw

This page was intentionally left blank.

The outcome of this Ph.D. project led to the production of new scientific knowledge with impact in the enzymatic mechanisms field, particularly for PLP-dependent and Mo-containing enzymes. Moreover, some contributions were also provided in terms of scientific software that are widely available for the community. Such software development aims to fulfill the requirement for some tools that were not already available. Finally, pedagogic technologies were also developed to shorten the gap between highly specific scientific knowledge and the general population. In sum, all these works support three fundamental pillars: (1) providing tools for scientific research, (2) new scientific knowledge in terms of enzymatic mechanisms, and (3) bringing science closer to society.

Regarding the study of catalytic mechanisms, enzymes from different groups were targeted in this thesis, namely enzymes employing organic or inorganic cofactors, and an enzyme that does not rely on cofactors.

Four studies were conducted on three different PLP-dependent enzymes, which are enzymes that use an organic PLP cofactor. The PLP cofactor is the most ubiquitous organic cofactor in nature. Interestingly, enzymes that employ PLP as a cofactor are not limited to a narrow set of distinct catalytic activities. Instead, PLP-dependent enzymes can catalyze a myriad of different types of reactions involving mainly amino acids as substrate. So far, at least six different types of reactions are assigned to PLP-dependent enzymes: decarboxylation, racemization, transamination, α-, β-, and γ-elimination/replacement. The primary goals of this Ph.D. project were to describe the catalytic mechanism of decarboxylation, α-, and γ-elimination, as well as rationalize how enzymes modulate the PLP's chemistry to catalyze such distinct reactions for quite similar substrates. The obtained results allowed an atomistic description of each studied mechanism, also providing a Gibbs free energy profile for each entire catalytic process. These results disclosed the importance of key residues for the specificity and catalytic activity of each enzyme. In addition, the role of PLP as a cofactor spans beyond its electron sink feature. The results also revealed some interesting features like the fact that the PLP cofactor is capable of transiently store protons through two distinct mechanisms, and the phosphate group can also act as a base and it is not limited to the anchoring role within the active site.

Apart from the study of PLP-dependent enzymes, two different catalytic mechanisms of XO were studied in strict collaboration with the results obtained experimentally. XO is an enzyme that employs an inorganic cofactor, the Mo ion, and catalyzes the oxidation of hypoxanthine to xanthine and xanthine to uric acid. Currently, the XO activity is already controlled through the use of inhibitors in cases of uric acid-related diseases, e.g., gout. However, the complete description of how XO catalyzes the oxidation of xanthine into

uric acid was not fully known and it can impact the development of better inhibitors. The computational study conducted on the catalytic mechanism of XO described its entire process, including details about key residues for the mechanism. This new data could help in the development of better inhibitors to treat a set of diseases with a wide incidence worldwide. Meanwhile, some experiments have reported an alternative mechanism for XO. In some conditions, XO could also act as a reductase, catalyzing the reduction of nitrate into nitrite and nitrite into nitric oxide. This reaction is seen as a critical source of nitric oxide, which is an important biological signaling molecule. Therefore, the mechanisms underlying these two reactions were also studied in the course of this thesis. The results provided, for the first time, the atomistic description of the reduction of nitrate and nitrite by XO. This description provides a better understanding of how XO contributes to ameliorate the impact of ischemic events. Moreover, it also raises an important alert for the possible consequences of the use of XO inhibitors.

In addition to the study of enzymes employing an organic or inorganic cofactor, and as a result of the current pandemic situation by COVID-19, the catalytic mechanism of the SARS-CoV-2 M^{pro} was studied. The M^{pro} is a protease essential for the replication of the SARS-CoV-2 virus and it is seen as a potential drug target against this viral infection. A computational protocol was followed to obtain the atomistic description of M^{pro} catalytic mechanism, looking for valuable insights for the rational design of inhibitors. The mechanistic proposal resulting from the calculations unveiled the importance of a set of residues and interactions that can now be explored for design specific inhibitors.

The computational methods are a valuable asset for the description of chemical reactions with atomistic detail, which is something almost impossible using other experimental methodologies. Although the pinnacle of a computational methodology would be the use *ab initio* methods to describe, with an extensive conformational sample, the chemistry of an entire molecular system, this is still not possible based on the currently available computer hardware. Therefore, some approaches need to be taken into consideration to achieve a good commitment between speed and accuracy. The most common strategy is using the Nobel-awarded methodology of running hybrid simulations, e.g., ONIOM QM/MM. The development of QM/MM methods allowed the application of computational methodologies to biomolecular systems with extreme relevance in the health, environmental, and industrial fields. In this strategy, the most important chemical species for the reaction are included in the region treated with QM, whereas the remaining system is treated with MM. Thus, the QM region should be big enough to accommodate all relevant species for the catalysis but needs to consider the hardware limitations in order to get a good balance between computational cost and

accuracy. For example, in PLP-dependent enzymes, a net of important residues that interact directly with the PLP cofactor need to be included in addition to the ones absolutely essential for catalysis. Consequently, the atoms included in the QM regions need to be carefully selected to provide an accurate description in a feasible time scale. Moreover, when the available data about the mechanism and the enzyme is limited, it is crucial to ensure that the reactional space is properly explored. This means that several different mechanistic hypotheses need to be assessed in order to ensure that the proposed mechanism is the most likely one to occur *in vivo*. Consequently, several different reactions are tested for each step of the mechanism, looking for the most favored one in terms of kinetic and thermodynamic magnitudes. From our experience, wrong mechanistic proposals lead to high energy barriers which are incompatible with the mild conditions in the cell. Moreover, the available experimental kinetic data provides an important checkpoint to validate the energies calculated computationally. The accuracy of energy calculations is imperative to establish a reasonable comparison between the computed energies and the experimental data. For example, activation Gibbs free energies of 16 kcal/mol, 17 kcal/mol, and 18 kcal/mol correspond to turnover numbers of 18.4 s^{-1}, 3.5 s^{-1}, and 0.7 s^{-1}, respectively (Eyring equation; Temperature = 30 °C). A quite small difference in terms of energy estimation originates significantly differences in turnover numbers. Therefore, the energy calculations performed in this project were refined with the recently available DLPNO-CCSD(T) approach that allows decreasing the energy error to less than 1 kcal/mol with an affordable computational cost (according to the published benchmarks [235-237]).

It is true that the powerfulness of computational studies has been rising due to the development of more powerful computational hardware. However, this advance is also supported by the scientific community through the development and update of software and algorithms. This is essential to take full profit from the most recent hardware implementations but also to provide more accurate and efficient code. Unlike other widely used software, scientific software usually targets a significantly smaller community of users. Consequently, there are few companies devoted to the development of scientific software, and a significant part of the software is developed by the scientific community itself. For example, at the beginning of this Ph.D. project, no powerful and efficient software was freely available for the analysis and preparation of Gaussian-related files. Therefore, molUP was developed, which is a plugin that allows taking advantage of a powerful molecular visualization software (VMD), providing the required tools for the input files preparation, analysis of calculation outputs, or manipulation of molecular structures, for example. This open-source software is now

used in all QM/MM studies of our research group and worldwide by several other researchers. This is just an example of the importance of software development within the scientific community. Researchers usually have a better understanding of their needs and have a deeper knowledge about their particular field. If they can develop skills on code development, researchers would have all the required tools to achieve a piece of software that fulfills their needs. All research fields can benefit from the use of software to improve how data is acquired, analyzed, or visualized. For a computational chemist, this statement is even more relevant, making software development an important aspect in this research field.

In addition to scientific software, some pedagogic technologies were developed during the course of this Ph.D. project. A (scientifically) cultured society is a better society because it makes people more aware of global problems, such as climate changes or the importance of reducing the use of antibiotics for strictly necessary situations. Therefore, experts from different scientific fields can also contribute to the scientific literacy of the general population. In that sense, technologies as BioSIM^{AR} and Protein Wars were developed to improve how some complex chemical knowledge is explained to non-expert people.

E. References

This page was intentionally left blank.

1. Nelson, D.L. and M.M. Cox, *Lehninger Principles of Biochemistry SEVENTH EDITION*. 2017, WH Freeman and Company.
2. Arcus, V.L., M.W. van der Kamp, C.R. Pudney, and A.J. Mulholland, *Enzyme evolution and the temperature dependence of enzyme catalysis.* Curr Opin Struct Biol, 2020. **65**: p. 96-101.
3. Punekar, N.S., *Enzymes: Their Place in Biology*, in *ENZYMES: Catalysis, Kinetics and Mechanisms*. 2018. p. 3-4.
4. Punekar, N.S., *On Enzyme Nomenclature and Classification*, in *ENZYMES: Catalysis, Kinetics and Mechanisms*. 2018. p. 33-41.
5. Harding, M.M., M.W. Nowicki, and M.D. Walkinshaw, *Metals in protein structures: a review of their principal features.* Crystallography Reviews, 2010. **16**(4): p. 247-302.
6. Fernandes, H.S., C.S.S. Teixeira, S.F. Sousa, and N. Cerqueira, *Formation of Unstable and very Reactive Chemical Species Catalyzed by Metalloenzymes: A Mechanistic Overview.* Molecules, 2019. **24**(13).
7. Goodsell, D.S., *Glucose Oxidase.* RCSB Protein Data Bank, 2006.
8. Griffiths, H., *Plant biology: designs on Rubisco.* Nature, 2006. **441**(7096): p. 940-1.
9. Nishigaki, I., H. Ichinose, K. Tamai, and T. Nagatsu, *Purification of aromatic L-amino acid decarboxylase from bovine brain with a monoclonal antibody.* The Biochemical journal, 1988. **252**(2): p. 331-5.
10. Papatsirou, M., P.G. Adamopoulos, P.I. Artemaki, V.P. Georganti, A. Scorilas, D. Vassilacopoulou, and C.K. Kontos, *Next-generation sequencing reveals alternative L-DOPA decarboxylase (DDC) splice variants bearing novel exons, in human hepatocellular and lung cancer cells.* Gene, 2021. **768**: p. 145262.
11. di Salvo, M.L., R. Contestabile, and M.K. Safo, *Vitamin B(6) salvage enzymes: mechanism, structure and regulation.* Biochim Biophys Acta, 2011. **1814**(11): p. 1597-608.
12. Percudani, R. and A. Peracchi, *A genomic overview of pyridoxal-phosphate-dependent enzymes.* EMBO Rep, 2003. **4**(9): p. 850-4.
13. Schomburg, I., A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg, *BRENDA, the enzyme database: updates and major new developments.* Nucleic acids research, 2004. **32**(Database issue): p. D431-3.
14. McCormick, D.B. and H. Chen, *Update on interconversions of vitamin B-6 with its coenzyme.* J Nutr, 1999. **129**(2): p. 325-7.
15. Clayton, P.T., *B6-responsive disorders: a model of vitamin dependency.* J Inherit Metab Dis, 2006. **29**(2-3): p. 317-26.
16. Mills, P.B., R.A. Surtees, M.P. Champion, C.E. Beesley, N. Dalton, P.J. Scambler, S.J. Heales, A. Briddon, I. Scheimberg, G.F. Hoffmann, J. Zschocke, and P.T. Clayton, *Neonatal epileptic encephalopathy caused by mutations in the PNPO gene encoding pyridox(am)ine 5'-phosphate oxidase.* Hum Mol Genet, 2005. **14**(8): p. 1077-86.
17. Coburn, S.P., J.D. Mahuren, and W.E. Schaltenbrand, *Increased activity of pyridoxal kinase in tongue in Down's syndrome.* J Ment Defic Res, 1991. **35 (Pt 6)**: p. 543-7.
18. Song, H., S. Ueno, S. Numata, J. Iga, S. Shibuya-Tayoshi, M. Nakataki, S. Tayoshi, K. Yamauchi, S. Sumitani, T. Tomotake, T. Tada, T. Tanahashi, M.

Itakura, and T. Ohmori, *Association between PNPO and schizophrenia in the Japanese population.* Schizophr Res, 2007. **97**(1-3): p. 264-70.

19. Laine-Cessac, P., A. Cailleux, and P. Allain, *Mechanisms of the inhibition of human erythrocyte pyridoxal kinase by drugs.* Biochemical pharmacology, 1997. **54**(8): p. 863-70.
20. Kastner, U., C. Hallmen, M. Wiese, E. Leistner, and C. Drewke, *The human pyridoxal kinase, a plausible target for ginkgotoxin from Ginkgo biloba.* Febs j, 2007. **274**(4): p. 1036-45.
21. Seto, T., H. Inada, N. Kobayashi, H. Tada, K. Furukawa, K. Hayashi, H. Hattori, O. Matsuoka, and G. Isshiki, *[Depression of serum pyridoxal levels in theophylline-related seizures].* No To Hattatsu, 2000. **32**(4): p. 295-300.
22. Delport, R., J.B. Ubbink, W.J. Vermaak, and P.J. Becker, *Theophylline increases pyridoxal kinase activity independently from vitamin B6 nutritional status.* Res Commun Chem Pathol Pharmacol, 1993. **79**(3): p. 325-33.
23. Kuwahara, H., Y. Noguchi, A. Inaba, and H. Mizusawa, *Case of an 81-year-old woman with theophylline-associated seizures followed by partial seizures due to vitamin B6 deficiency.* Rinsho shinkeigaku= Clinical neurology, 2008. **48**(2): p. 125-129.
24. Alao, A.O. and J.C. Yolles, *Isoniazid-induced psychosis.* Ann Pharmacother, 1998. **32**(9): p. 889-91.
25. Windebank, A.J., *Neurotoxicity of Pyridoxine Analogs Is Related to Coenzyme Structure.* Neurochemical pathology, 1985. **3**(3): p. 159-167.
26. Toney, M.D., *Controlling reaction specificity in pyridoxal phosphate enzymes.* Biochim Biophys Acta, 2011. **1814**(11): p. 1407-18.
27. Oliveira, E.F., N.M. Cerqueira, P.A. Fernandes, and M.J. Ramos, *Mechanism of formation of the internal aldimine in pyridoxal 5'-phosphate-dependent enzymes.* J Am Chem Soc, 2011. **133**(39): p. 15496-505.
28. Cerqueira, N.M., P.A. Fernandes, and M.J. Ramos, *Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5'-Phosphate-Requiring Enzymes.* J Chem Theory Comput, 2011. **7**(5): p. 1356-68.
29. Li, Z., Y. Zhao, H. Zhou, H.B. Luo, and C.G. Zhan, *Catalytic Roles of Coenzyme Pyridoxal-5'-phosphate (PLP) in PLP-dependent Enzymes: Reaction Pathway for Methionine-gamma-lyase-catalyzed L-methionine Depletion.* ACS Catal, 2020. **10**(3): p. 2198-2210.
30. Dutta Banik, S., A. Bankura, and A. Chandra, *A QM/MM simulation study of transamination reaction at the active site of aspartate aminotransferase: Free energy landscape and proton transfer pathways.* J Comput Chem, 2020. **41**(32): p. 2684-2694.
31. Soniya, K. and A. Chandra, *Free energy landscapes of prototropic tautomerism in pyridoxal 5'-phosphate schiff bases at the active site of an enzyme in aqueous medium.* J Comput Chem, 2018. **39**(21): p. 1629-1638.
32. Soniya, K., S. Awasthi, N.N. Nair, and A. Chandra, *Transimination Reaction at the Active Site of Aspartate Aminotransferase: A Proton Hopping Mechanism through Pyridoxal 5 '-Phosphate.* ACS Catalysis, 2019. **9**(7): p. 6276-6283.
33. Locasale, J.W., *Serine, glycine and one-carbon units: cancer metabolism in full circle.* Nat Rev Cancer, 2013. **13**(8): p. 572-83.

34. Rossi, M., A. Amaretti, and S. Raimondi, *Folate production by probiotic bacteria.* Nutrients, 2011. **3**(1): p. 118-34.

35. Abali, E.E., N.E. Skacel, H. Celikkaya, and Y.C. Hsieh, *Chapter 9 Regulation of Human Dihydrofolate Reductase Activity and Expression*, in *Folic Acid and Folates*. 2008, Academic Press. p. 267-292.

36. Newman, A.C. and O.D.K. Maddocks, *One-carbon metabolism in cancer.* Br J Cancer, 2017. **116**(12): p. 1499-1504.

37. Carreras, C.W. and D.V. Santi, *The catalytic mechanism and structure of thymidylate synthase.* Annu Rev Biochem, 1995. **64**: p. 721-62.

38. Smulders, Y.M. and C.D. Stehouwer, *Folate metabolism and cardiovascular disease.* Semin Vasc Med, 2005. **5**(2): p. 87-97.

39. Scaglione, F. and G. Panzavolta, *Folate, folic acid and 5-methyltetrahydrofolate are not the same thing.* Xenobiotica, 2014. **44**(5): p. 480-8.

40. Maia, L.B., I. Moura, and J.J.G. Moura, *Molybdenum and tungsten-containing formate dehydrogenases: Aiming to inspire a catalyst for carbon dioxide utilization.* Inorganica Chimica Acta, 2017. **455**: p. 350-363.

41. Zhang, Y., S. Rump, and V.N. Gladyshev, *Comparative Genomics and Evolution of Molybdenum Utilization.* Coord Chem Rev, 2011. **255**(9-10): p. 1206-1217.

42. Hille, R., *The Mononuclear Molybdenum Enzymes.* Chem Rev, 1996. **96**(7): p. 2757-2816.

43. Schwarz, G., R.R. Mendel, and M.W. Ribbe, *Molybdenum cofactors, enzymes and pathways.* Nature, 2009. **460**(7257): p. 839-47.

44. Mendel, R.R. and T. Kruse, *Cell biology of molybdenum in plants and humans.* Biochim Biophys Acta, 2012. **1823**(9): p. 1568-79.

45. Hille, R., *The molybdenum oxotransferases and related enzymes.* Dalton Trans, 2013. **42**(9): p. 3029-42.

46. Hille, R., J. Hall, and P. Basu, *The mononuclear molybdenum enzymes.* Chem Rev, 2014. **114**(7): p. 3963-4038.

47. Peters, J.W. and R.K. Szilagyi, *Exploring new frontiers of nitrogenase structure and mechanism.* Curr Opin Chem Biol, 2006. **10**(2): p. 101-8.

48. Green, D.E. and H. Beinert, *Xanthine oxidase, a molybdo-flavoprotein.* Biochim Biophys Acta, 1953. **11**(4): p. 599-600.

49. Renzo, E.d., E. Kaleita, P. Heytler, J. Oleson, B. Hutchings, and J. Williams, *The nature of the xanthine oxidase factor.* Journal of the American Chemical Society, 1953. **75**(3): p. 753-753.

50. Richert, D.A. and W.W. Westerfeld, *Isolation and Identification of the Xanthine Oxidase Factor as Molybdenum.* Journal of Biological Chemistry, 1953. **203**(2): p. 915-923.

51. Totter, J.R., W.T. Burnett, R.A. Monroe, I.B. Whitney, and C.L. Comar, *Evidence That Molybdenum Is a Nondialyzable Component of Xanthine Oxidase.* Science, 1953. **118**(3071): p. 555-555.

52. Avis, P.G., F. Bergel, R.C. Bray, and K.V. Shooter, *A crystalline material with xanthine oxidase activity.* Nature, 1954. **173**(4417): p. 1230-1.

53. Bray, R.C., B.G. Malmstrom, and T. Vanngard, *The chemistry of xanthine oxidase. Electron-spin resonance of xanthine oxidase solutions.* The Biochemical journal, 1959. **73**(1): p. 193-7.

54. Bray, R.C., *The chemistry of xanthine oxidase. 8. Electronspin-resonance measurements during the enzymic reaction.* The Biochemical journal, 1961. **81**(1): p. 196-9.

55. Maia, L.B., I. Moura, and J.J.G. Moura, *CHAPTER 1. Molybdenum and Tungsten-Containing Enzymes: An Overview*, in *Molybdenum and Tungsten Enzymes*. 2016, The Royal Society of Chemistry. p. 1-80.

56. Schardinger, F., *Ueber das Verhalten der Kuhmilch gegen Methylenblau und seine Verwendung zur Unterscheidung von ungekochter und gekochter Milch.* Zeitschrift für Untersuchung der Nahrungs- und Genussmittel, 1902. **5**(22): p. 1113-1121.

57. Morgan, E.J., C.P. Stewart, and F.G. Hopkins, *On the anaerobic and aerobic oxidation of xanthin and hypoxanthin by tissues and by milk.* Proceedings of the Royal Society of London Series B-Containing Papers of a Biological Character, 1922. **94**(657): p. 109-131.

58. Majumdar, A., *Structural and functional models in molybdenum and tungsten bioinorganic chemistry: description of selected model complexes, present scenario and possible future scopes.* Dalton Trans, 2014. **43**(24): p. 8990-9003.

59. Basu, P. and S.J. Burgmayer, *Recent developments in the study of molybdoenzyme models.* J Biol Inorg Chem, 2015. **20**(2): p. 373-83.

60. Okamoto, K., T. Kusano, and T. Nishino, *Chemical nature and reaction mechanisms of the molybdenum cofactor of xanthine oxidoreductase.* Curr Pharm Des, 2013. **19**(14): p. 2606-14.

61. Corte, E.D. and F. Stirpe, *Regulation of xanthine oxidase in rat liver: modifications of the enzyme activity of rat liver supernatant on storage at 20 degrees.* Biochemical Journal, 1968. **108**(2): p. 349-351.

62. Stirpe, F. and E. Della Corte, *The regulation of rat liver xanthine oxidase. Conversion in vitro of the enzyme activity from dehydrogenase (type D) to oxidase (type O).* J Biol Chem, 1969. **244**(14): p. 3855-63.

63. Battelli, M.G., E. Lorenzoni, and F. Stirpe, *Milk Xanthine-Oxidase Type D (Dehydrogenase) and Type O (Oxidase) - Purification, Interconversion and Some Properties.* Biochemical Journal, 1973. **131**(2): p. 191-198.

64. Waud, W.R. and K.V. Rajagopalan, *The mechanism of conversion of rat liver xanthine dehydrogenase from an NAD+-dependent form (type D) to an O2-dependent form (type O).* Arch Biochem Biophys, 1976. **172**(2): p. 365-79.

65. Kaminski, Z.W. and M.M. Jezewska, *Involvement of a single thiol group in the conversion of the NAD+-dependent activity of rat liver xanthine oxidoreductase to the O2-dependent activity.* The Biochemical journal, 1982. **207**(2): p. 341-6.

66. Nakamura, M. and I. Yamazaki, *Preparation of Bovine-Milk Xanthine-Oxidase as a Dehydrogenase Form.* Journal of Biochemistry, 1982. **92**(4): p. 1279-1286.

67. Kuwabara, Y., T. Nishino, K. Okamoto, T. Matsumura, B.T. Eger, E.F. Pai, and T. Nishino, *Unique amino acids cluster for switching from the dehydrogenase to oxidase form of xanthine oxidoreductase.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8170-5.

68. Ishikita, H., B.T. Eger, K. Okamoto, T. Nishino, and E.F. Pai, *Protein conformational gating of enzymatic activity in xanthine oxidoreductase.* J Am Chem Soc, 2012. **134**(2): p. 999-1009.

69. Enroth, C., B.T. Eger, K. Okamoto, T. Nishino, T. Nishino, and E.F. Pai, *Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase:*

structure-based mechanism of conversion. Proc Natl Acad Sci U S A, 2000. **97**(20): p. 10723-8.

70. Okamoto, K., K. Matsumoto, R. Hille, B.T. Eger, E.F. Pai, and T. Nishino, *The crystal structure of xanthine oxidoreductase during catalysis: implications for reaction mechanism and enzyme inhibition.* Proc Natl Acad Sci U S A, 2004. **101**(21): p. 7931-6.

71. Stein, B.W. and M.L. Kirk, *Electronic structure contributions to reactivity in xanthine oxidase family enzymes.* J Biol Inorg Chem, 2015. **20**(2): p. 183-94.

72. Yamaguchi, Y., T. Matsumura, K. Ichida, K. Okamoto, and T. Nishino, *Human xanthine oxidase changes its substrate specificity to aldehyde oxidase type upon mutation of amino acid residues in the active site: roles of active site residues in binding and activation of purine substrate.* J Biochem, 2007. **141**(4): p. 513-24.

73. Metz, S. and W. Thiel, *A combined QM/MM study on the reductive half-reaction of xanthine oxidase: substrate orientation and mechanism.* J Am Chem Soc, 2009. **131**(41): p. 14885-902.

74. Metz, S. and W. Thiel, *QM/MM studies of xanthine oxidase: variations of cofactor, substrate, and active-site Glu802.* J Phys Chem B, 2010. **114**(3): p. 1506-17.

75. Lundberg, J.O., E. Weitzberg, and M.T. Gladwin, *The nitrate-nitrite-nitric oxide pathway in physiology and therapeutics.* Nat Rev Drug Discov, 2008. **7**(2): p. 156-67.

76. Maia, L.B. and J.J.G. Moura, *Putting xanthine oxidoreductase and aldehyde oxidase on the NO metabolism map: Nitrite reduction by molybdoenzymes.* Redox Biol, 2018. **19**: p. 274-289.

77. Maia, L.B., V. Pereira, L. Mira, and J.J. Moura, *Nitrite reductase activity of rat and human xanthine oxidase, xanthine dehydrogenase, and aldehyde oxidase: evaluation of their contribution to NO formation in vivo.* Biochemistry, 2015. **54**(3): p. 685-710.

78. Maia, L.B. and J.J.G. Moura, *CHAPTER 17. Lessons from Denitrification for the Human Metabolism of Signalling Nitric Oxide*, in *Metalloenzymes in Denitrification*. 2016. p. 419-443.

79. Webb, A., R. Bond, P. McLean, R. Uppal, N. Benjamin, and A. Ahluwalia, *Reduction of nitrite to nitric oxide during ischemia protects against myocardial ischemia-reperfusion damage.* Proc Natl Acad Sci U S A, 2004. **101**(37): p. 13683-8.

80. Baker, J.E., J. Su, X. Fu, A. Hsu, G.J. Gross, J.S. Tweddell, and N. Hogg, *Nitrite confers protection against myocardial infarction: role of xanthine oxidoreductase, NADPH oxidase and K(ATP) channels.* J Mol Cell Cardiol, 2007. **43**(4): p. 437-44.

81. Hendgen-Cotta, U.B., M.W. Merx, S. Shiva, J. Schmitz, S. Becher, J.P. Klare, H.J. Steinhoff, A. Goedecke, J. Schrader, M.T. Gladwin, M. Kelm, and T. Rassaf, *Nitrite reductase activity of myoglobin regulates respiration and cellular viability in myocardial ischemia-reperfusion injury.* Proc Natl Acad Sci U S A, 2008. **105**(29): p. 10256-61.

82. Shiva, S., Z. Huang, R. Grubina, J. Sun, L.A. Ringwood, P.H. MacArthur, X. Xu, E. Murphy, V.M. Darley-Usmar, and M.T. Gladwin, *Deoxymyoglobin is a nitrite reductase that generates nitric oxide and regulates mitochondrial respiration.* Circ Res, 2007. **100**(5): p. 654-61.

83. Rassaf, T., U. Flogel, C. Drexhage, U. Hendgen-Cotta, M. Kelm, and J. Schrader, *Nitrite reductase function of deoxymyoglobin: oxygen sensor and regulator of cardiac energetics and function.* Circ Res, 2007. **100**(12): p. 1749-54.

84. Maia, L.B. and J.J. Moura, *Nitrite reduction by xanthine oxidase family enzymes: a new class of nitrite reductases.* J Biol Inorg Chem, 2011. **16**(3): p. 443-60.

85. Millar, T.M., C.R. Stevens, N. Benjamin, R. Eisenthal, R. Harrison, and D.R. Blake, *Xanthine oxidoreductase catalyses the reduction of nitrates and nitrite to nitric oxide under hypoxic conditions.* FEBS Lett, 1998. **427**(2): p. 225-8.

86. Zhang, Z., D. Naughton, P.G. Winyard, N. Benjamin, D.R. Blake, and M.C. Symons, *Generation of nitric oxide by a nitrite reductase activity of xanthine oxidase: a potential pathway for nitric oxide formation in the absence of nitric oxide synthase activity.* Biochem Biophys Res Commun, 1998. **249**(3): p. 767-72.

87. Godber, B.L., J.J. Doel, G.P. Sapkota, D.R. Blake, C.R. Stevens, R. Eisenthal, and R. Harrison, *Reduction of nitrite to nitric oxide catalyzed by xanthine oxidoreductase.* J Biol Chem, 2000. **275**(11): p. 7757-63.

88. Li, H., A. Samouilov, X. Liu, and J.L. Zweier, *Characterization of the magnitude and kinetics of xanthine oxidase-catalyzed nitrite reduction. Evaluation of its role in nitric oxide generation in anoxic tissues.* J Biol Chem, 2001. **276**(27): p. 24482-9.

89. Li, H., A. Samouilov, X. Liu, and J.L. Zweier, *Characterization of the effects of oxygen on xanthine oxidase-mediated nitric oxide formation.* J Biol Chem, 2004. **279**(17): p. 16939-46.

90. Li, H., H. Cui, T.K. Kundu, W. Alzawahra, and J.L. Zweier, *Nitric oxide production from nitrite occurs primarily in tissues not in the blood: critical role of xanthine oxidase and aldehyde oxidase.* J Biol Chem, 2008. **283**(26): p. 17855-63.

91. Li, H., T.K. Kundu, and J.L. Zweier, *Characterization of the magnitude and mechanism of aldehyde oxidase-mediated nitric oxide production from nitrite.* J Biol Chem, 2009. **284**(49): p. 33850-8.

92. Maia, L.B. and J.J. Moura, *Detection of Nitric Oxide by Electron Paramagnetic Resonance Spectroscopy: Spin-Trapping with Iron-Dithiocarbamates.* Methods Mol Biol, 2016. **1424**: p. 81-102.

93. Gashaw, I., P. Ellinghaus, A. Sommer, and K. Asadullah, *What makes a good drug target?* Drug Discov Today, 2011. **16**(23-24): p. 1037-43.

94. Brule, D., G. Sarwar, and L. Savoie, *Changes in serum and urinary uric acid levels in normal human subjects fed purine-rich foods containing different amounts of adenine and hypoxanthine.* J Am Coll Nutr, 1992. **11**(3): p. 353-8.

95. Becker, M.A., H.R. Schumacher, Jr., R.L. Wortmann, P.A. MacDonald, D. Eustace, W.A. Palo, J. Streit, and N. Joseph-Ridge, *Febuxostat compared with allopurinol in patients with hyperuricemia and gout.* N Engl J Med, 2005. **353**(23): p. 2450-61.

96. Becker, M.A., H.R. Schumacher, Jr., R.L. Wortmann, P.A. MacDonald, W.A. Palo, D. Eustace, L. Vernillet, and N. Joseph-Ridge, *Febuxostat, a novel nonpurine selective inhibitor of xanthine oxidase: a twenty-eight-day, multicenter, phase II, randomized, double-blind, placebo-controlled, dose-response clinical trial examining safety and efficacy in patients with gout.* Arthritis Rheum, 2005. **52**(3): p. 916-23.

97. Schumacher, H.R., Jr., *Febuxostat: a non-purine, selective inhibitor of xanthine oxidase for the management of hyperuricaemia in patients with gout.* Expert Opin Investig Drugs, 2005. **14**(7): p. 893-903.

98. Chen, L.X. and H.R. Schumacher, *Gout: an evidence-based review.* J Clin Rheumatol, 2008. **14**(5 Suppl): p. S55-62.

99. Richette, P. and T. Bardin, *Gout.* Lancet, 2010. **375**(9711): p. 318-28.

100. Gutowska-Owsiak, D., L. Greenwald, C. Watson, T.A. Selvakumar, X. Wang, and G.S. Ogg, *The histamine-synthesizing enzyme histidine decarboxylase is upregulated by keratinocytes in atopic skin.* Br J Dermatol, 2014. **171**(4): p. 771-8.

101. Tamaka, K., M. Seike, T. Hagiwara, A. Sato, and H. Ohtsu, *Histamine suppresses regulatory T cells mediated by TGF-beta in murine chronic allergic contact dermatitis.* Exp Dermatol, 2015. **24**(4): p. 280-4.

102. Gervasini, G., J.A. Agundez, J. Garcia-Menaya, C. Martinez, C. Cordobes, P. Ayuso, J.A. Cornejo, M. Blanca, and E. Garcia-Martin, *Variability of the L-Histidine decarboxylase gene in allergic rhinitis.* Allergy, 2010. **65**(12): p. 1576-84.

103. Guth, P.H. and P. Hall, *Microcirculatory and mast cell changes in restraint-induced gastric ulcer.* Gastroenterology, 1966. **50**(4): p. 562-70.

104. Parmar, N.S., G. Hennings, and O.P. Gulati, *Histidine decarboxylase inhibition: a novel approach towards the development of an effective and safe gastric anti-ulcer drug.* Agents Actions, 1984. **15**(5-6): p. 494-9.

105. Alkan, M., F. Machavoine, R. Rignault, J. Dam, M. Dy, and N. Thieblemont, *Histidine Decarboxylase Deficiency Prevents Autoimmune Diabetes in NOD Mice.* J Diabetes Res, 2015. **2015**: p. 965056.

106. Mackins, C.J., S. Kano, N. Seyedi, U. Schafer, A.C. Reid, T. Machida, R.B. Silver, and R. Levi, *Cardiac mast cell-derived renin promotes local angiotensin formation, norepinephrine release, and arrhythmias in ischemia/reperfusion.* The Journal of clinical investigation, 2006. **116**(4): p. 1063-70.

107. Darvas, Z., E. Sakurai, H.G. Schwelberger, H. Hegyesi, E. Rivera, H. Othsu, T. Watanabe, E. Pallinger, and A. Falus, *Autonomous histamine metabolism in human melanoma cells.* Melanoma Res, 2003. **13**(3): p. 239-46.

108. He, G.H., J.J. Lin, W.K. Cai, W.M. Xu, Z.P. Yu, S.J. Yin, C.H. Zhao, and G.L. Xu, *Associations of polymorphisms in histidine decarboxylase, histamine N-methyltransferase and histamine receptor H3 genes with breast cancer.* PLOS ONE, 2014. **9**(5): p. e97728.

109. Boer, K., Z. Darvas, M. Baki, I. Kaszas, Z. Pal, and A. Falus, *Expression of histidine decarboxylase in human colonic cancer cells and adenomatous polyps.* Inflamm Res, 2003. **52 Suppl 1**(1): p. S76-7.

110. Tanimoto, A., Y. Matsuki, T. Tomita, T. Sasaguri, S. Shimajiri, and Y. Sasaguri, *Histidine decarboxylase expression in pancreatic endocrine cells and related tumors.* Pathology international, 2004. **54**(6): p. 408-12.

111. Cui, J., G. Xu, J. Liu, Z. Pang, J. Florholmen, and G. Cui, *The expression of non-mast histamine in tumor associated microvessels in human colorectal cancers.* Pathology oncology research : POR, 2013. **19**(2): p. 311-6.

112. Robertson, J.G., *Enzymes as a special class of therapeutic target: clinical drugs and modes of action.* Curr Opin Struct Biol, 2007. **17**(6): p. 674-9.

113. Hopkins, A.L. and C.R. Groom, *The druggable genome.* Nat Rev Drug Discov, 2002. **1**(9): p. 727-30.

114. Ramsay, R.R. and K.F. Tipton, *Assessment of Enzyme Inhibition: A Review with Examples from the Development of Monoamine Oxidase and Cholinesterase Inhibitory Drugs.* Molecules, 2017. **22**(7).

115. Drews, J., *Drug discovery: a historical perspective.* Science, 2000. **287**(5460): p. 1960-4.

116. Brown, D.G. and J. Bostrom, *Where Do Recent Small Molecule Clinical Development Candidates Come From?* J Med Chem, 2018. **61**(21): p. 9442-9468.

117. Lloyd, M.D., *High-Throughput Screening for the Discovery of Enzyme Inhibitors.* J Med Chem, 2020. **63**(19): p. 10742-10772.

118. Schramm, V.L., *Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes.* Annu Rev Biochem, 2011. **80**: p. 703-32.

119. Garcia-Viloca, M., J. Gao, M. Karplus, and D.G. Truhlar, *How enzymes work: Analysis by modern rate theory and computer simulations.* Science, 2004. **303**(5655): p. 186-195.

120. Kamerlin, S.C., J. Mavri, and A. Warshel, *Examining the case for the effect of barrier compression on tunneling, vibrationally enhanced catalysis, catalytic entropy and related issues.* FEBS Lett, 2010. **584**(13): p. 2759-66.

121. Nagel, Z.D. and J.P. Klinman, *A 21 st century revisionist's view at a turning point in enzymology.* Nature chemical biology, 2009. **5**(8): p. 543-550.

122. Schwartz, S.D. and V.L. Schramm, *Enzymatic transition states and dynamic motion in barrier crossing.* Nat Chem Biol, 2009. **5**(8): p. 551-8.

123. Robertson, J.G., *Mechanistic basis of enzyme-targeted drugs.* Biochemistry, 2005. **44**(15): p. 5561-71.

124. Fernandes, H.S., C.S. Silva Teixeira, P.A. Fernandes, M.J. Ramos, and N.M. Cerqueira, *Amino acid deprivation using enzymes as a targeted therapy for cancer and viral infections.* Expert Opin Ther Pat, 2017. **27**(3): p. 283-297.

125. Roberts, J., *Method of purifying l-asparaginase*. 1971, Google Patents.

126. Ueno, T., K. Ohtawa, K. Mitsui, Y. Kodera, M. Hiroto, A. Matsushima, Y. Inada, and H. Nishimura, *Cell cycle arrest and apoptosis of leukemia cells induced by L-asparaginase.* Leukemia, 1997. **11**(11): p. 1858-61.

127. Emadi, A., H. Zokaee, and E.A. Sausville, *Asparaginase in the treatment of non-ALL hematologic malignancies.* Cancer Chemother Pharmacol, 2014. **73**(5): p. 875-83.

128. Hoffman, R.M., Y. Tan, S. Li, Q. Han, S. Yagi, T. Takakura, A. Takimoto, K. Inagaki, and D. Kudou, *Development of Recombinant Methioninase for Cancer Treatment.* Methods Mol Biol, 2019. **1866**: p. 107-131.

129. Sharma, B., S. Singh, and S.S. Kanwar, *L-methionase: a therapeutic enzyme to treat malignancies.* Biomed Res Int, 2014. **2014**: p. 506287.

130. Tan, Y.Y., J. Zavala, M.X. Xu, J. Zavala, and R.M. Hoffman, *Serum methionine depletion without side effects by methioninase in metastatic breast cancer patients.* Anticancer research, 1996. **16**(6c): p. 3937-3942.

131. Tan, Y., J. Zavala, Sr., Q. Han, M. Xu, X. Sun, X. Tan, X. Tan, R. Magana, J. Geller, and R.M. Hoffman, *Recombinant methioninase infusion reduces the*

biochemical endpoint of serum methionine with minimal toxicity in high-stage cancer patients. Anticancer Res, 1997. **17**(5B): p. 3857-60.

132. Kawaguchi, K., K. Miyake, Q.H. Han, S.K. Li, Y.Y. Tan, K. Igarashi, T. Kiyuna, M. Miyake, T. Higuchi, H. Oshiro, Z.Y. Zhang, S. Razmjooei, S. Wangsiricharoen, M. Bouvet, S.R. Singh, M. Unno, and R.M. Hoffman, *Oral recombinant methioninase (o-rMETase) is superior to injectable rMETase and overcomes acquired gemcitabine resistance in pancreatic cancer.* Cancer Letters, 2018. **432**: p. 251-259.

133. Tan, Y., Z. Yang, X. Sun, S. Li, Q. Han, and M. Xu, *Lyase Treatment For P. Carinii*. 2007, ANTICANCER INC: US.

134. Hoffman, R.M. and Q. Han, *Oral Methioninase for Covid-19 Methionine-restriction Therapy.* In Vivo, 2020. **34**(3 Suppl): p. 1593-1596.

135. Dillon, B.J., V.G. Prieto, S.A. Curley, C.M. Ensor, F.W. Holtsberg, J.S. Bomalaski, and M.A. Clark, *Incidence and distribution of argininosuccinate synthetase deficiency in human cancers: a method for identifying cancers sensitive to arginine deprivation.* Cancer, 2004. **100**(4): p. 826-33.

136. Qiu, F., J. Huang, and M. Sui, *Targeting arginine metabolism pathway to treat arginine-dependent cancers.* Cancer Lett, 2015. **364**(1): p. 1-7.

137. Avramis, V.I. and L. Cohen, *Pharmaceutical compositions comprising peg-asparaginase for the treatment of hiv infections*. 1999: WO.

138. Tripathi, P. and S. Sinha, *Industrial Biocatalysis: an Insight into Trends and Future Directions.* Current Sustainable/Renewable Energy Reports, 2020. **7**(3): p. 66-72.

139. Schmid, A., J.S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, and B. Witholt, *Industrial biocatalysis today and tomorrow.* Nature, 2001. **409**(6817): p. 258-68.

140. Arnold, F.H., *Directed evolution: Creating biocatalysts for the future.* Chemical engineering science, 1996. **51**(23): p. 5091-5102.

141. *The Nobel Prize | Women who changed science | Frances H. Arnold*. 2018 [cited 2020 2020-12-02]; Available from: https://www.nobelprize.org/womenwhochangedscience/stories/frances-arnold.

142. Lutz, S., *Beyond directed evolution--semi-rational protein engineering and design.* Curr Opin Biotechnol, 2010. **21**(6): p. 734-43.

143. Hernandez, K., I. Zelen, G. Petrillo, I. Uson, C.M. Wandtke, J. Bujons, J. Joglar, T. Parella, and P. Clapes, *Engineered L-serine hydroxymethyltransferase from Streptococcus thermophilus for the synthesis of alpha,alpha-dialkyl-alpha-amino acids.* Angew Chem Int Ed Engl, 2015. **54**(10): p. 3013-7.

144. Cramer, C.J., *Essentials of Computational Chemistry: Theories and Models*. 2nd ed. 2004: Wiley.

145. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**(5): p. 1589-615.

146. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.

147. Wang, J., W. Wang, P.A. Kollman, and D.A. Case, *Automatic atom type and bond type perception in molecular mechanical calculations.* J Mol Graph Model, 2006. **25**(2): p. 247-60.

148. Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-74.

149. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.

150. Maier, J.A., C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, and C. Simmerling, *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.

151. Brooks, B.R., C.L. Brooks, 3rd, A.D. Mackerell, Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: the biomolecular simulation program.* J Comput Chem, 2009. **30**(10): p. 1545-614.

152. Jorgensen, W.L., D.S. Maxwell, and J. TiradoRives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.* Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.

153. Jorgensen, W.L. and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.* J Am Chem Soc, 1988. **110**(6): p. 1657-66.

154. Schmid, N., A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7.* Eur Biophys J, 2011. **40**(7): p. 843-56.

155. D.A. Case, K.B., I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F.Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R.Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman, *AMBER 2020*. 2020, University of California: San Francisco.

156. Lorentz, H.A., *Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase.* Annalen der Physik, 1881. **248**(1): p. 127-136.

157. Mackerell, A.D., Jr., *Empirical force fields for biological macromolecules: overview and issues.* J Comput Chem, 2004. **25**(13): p. 1584-604.

158. Debye, P., *Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index.* Mathematische Annalen, 1909. **67**(4): p. 535-558.

159. Haug, E.J., J.S. Arora, and K. Matsui, *A steepest-descent method for optimization of mechanical systems.* Journal of Optimization Theory and Applications, 1976. **19**(3): p. 401-424.

160. Hestenes, M.R. and E. Stiefel, *Methods of Conjugate Gradients for Solving Linear Systems.* Journal of Research of the National Bureau of Standards, 1952. **49**(6): p. 409-436.

161. Jensen, F., *Introduction to Computational Chemistry*. 2 ed. 2007, England: John Wiley & Sons Ltd.

162. Berendsen, H.J.C., J.P.M. Postma, W.F. Vangunsteren, A. Dinola, and J.R. Haak, *Molecular-Dynamics with Coupling to an External Bath.* Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.

163. Langevin, P., *Sur la théorie du mouvement brownien.* Comptes-rendus de l'Académie des Sciences, 1908. **146**: p. 530-532.

164. Lemons, D.S. and A. Gythiel, *Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)].* American Journal of Physics, 1997. **65**(11): p. 1079-1081.

165. Izaguirre, J.A., D.P. Catarello, J.M. Wozniak, and R.D. Skeel, *Langevin stabilization of molecular dynamics.* The Journal of Chemical Physics, 2001. **114**(5): p. 2090-2098.

166. Paquet, E. and H.L. Viktor, *Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review.* Biomed Res Int, 2015. **2015**: p. 183918.

167. Jensen, F., *Introduction to Computational Chemistry*. 3rd ed. 2017: Wiley.

168. Friedman, P.A.R., *Molecular Quantum Mechanics*. 5th ed. 2011: Oxford University Press.

169. Levine, I.N., *Quantum Chemistry*. 7th ed. 2014: Pearson.

170. Lewars, E.G., *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. 3rd ed. 2016: Springer.

171. Kulik, H.J., J. Zhang, J.P. Klinman, and T.J. Martinez, *How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase.* J Phys Chem B, 2016. **120**(44): p. 11381-11394.

172. Hertz, H., *Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung.* Annalen der Physik und Chemie, 1887. **267**(8): p. 983-1000.

173. Planck, M., *Ueber die Elementarquanta der Materie und der Elektricität.* Annalen der Physik, 1901. **309**(3): p. 564-566.

174. Einstein, A., *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt.* Annalen der Physik, 1905. **322**(6): p. 132-148.

175. Bohr, N., *XXXVII. On the constitution of atoms and molecules.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2009. **26**(153): p. 476-502.

176. Bohr, N., *I. On the constitution of atoms and molecules.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2009. **26**(151): p. 1-25.

177. Weinberger, P., *Revisiting Louis de Broglie's famous 1924 paper in thePhilosophical Magazine.* Philosophical Magazine Letters, 2006. **86**(7): p. 405-410.

178. Juffmann, T., A. Milic, M. Mullneritsch, P. Asenbaum, A. Tsukernik, J. Tuxen, M. Mayor, O. Cheshnovsky, and M. Arndt, *Real-time single-molecule imaging of quantum interference.* Nat Nanotechnol, 2012. **7**(5): p. 297-300.

179. Heisenberg, W., *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, 1927. **43**(3-4): p. 172-198.

180. Schrödinger, E., *Quantisierung als Eigenwertproblem.* Annalen der Physik, 1926. **384**(4): p. 361-376.

181. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln.* Annalen der Physik, 1927. **389**(20): p. 457-484.

182. Ditchfield, R., W.J. Hehre, and J.A. Pople, *Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules.* The Journal of Chemical Physics, 1971. **54**(2): p. 724-728.

183. Schäfer, A., C. Huber, and R. Ahlrichs, *Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr.* The Journal of Chemical Physics, 1994. **100**(8): p. 5829-5835.

184. Schäfer, A., H. Horn, and R. Ahlrichs, *Fully optimized contracted Gaussian basis sets for atoms Li to Kr.* The Journal of Chemical Physics, 1992. **97**(4): p. 2571-2577.

185. Dunning, T.H., *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen.* The Journal of Chemical Physics, 1989. **90**(2): p. 1007-1023.

186. Kendall, R.A., T.H. Dunning, and R.J. Harrison, *Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions.* The Journal of Chemical Physics, 1992. **96**(9): p. 6796-6806.

187. Woon, D.E. and T.H. Dunning, *Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon.* The Journal of Chemical Physics, 1993. **98**(2): p. 1358-1371.

188. Fernandes, H.S., M.J. Ramos, and N.M.F.S.A. Cerqueira, *Catalytic Mechanism of the Serine Hydroxymethyltransferase: A Computational ONIOM QM/MM Study.* ACS Catalysis, 2018. **8**(11): p. 10096-10110.

189. Fernandes, H.S., M.J. Ramos, and N. Cerqueira, *The Catalytic Mechanism of the Pyridoxal-5'-phosphate-Dependent Enzyme, Histidine Decarboxylase: A Computational Study.* Chemistry (Weinheim an der Bergstrasse, Germany), 2017. **23**(38): p. 9162-9173.

190. Silva, C.F., K.B. Borges, and C.S. Nascimento, Jr., *Computational study on acetamiprid-molecular imprinted polymer.* J Mol Model, 2019. **25**(4): p. 104.

191. Greenberg, A., A.R. Green, and J.F. Liebman, *Computational Study of Selected Amine and Lactam N-Oxides Including Comparisons of NO Bond Dissociation Enthalpies with Those of Pyridine N-Oxides.* Molecules, 2020. **25**(16): p. 3703.

192. Luo, D., R. Jin, X. Han, and K. Li, *Computational study on optoelectronic and charge transport properties of diketopyrrolopyrrole-based A-D-A-D-A structure molecules for organic solar cells.* J Mol Model, 2019. **25**(11): p. 339.

193. Yuqi, B. and L. Chang, *Response of inorganic Al12N12 nano-cage toward nitrosyl hydride: A computational study.* Inorganic Chemistry Communications, 2020. **119**: p. 108069.

194. Almatarneh, M.H., I.A. Elayan, M. Altarawneh, and J.W. Hollett, *A computational study of the ozonolysis of sabinene.* Theoretical Chemistry Accounts, 2019. **138**(2): p. 1-14.

195. Ribeiro, P.M., H.S. Fernandes, L.B. Maia, S. Sousa, J.J.J. Moura, and N.M. Cerqueira, *The complete catalytic mechanism of Xanthine Oxidase: a computational study.* Inorganic Chemistry Frontiers, 2020.

196. Teixeira, C.S.S., S.F. Sousa, and N.M. Cerqueira. *Molecular dynamic simulations and QM/MM studies addressed to build an active Tryptophan Synthase model. A critical enzyme to treat tuberculosis*. in *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG)*. 2019. IEEE.

197. Wadt, W. and H. PJ, *Ab initio effective core potentials formolecular calculations. Potentials for K to Au including theoutermost core orbitals.* J Chem. Phys, 1985. **82**: p. 299-305.

198. Hehre, W.J., *Ab-Initio Molecular-Orbital Theory.* Accounts of Chemical Research, 1976. **9**(11): p. 399-406.

199. Hay, P.J., *A binitio studies of excited states of polyatomic molecules including spin-orbit and multiplet effects: The electronic states of UF6.* The Journal of Chemical Physics, 1983. **79**(11): p. 5469-5482.

200. Shavitt, I., *Methods of Electronic Structure Theory (Modern theoretical chemistry)*. 1977, Springer.

201. Ortiz, J., P.J. Hay, and R.L. Martin, *Role of d and f orbitals in the geometries of low-valent actinide compounds. Ab initio studies of U (CH3) 3, Np (CH3) 3, and Pu (CH3) 3.* Journal of the American Chemical Society, 1992. **114**(7): p. 2736-2737.

202. Dinda, S., *Computational Study of Enantioselectivity in the Asymmetric Allylation of Aldehydes with Chiral Pt(II) Phosphinite Complexes.* J Org Chem, 2018. **83**(22): p. 13911-13921.

203. Samanta, P.N. and K.K. Das, *Comparative QM/MM studies of H2*

adsorption on lithium doped single walled armchair and zigzag nanotubes: SiCNT, GeCNT, and SnCNT. International Journal of Quantum Chemistry, 2016. **116**(20): p. 1467-1476.

204. Drici, N. and M.A. Krallafa, *Effect of mutation on the stabilization energy of HIV-1 zinc fingers: a hybrid local self-consistent field/molecular mechanics investigation.* J Biol Inorg Chem, 2017. **22**(1): p. 109-119.

205. Samanta, P.N. and K.K. Das, *QM/MM study of the interaction between zigzag SnC nanotube and small toxic gas molecules.* International Journal of Quantum Chemistry, 2016. **116**(6): p. 411-420.

206. Zou, H.Y., D.X. Zhao, and Z.Z. Yang, *Studies on the Structural Properties of NAMI-A in Aqueous Solution by Combined QM and ABEEM/MM Method.* Acta Chimica Sinica, 2013. **71**(11): p. 1547-1552.

207. Kumar, A. and A.G. Samuelson, *Computational tools for mechanistic discrimination in the reductive and metathesis coupling reactions mediated by titanium(IV) isopropoxide.* Journal of Chemical Sciences, 2013. **124**(6): p. 1343-1352.

208. Dans, P.D., A. Crespo, D.A. Estrin, and E.L. Coitino, *Structural and Energetic Study of Cisplatin and Derivatives: Comparison of the Performance of Density Funtional Theory Implementations.* J Chem Theory Comput, 2008. **4**(5): p. 740-50.

209. Pauli, W., *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren.* Zeitschrift für Physik, 1925. **31**(1): p. 765-783.

210. Slater, J.C., *The theory of complex spectra.* Physical Review, 1929. **34**(10): p. 1293-1322.

211. Hartree, D.R., *The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion.* Mathematical Proceedings of the Cambridge Philosophical Society, 2008. **24**(1): p. 111-132.

212. Møller, C. and M.S. Plesset, *Note on an Approximation Treatment for Many-Electron Systems.* Physical Review, 1934. **46**(7): p. 618-622.

213. Čížek, J., *On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods.* The Journal of Chemical Physics, 1966. **45**(11): p. 4256-4266.

214. Grimme, S. and P.R. Schreiner, *Computational Chemistry: The Fate of Current Methods and Future Challenges.* Angew Chem Int Ed Engl, 2018. **57**(16): p. 4170-4176.

215. Lyakh, D.I., M. Musial, V.F. Lotrich, and R.J. Bartlett, *Multireference nature of chemistry: the coupled-cluster view.* Chem Rev, 2012. **112**(1): p. 182-243.

216. Mallick, S., B. Roy, and P. Kumar, *A comparison of DLPNO-CCSD(T) and CCSD(T) method for the determination of the energetics of hydrogen atom transfer reactions.* Computational and Theoretical Chemistry, 2020. **1187**.

217. Liakos, D.G. and F. Neese, *Is It Possible To Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory.* J Chem Theory Comput, 2015. **11**(9): p. 4054-63.

218. Purvis, G.D. and R.J. Bartlett, *A full coupled-cluster singles and doubles model: The inclusion of disconnected triples.* The Journal of Chemical Physics, 1982. **76**(4): p. 1910-1918.

219. Sahu, N., G. Singh, A. Nandi, and S.R. Gadre, *Toward an Accurate and Inexpensive Estimation of CCSD(T)/CBS Binding Energies of Large Water Clusters.* J Phys Chem A, 2016. **120**(28): p. 5706-14.

220. Adler, T.B., G. Knizia, and H.J. Werner, *A simple and efficient CCSD(T)-F12 approximation.* J Chem Phys, 2007. **127**(22): p. 221106.

221. Fliegl, H., W. Klopper, and C. Hattig, *Coupled-cluster theory with simplified linear-r(12) corrections: the CCSD(R12) model.* J Chem Phys, 2005. **122**(8): p. 84107.

222. Knizia, G., T.B. Adler, and H.J. Werner, *Simplified CCSD(T)-F12 methods: theory and benchmarks.* J Chem Phys, 2009. **130**(5): p. 054104.

223. Tew, D.P., W. Klopper, C. Neiss, and C. Hattig, *Quintuple-zeta quality coupled-cluster correlation energies with triple-zeta basis sets.* Phys Chem Chem Phys, 2007. **9**(16): p. 1921-30.

224. Werner, H.J., G. Knizia, and F.R. Manby, *Explicitly correlated coupled cluster methods with pair-specific geminals.* Molecular Physics, 2011. **109**(3): p. 407-417.

225. Scuseria, G.E. and P.Y. Ayala, *Linear scaling coupled cluster and perturbation theories in the atomic orbital basis.* Journal of Chemical Physics, 1999. **111**(18): p. 8330-8343.

226. Pulay, P., *Localizability of Dynamic Electron Correlation.* Chemical Physics Letters, 1983. **100**(2): p. 151-154.

227. Schütz, M., *Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T).* The Journal of Chemical Physics, 2000. **113**(22): p. 9986-10001.

228. Hetzer, G., M. Schutz, H. Stoll, and H.J. Werner, *Low-order scaling local correlation methods II: Splitting the Coulomb operator in linear scaling local second-order Moller-Plesset perturbation theory.* Journal of Chemical Physics, 2000. **113**(21): p. 9443-9455.

229. Adler, T.B. and H.J. Werner, *Local explicitly correlated coupled-cluster methods: efficient removal of the basis set incompleteness and domain errors.* J Chem Phys, 2009. **130**(24): p. 241101.

230. Adler, T.B. and H.J. Werner, *An explicitly correlated local coupled cluster method for calculations of large molecules close to the basis set limit.* J Chem Phys, 2011. **135**(14): p. 144117.

231. Ma, Q.L. and H.J. Werner, *Explicitly correlated local coupled-cluster methods using pair natural orbitals.* Wiley Interdisciplinary Reviews-Computational Molecular Science, 2018. **8**(6): p. e1371.

232. Guo, Y., C. Riplinger, U. Becker, D.G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese, *Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)].* J Chem Phys, 2018. **148**(1): p. 011101.

233. Riplinger, C. and F. Neese, *An efficient and near linear scaling pair natural orbital based local coupled cluster method.* J Chem Phys, 2013. **138**(3): p. 034106.

234. Riplinger, C., B. Sandhoefer, A. Hansen, and F. Neese, *Natural triple excitations in local coupled cluster calculations with pair natural orbitals.* J Chem Phys, 2013. **139**(13): p. 134101.

235. Liakos, D.G., M. Sparta, M.K. Kesharwani, J.M. Martin, and F. Neese, *Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory.* J Chem Theory Comput, 2015. **11**(4): p. 1525-39.

236. Wappett, D.A. and L. Goerigk, *Toward a Quantum-Chemical Benchmark Set for Enzymatically Catalyzed Reactions: Important Steps and Insights.* J Phys Chem A, 2019. **123**(32): p. 7057-7074.

237. Paiva, P., M.J. Ramos, and P.A. Fernandes, *Assessing the validity of DLPNO-CCSD(T) in the calculation of activation and reaction energies of ubiquitous enzymatic reactions.* J Comput Chem, 2020. **41**(29): p. 2459-2468.

238. Thomas, L.H., *The calculation of atomic fields.* Mathematical Proceedings of the Cambridge Philosophical Society, 2008. **23**(5): p. 542-548.

239. Fermi, E., *Un metodo statistico per la determinazione di alcune priorieta dell'atome.* Rend. Accad. Naz. Lincei, 1927. **6**(602-607): p. 32.

240. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas.* Physical Review B, 1964. **136**(3b): p. B864-+.

241. Kohn, W. and L.J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects.* Physical Review, 1965. **140**(4a): p. 1133-&.

242. Kohn, W. and L.J. Sham, *Quantum Density Oscillations in an Inhomogeneous Electron Gas.* Physical Review, 1965. **137**(6a): p. 1697-&.

243. Perdew, J.P. and Y. Wang, *Accurate and simple analytic representation of the electron-gas correlation energy.* Phys Rev B Condens Matter, 1992. **45**(23): p. 13244-13249.

244. Becke, A.D., *A new mixing of Hartree–Fock and local density-functional theories.* The Journal of Chemical Physics, 1993. **98**(2): p. 1372-1377.

245. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Phys Rev B Condens Matter, 1988. **37**(2): p. 785-789.

246. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Phys Rev A Gen Phys, 1988. **38**(6): p. 3098-3100.

247. Becke, A.D., *Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing.* The Journal of Chemical Physics, 1996. **104**(3): p. 1040-1046.

248. Cerqueira, N., P.A. Fernandes, and M.J. Ramos, *Protocol for Computational Enzymatic Reactivity Based on Geometry Optimisation.* ChemPhysChem, 2018. **19**(6): p. 669-689.

249. Ochterski, J.W., *Thermochemistry in Gaussian*. 2000, Gaussian, Inc.: gaussian.com. p. 19.

250. Frank Neese, F.W., Daniel Aravena, Michael Atanasov, Ute Becker, Giovanni Bistoni, Dmytro Bykov, Vijay G. Chilkuri, Dipayan Datta, Achintya Kumar Dutta, Dmitry Ganyushin, Miquel Garcia, Yang Guo, Andreas Hansen, Benjamin Helmich-Paris, Lee Huntington, Róbert Izsák, Christian Kollmar, Simone Kossmann, Martin Krupicka, Lucas Lang, Dagmar Lenk, Dimitrios G. Liakos, Dimitrios Manganas, Dimitrios A. Pantazis, Taras Petrenko, Peter Pinski, Christoph Reimann, Marius Retegan, Christoph Riplinger, Tobias Risthaus, Michael Roemelt, Masaaki Saitow, Barbara Sandhöfer, Igor Schapiro, Kantharuban Sivalingam, Bernardo de Souza, Georgi Stoychev, Willem Van den Heuvel, Boris Wezisla, Mihály Kállay, Stefan Grimme, Edward Valeev, Garnet Chan, Jiri Pittner, Martin Brehm, Lars Goerigk, Vilhjálmur Asgeirsson, Wolfgang Schneider, *ORCA - An ab initio, DFT and semiempirical SCF-MO package*. 2019. p. 1048.

251. Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme.* J Mol Biol, 1976. **103**(2): p. 227-49.

252. Reuter, N., A. Dejaegere, B. Maigret, and M. Karplus, *Frontier bonds in QM/MM methods: A comparison of different approaches.* Journal of Physical Chemistry A, 2000. **104**(8): p. 1720-1735.

253. Cui, Q., M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, *A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method.* Journal of Physical Chemistry B, 2001. **105**(2): p. 569-585.

254. Maseras, F. and K. Morokuma, *Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States.* Journal of Computational Chemistry, 1995. **16**(9): p. 1170-1179.

255. Arrhenius, S., *‹ber die Dissociationsw‰rme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte.* Zeitschrift f¸r Physikalische Chemie. **4U**: p. 116 - 96.

256. Eyring, H., *The activated complex in chemical reactions.* Journal of Chemical Physics, 1935. **3**(2): p. 107-115.

257. Fajardo, I., J.L. Urdiales, M.A. Medina, and F. Sanchez-Jimenez, *Effects of phorbol ester and dexamethasone treatment on histidine decarboxylase and*

ornithine decarboxylase in basophilic cells. Biochemical pharmacology, 2001. **61**(9): p. 1101-6.

258. Niijima-Yaoita, F., M. Tsuchiya, H. Ohtsu, K. Yanai, S. Sugawara, Y. Endo, and T. Tadano, *Roles of histamine in exercise-induced fatigue: favouring endurance and protecting against exhaustion.* Biological & pharmaceutical bulletin, 2012. **35**(1): p. 91-7.

259. Na, J.I., K.M. Nam, H.R. Choi, S. Inui, S.W. Youn, C.H. Huh, and K.C. Park, *Histidine decarboxylase expression influences the neofolliculogenesis of newborn mouse dermal cells.* J Dermatol Sci, 2012. **67**(2): p. 95-100.

260. Nakazawa, S., M. Sakanaka, K. Furuta, M. Natsuhara, H. Takano, S. Tsuchiya, Y. Okuno, H. Ohtsu, M. Nishibori, R.L. Thurmond, N. Hirasawa, K. Nakayama, A. Ichikawa, Y. Sugimoto, and S. Tanaka, *Histamine synthesis is required for granule maturation in murine mast cells.* Eur J Immunol, 2014. **44**(1): p. 204-14.

261. Komori, H., Y. Nitta, H. Ueno, and Y. Higuchi, *Structural study reveals that Ser-354 determines substrate specificity on human histidine decarboxylase.* J Biol Chem, 2012. **287**(34): p. 29175-83.

262. Alcaniz, L., A. Vega, P. Chacon, R. El Bekay, I. Ventura, R. Aroca, M. Blanca, D.T. Bergstralh, and J. Monteseirin, *Histamine production by human neutrophils.* FASEB J, 2013. **27**(7): p. 2902-10.

263. Pearce, F.L., *Biological effects of histamine: an overview.* Agents Actions, 1991. **33**(1-2): p. 4-7.

264. Moya-Garcia, A.A., M.A. Medina, and F. Sanchez-Jimenez, *Mammalian histidine decarboxylase: from structure to function.* BioEssays, 2005. **27**(1): p. 57-63.

265. Engel, N., M.T. Olmo, C.S. Coleman, M.A. Medina, A.E. Pegg, and F. Sanchez-Jimenez, *Experimental evidence for structure-activity features in common between mammalian histidine decarboxylase and ornithine decarboxylase.* The Biochemical journal, 1996. **320 (Pt 2)**: p. 365-8.

266. Olmo, M.T., F. Sanchez-Jimenez, M.A. Medina, and H. Hayashi, *Spectroscopic analysis of recombinant rat histidine decarboxylase.* J Biochem, 2002. **132**(3): p. 433-9.

267. Malcolm, B.A. and J.F. Kirsch, *Site-directed mutagenesis of aspartate aminotransferase from E. coli.* Biochem Biophys Res Commun, 1985. **132**(3): p. 915-21.

268. Fleming, J.V., F. Sanchez-Jimenez, A.A. Moya-Garcia, M.R. Langlois, and T.C. Wang, *Mapping of catalytically important residues in the rat L-histidine decarboxylase enzyme using bioinformatic and site-directed mutagenesis approaches.* The Biochemical journal, 2004. **379**(Pt 2): p. 253-61.

269. Kuramitsu, S., Y. Inoue, S. Tanase, Y. Morino, and H. Kagamiyama, *Substitution of an arginyl residue for the active site lysyl residue (Lys258) of aspartate aminotransferase.* Biochem Biophys Res Commun, 1987. **146**(2): p. 416-21.

270. Moya-Garcia, A.A., J. Ruiz-Pernia, S. Marti, F. Sanchez-Jimenez, and I. Tunon, *Analysis of the decarboxylation step in mammalian histidine decarboxylase. A computational study.* J Biol Chem, 2008. **283**(18): p. 12393-401.

271. Rodriguez-Caso, C., D. Rodriguez-Agudo, A.A. Moya-Garcia, I. Fajardo, M.A. Medina, V. Subramaniam, and F. Sanchez-Jimenez, *Local changes in the catalytic site of mammalian histidine decarboxylase can affect its global conformation and stability.* European journal of biochemistry, 2003. **270**(21): p. 4376-87.

272. Komori, H., Y. Nitta, H. Ueno, and Y. Higuchi, *Purification, crystallization and preliminary X-ray analysis of human histidine decarboxylase.* Acta crystallographica. Section F, Structural biology and crystallization communications, 2012. **68**(Pt 6): p. 675-7.

273. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of Simple Potential Functions for Simulating Liquid Water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

274. Price, D.J. and C.L. Brooks, 3rd, *A modified TIP3P water potential for simulation with Ewald summation.* J Chem Phys, 2004. **121**(20): p. 10096-103.

275. Sondergaard, C.R., M.H. Olsson, M. Rostkowski, and J.H. Jensen, *Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values.* J Chem Theory Comput, 2011. **7**(7): p. 2284-95.

276. Olsson, M.H., C.R. Sondergaard, M. Rostkowski, and J.H. Jensen, *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* J Chem Theory Comput, 2011. **7**(2): p. 525-37.

277. Case, D.A., T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman, *AMBER 12*. 2012, University of California, San Francisco.

278. Sorin, E.J. and V.S. Pande, *Exploring the helix-coil transition via all-atom equilibrium ensemble simulations.* Biophys J, 2005. **88**(4): p. 2472-93.

279. Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, and D.E. Shaw, *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-8.

280. Frisch, M.J., G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N.J. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.

281. Spitznagel, G.W., T. Clark, P. von Ragué Schleyer, and W.J. Hehre, *An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl.* Journal of Computational Chemistry, 1987. **8**(8): p. 1109-1116.

282. Hariharan, P.C. and J.A. Pople, *The influence of polarization functions on molecular orbital hydrogenation energies.* Theoretica Chimica Acta, 1973. **28**(3): p. 213-222.

283. Hehre, W.J., R. Ditchfield, and J.A. Pople, *Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in*

Molecular Orbital Studies of Organic Molecules. The Journal of Chemical Physics, 1972. **56**(5): p. 2257-2261.

284. Francl, M.M., W.J. Pietro, W.J. Hehre, J.S. Binkley, M.S. Gordon, D.J. DeFrees, and J.A. Pople, *Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements.* The Journal of Chemical Physics, 1982. **77**(7): p. 3654-3665.

285. Gordon, M.S., J.S. Binkley, J.A. Pople, W.J. Pietro, and W.J. Hehre, *Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements.* Journal of the American Chemical Society, 1982. **104**(10): p. 2797-2803.

286. Zhao, Y. and D.G. Truhlar, *The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals.* Theoretical Chemistry Accounts, 2007. **120**(1-3): p. 215-241.

287. McLean, A.D. and G.S. Chandler, *Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18.* The Journal of Chemical Physics, 1980. **72**(10): p. 5639-5648.

288. Clark, T., J. Chandrasekhar, G.W. Spitznagel, and P.V. Schleyer, *Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. Iii. The 3-21+G Basis Set for First-Row Elements, Li-F.* Journal of Computational Chemistry, 1983. **4**(3): p. 294-301.

289. Frisch, M.J., J.A. Pople, and J.S. Binkley, *Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets.* The Journal of Chemical Physics, 1984. **80**(7): p. 3265-3269.

290. Krishnan, R., J.S. Binkley, R. Seeger, and J.A. Pople, *Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions.* The Journal of Chemical Physics, 1980. **72**(1): p. 650-654.

291. Singh, U.C. and P.A. Kollman, *An Approach to Computing Electrostatic Charges for Molecules.* Journal of Computational Chemistry, 1984. **5**(2): p. 129-145.

292. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

293. Roy Denningto, T.A.K., John M. Millam, *GaussView*. 2009, Semichem Inc. Shawnee Mission KS.

294. Hayashi, H., K. Inoue, T. Nagata, S. Kuramitsu, and H. Kagamiyama, *Escherichia coli aromatic amino acid aminotransferase: characterization and comparison with aspartate aminotransferase.* Biochemistry, 1993. **32**(45): p. 12229-39.

295. Dunathan, H.C., *Conformation and reaction specificity in pyridoxal phosphate enzymes.* Proc Natl Acad Sci U S A, 1966. **55**(4): p. 712-6.

296. Dunathan, H.C., *Stereochemical aspects of pyridoxal phosphate catalysis.* Adv Enzymol Relat Areas Mol Biol, 1971. **35**: p. 79-134.

297. Haldar, K., S. Bhattacharjee, and I. Safeukui, *Drug resistance in Plasmodium.* Nat Rev Microbiol, 2018. **16**(3): p. 156-170.

298. *World malaria report 2019*. 2019, World Health Organization: Geneva.

299. Baragana, B., I. Hallyburton, M.C. Lee, N.R. Norcross, R. Grimaldi, T.D. Otto, W.R. Proto, A.M. Blagborough, S. Meister, G. Wirjanata, A. Ruecker, L.M. Upton, T.S. Abraham, M.J. Almeida, A. Pradhan, A. Porzelle, T. Luksch, M.S. Martinez,

T. Luksch, J.M. Bolscher, A. Woodland, S. Norval, F. Zuccotto, J. Thomas, F. Simeons, L. Stojanovski, M. Osuna-Cabello, P.M. Brock, T.S. Churcher, K.A. Sala, S.E. Zakutansky, M.B. Jimenez-Diaz, L.M. Sanz, J. Riley, R. Basak, M. Campbell, V.M. Avery, R.W. Sauerwein, K.J. Dechering, R. Noviyanti, B. Campo, J.A. Frearson, I. Angulo-Barturen, S. Ferrer-Bazaga, F.J. Gamo, P.G. Wyatt, D. Leroy, P. Siegl, M.J. Delves, D.E. Kyle, S. Wittlin, J. Marfurt, R.N. Price, R.E. Sinden, E.A. Winzeler, S.A. Charman, L. Bebrevska, D.W. Gray, S. Campbell, A.H. Fairlamb, P.A. Willis, J.C. Rayner, D.A. Fidock, K.D. Read, and I.H. Gilbert, *A novel multiple-stage antimalarial agent that inhibits protein synthesis.* Nature, 2015. **522**(7556): p. 315-20.

300. McCarthy, J.S., J. Lotharius, T. Ruckle, S. Chalon, M.A. Phillips, S. Elliott, S. Sekuloski, P. Griffin, C.L. Ng, D.A. Fidock, L. Marquart, N.S. Williams, N. Gobeau, L. Bebrevska, M. Rosario, K. Marsh, and J.J. Mohrle, *Safety, tolerability, pharmacokinetics, and activity of the novel long-acting antimalarial DSM265: a two-part first-in-human phase 1a/1b randomised study.* Lancet Infect Dis, 2017. **17**(6): p. 626-635.

301. Kronenberger, T., J. Lindner, K.A. Meissner, F.M. Zimbres, M.A. Coronado, F.M. Sauer, I. Schettert, and C. Wrenger, *Vitamin B6-dependent enzymes in the human malaria parasite Plasmodium falciparum: a druggable target?* Biomed Res Int, 2014. **2014**: p. 108516.

302. Schwertz, G., M.C. Witschel, M. Rottmann, R. Bonnert, U. Leartsakulpanich, P. Chitnumsub, A. Jaruwat, W. Ittarat, A. Schafer, R.A. Aponte, S.A. Charman, K.L. White, A. Kundu, S. Sadhukhan, M. Lloyd, G.M. Freiberg, M. Srikumaran, M. Siggel, A. Zwyssig, P. Chaiyen, and F. Diederich, *Antimalarial Inhibitors Targeting Serine Hydroxymethyltransferase (SHMT) with in Vivo Efficacy and Analysis of their Binding Mode Based on X-ray Cocrystal Structures.* J Med Chem, 2017. **60**(12): p. 4840-4860.

303. Schwertz, G., M.S. Frei, M.C. Witschel, M. Rottmann, U. Leartsakulpanich, P. Chitnumsub, A. Jaruwat, W. Ittarat, A. Schafer, R.A. Aponte, N. Trapp, K. Mark, P. Chaiyen, and F. Diederich, *Conformational Aspects in the Design of Inhibitors for Serine Hydroxymethyltransferase (SHMT): Biphenyl, Aryl Sulfonamide, and Aryl Sulfone Motifs.* Chemistry (Weinheim an der Bergstrasse, Germany), 2017. **23**(57): p. 14345-14357.

304. Pornthanakasem, W., D. Kongkasuriyachai, C. Uthaipibull, Y. Yuthavong, and U. Leartsakulpanich, *Plasmodium serine hydroxymethyltransferase: indispensability and display of distinct localization.* Malar J, 2012. **11**(1): p. 387.

305. Franca, T.C., P.G. Pascutti, T.C. Ramalho, and J.D. Figueroa-Villar, *A three-dimensional structure of Plasmodium falciparum serine hydroxymethyltransferase in complex with glycine and 5-formyl-tetrahydrofolate. Homology modeling and molecular dynamics.* Biophys Chem, 2005. **115**(1): p. 1-10.

306. Florio, R., R. Chiaraluce, V. Consalvi, A. Paiardini, B. Catacchio, F. Bossa, and R. Contestabile, *The role of evolutionarily conserved hydrophobic contacts in the quaternary structure stability of Escherichia coli serine hydroxymethyltransferase.* Febs j, 2009. **276**(1): p. 132-43.

307. Angelaccio, S., M.L. di Salvo, A. Parroni, A. Di Bello, R. Contestabile, and S. Pascarella, *Structural stability of cold-adapted serine hydroxymethyltransferase, a tool for β-hydroxy-α-amino acid biosynthesis.* Journal of Molecular Catalysis B: Enzymatic, 2014. **110**: p. 171-177.

308. Schirch, V. and D.M. Szebenyi, *Serine hydroxymethyltransferase revisited.* Curr Opin Chem Biol, 2005. **9**(5): p. 482-7.

309. Szebenyi, D.M., F.N. Musayev, M.L. di Salvo, M.K. Safo, and V. Schirch, *Serine hydroxymethyltransferase: role of glu75 and evidence that serine is cleaved by a retroaldol mechanism.* Biochemistry, 2004. **43**(22): p. 6865-76.

310. Scarsdale, J.N., S. Radaev, G. Kazanina, V. Schirch, and H.T. Wright, *Crystal structure at 2.4 A resolution of E. coli serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate.* J Mol Biol, 2000. **296**(1): p. 155-68.

311. Rao, J.V., V. Prakash, N.A. Rao, and H.S. Savithri, *The role of Glu74 and Tyr82 in the reaction catalyzed by sheep liver cytosolic serine hydroxymethyltransferase.* European journal of biochemistry, 2000. **267**(19): p. 5967-76.

312. Kallen, R.G. and W.P. Jencks, *The mechanism of the condensation of formaldehyde with tetrahydrofolic acid.* J Biol Chem, 1966. **241**(24): p. 5851-63.

313. Chitnumsub, P., W. Ittarat, A. Jaruwat, K. Noytanom, W. Amornwatcharapong, W. Pornthanakasem, P. Chaiyen, Y. Yuthavong, and U. Leartsakulpanich, *The structure of Plasmodium falciparum serine hydroxymethyltransferase reveals a novel redox switch that regulates its activities.* Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 6): p. 1517-27.

314. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic acids research, 1997. **25**(17): p. 3389-402.

315. Altschul, S.F., J.C. Wootton, E.M. Gertz, R. Agarwala, A. Morgulis, A.A. Schaffer, and Y.K. Yu, *Protein database searches using compositionally adjusted substitution matrices.* Febs j, 2005. **272**(20): p. 5101-9.

316. H, S.F., M.J. Ramos, and M.F.S.A.C. N, *molUP: A VMD plugin to handle QM and ONIOM calculations using the gaussian software.* J Comput Chem, 2018. **39**(19): p. 1344-1353.

317. Hättig, C., *Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core–valence and quintuple-ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr.* Phys. Chem. Chem. Phys., 2005. **7**(1): p. 59-66.

318. Weigend, F., A. Kohn, and C. Hattig, *Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations.* Journal of Chemical Physics, 2002. **116**(8): p. 3175-3183.

319. Truhlar, D.G., *Basis-set extrapolation.* Chemical Physics Letters, 1998. **294**(1-3): p. 45-48.

320. Ngo, H.P., N.M. Cerqueira, J.K. Kim, M.K. Hong, P.A. Fernandes, M.J. Ramos, and L.W. Kang, *PLP undergoes conformational changes during the course of an enzymatic reaction.* Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 2): p. 596-606.

321. Di Salvo, M.L., J.N. Scarsdale, G. Kazanina, R. Contestabile, V. Schirch, and H.T. Wright, *Structure-based mechanism for early PLP-mediated steps of rabbit cytosolic serine hydroxymethyltransferase reaction.* Biomed Res Int, 2013. **2013**: p. 458571.

322. Jain, M., R. Nilsson, S. Sharma, N. Madhusudhan, T. Kitami, A.L. Souza, R. Kafri, M.W. Kirschner, C.B. Clish, and V.K. Mootha, *Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation.* Science, 2012. **336**(6084): p. 1040-4.

323. Fesko, K., L. Giger, and D. Hilvert, *Synthesis of beta-hydroxy-alpha-amino acids with a reengineered alanine racemase.* Bioorg Med Chem Lett, 2008. **18**(22): p. 5987-90.

324. Fesko, K., M. Uhl, J. Steinreiber, K. Gruber, and H. Griengl, *Biocatalytic access to alpha,alpha-dialkyl-alpha-amino acids by a mechanism-based approach.* Angew Chem Int Ed Engl, 2010. **49**(1): p. 121-4.

325. Windle, C.L., M. Muller, A. Nelson, and A. Berry, *Engineering aldolases as biocatalysts.* Curr Opin Chem Biol, 2014. **19**: p. 25-33.

326. Müller, M., *Recent Developments in Enzymatic Asymmetric C□C Bond Formation.* Advanced Synthesis & Catalysis, 2012. **354**(17): p. 3161-3174.

327. Gutierrez, M.L., X. Garrabou, E. Agosta, S. Servi, T. Parella, J. Joglar, and P. Clapes, *Serine hydroxymethyl transferase from Streptococcus thermophilus and L-threonine aldolase from Escherichia coli as stereocomplementary biocatalysts for the synthesis of beta-hydroxy-alpha,omega-diamino acid derivatives.* Chemistry (Weinheim an der Bergstrasse, Germany), 2008. **14**(15): p. 4647-56.

328. Chiba, Y., T. Terada, M. Kameya, K. Shimizu, H. Arai, M. Ishii, and Y. Igarashi, *Mechanism for folate-independent aldolase reaction catalyzed by serine hydroxymethyltransferase.* Febs j, 2012. **279**(3): p. 504-14.

329. Vidal, L., J. Calveras, P. Clapes, P. Ferrer, and G. Caminal, *Recombinant production of serine hydroxymethyl transferase from Streptococcus thermophilus and its preliminary evaluation as a biocatalyst.* Appl Microbiol Biotechnol, 2005. **68**(4): p. 489-97.

330. Yanai, T., D.P. Tew, and N.C. Handy, *A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP).* Chemical Physics Letters, 2004. **393**(1-3): p. 51-57.

331. Chiku, T., D. Padovani, W. Zhu, S. Singh, V. Vitvitsky, and R. Banerjee, *H2S biogenesis by human cystathionine gamma-lyase leads to the novel sulfur metabolites lanthionine and homolanthionine and is responsive to the grade of hyperhomocysteinemia.* J Biol Chem, 2009. **284**(17): p. 11601-12.

332. Aitken, S.M. and J.F. Kirsch, *The enzymology of cystathionine biosynthesis: strategies for the control of substrate and reaction specificity.* Arch Biochem Biophys, 2005. **433**(1): p. 166-75.

333. Yamanishi, T. and S. Tuboi, *The mechanism of the L-cystine cleavage reaction catalyzed by rat liver gamma-cystathionase.* J Biochem, 1981. **89**(6): p. 1913-21.

334. Wang, R., *Two's company, three's a crowd: can H2S be the third endogenous gaseous transmitter?* The FASEB Journal, 2002. **16**(13): p. 1792-1798.

335. Moore, P.K., M. Bhatia, and S. Moochhala, *Hydrogen sulfide: from the smell of the past to the mediator of the future?* Trends Pharmacol Sci, 2003. **24**(12): p. 609-11.

336. Li, L., M. Bhatia, and P.K. Moore, *Hydrogen sulphide--a novel mediator of inflammation?* Curr Opin Pharmacol, 2006. **6**(2): p. 125-9.

337. Stipanuk, M.H. and P.W. Beck, *Characterization of the enzymic capacity for cysteine desulphhydration in liver and kidney of the rat.* The Biochemical journal, 1982. **206**(2): p. 267-77.

338. Kamoun, P., *Endogenous production of hydrogen sulfide in mammals.* Amino Acids, 2004. **26**(3): p. 243-54.

339. Ishii, I., N. Akahoshi, X.N. Yu, Y. Kobayashi, K. Namekata, G. Komaki, and H. Kimura, *Murine cystathionine gamma-lyase: complete cDNA and genomic sequences, promoter activity, tissue distribution and developmental expression.* The Biochemical journal, 2004. **381**(Pt 1): p. 113-23.

340. Eto, K., M. Ogasawara, K. Umemura, Y. Nagai, and H. Kimura, *Hydrogen sulfide is produced in response to neuronal excitation (Retracted Article. See vol 24, pg 5649, 2004).* Journal of Neuroscience, 2002. **22**(9): p. 3386-3391.

341. Yang, G., L. Wu, S. Bryan, N. Khaper, S. Mani, and R. Wang, *Cystathionine gamma-lyase deficiency and overproliferation of smooth muscle cells.* Cardiovasc Res, 2010. **86**(3): p. 487-95.

342. Yang, G., L. Wu, B. Jiang, W. Yang, J. Qi, K. Cao, Q. Meng, A.K. Mustafa, W. Mu, and S. Zhang, *H2S as a physiologic vasorelaxant: hypertension in mice with deletion of cystathionine γ-lyase.* Science, 2008. **322**(5901): p. 587-590.

343. Zhu, X.Y., S.J. Liu, Y.J. Liu, S. Wang, and X. Ni, *Glucocorticoids suppress cystathionine gamma-lyase expression and H2S production in lipopolysaccharide-treated macrophages.* Cell Mol Life Sci, 2010. **67**(7): p. 1119-32.

344. Untereiner, A.A., R. Wang, Y. Ju, and L. Wu, *Decreased Gluconeogenesis in the Absence of Cystathionine Gamma-Lyase and the Underlying Mechanisms.* Antioxid Redox Signal, 2016. **24**(3): p. 129-40.

345. Tripatara, P., N.S. Patel, V. Brancaleone, D. Renshaw, J. Rocha, B. Sepodes, H. Mota-Filipe, M. Perretti, and C. Thiemermann, *Characterisation of cystathionine gamma-lyase/hydrogen sulphide pathway in ischaemia/reperfusion injury of the mouse kidney: an in vivo study.* Eur J Pharmacol, 2009. **606**(1-3): p. 205-9.

346. Ali, A., Y. Zhang, M. Fu, Y. Pei, L. Wu, R. Wang, and G. Yang, *Cystathionine gamma-lyase/H2S system suppresses hepatic acetyl-CoA accumulation and nonalcoholic fatty liver disease in mice.* Life Sci, 2020. **252**: p. 117661.

347. Cheung, S.H., W.K. Kwok, K.F. To, and J.Y. Lau, *Anti-atherogenic effect of hydrogen sulfide by over-expression of cystathionine gamma-lyase (CSE) gene.* PLOS ONE, 2014. **9**(11): p. e113038.

348. Bibli, S., J. Hu, V. Randriamboavonjy, M. Siragusa, F. Sigala, S. Zukunft, J. Pfeilschifter, I. Wittig, A. Papapetropoulos, and I. Fleming, *P564 L-cystathionine metabolism by cystathionine gamma lyase generates H2S to preserve endothelial cell function and decelerate atherosclerosis development in humans.* Cardiovascular Research, 2018. **114**(suppl_1): p. S138-S139.

349. Wang, Y., X. Zhao, H. Jin, H. Wei, W. Li, D. Bu, X. Tang, Y. Ren, C. Tang, and J. Du, *Role of hydrogen sulfide in the development of atherosclerotic lesions in apolipoprotein E knockout mice.* Arterioscler Thromb Vasc Biol, 2009. **29**(2): p. 173-9.

350. Mard, S.A., I. Ahmadi, A. Ahangarpour, M.K. Gharib-Naseri, and M. Badavi, *Delayed gastric emptying in diabetic rats caused by decreased expression of cystathionine gamma lyase and H2S synthesis: in vitro and in vivo studies.* Neurogastroenterology and Motility, 2016. **28**(11): p. 1677-1689.

351. Wang, P., L. Wu, Y. Ju, M. Fu, T. Shuang, Z. Qian, and R. Wang, *Age-Dependent Allergic Asthma Development and Cystathionine Gamma-Lyase Deficiency.* Antioxid Redox Signal, 2017. **27**(13): p. 931-944.

352. Ahmad, A., N. Druzhyna, and C. Szabo, *Cystathionine-gamma-lyase deficient mice are protected against the development of multiorgan failure and exhibit reduced inflammatory response during burn.* Burns, 2017. **43**(5): p. 1021-1033.

353. Qu, Z., Y. Jiang, B.Q. Wu, Y.F. Duan, Z.D. Sun, and G.H. Luo, *Cystathionine-gamma-lyase inhibitor attenuates acute lung injury induced by acute pancreatitis in rats.* Arch Med Sci, 2014. **10**(4): p. 825-9.

354. Dam, V.P., J.L. Scott, A. Ross, and R.T. Kinobe, *Inhibition of cystathionine gamma-lyase and the biosynthesis of endogenous hydrogen sulphide ameliorates gentamicin-induced nephrotoxicity.* Eur J Pharmacol, 2012. **685**(1-3): p. 165-73.

355. Chen, L., C. Lu, and Y. Hua, *Cystathionine gamma-lyase aggravates periodontal damage in traumatic occlusion mouse models.* J Periodontal Res, 2020. **55**(5): p. 667-675.

356. Fu, M., W. Zhang, G. Yang, and R. Wang, *Is cystathionine gamma-lyase protein expressed in the heart?* Biochem Biophys Res Commun, 2012. **428**(4): p. 469-74.

357. Sun, Q., R. Collins, S. Huang, L. Holmberg-Schiavone, G.S. Anand, C.H. Tan, S. van-den-Berg, L.W. Deng, P.K. Moore, T. Karlberg, and J. Sivaraman, *Structural basis for the inhibition mechanism of human cystathionine gamma-lyase, an enzyme responsible for the production of H(2)S.* J Biol Chem, 2009. **284**(5): p. 3076-3085.

358. Zhu, W., A. Lin, and R. Banerjee, *Kinetic properties of polymorphic variants and pathogenic mutants in human cystathionine gamma-lyase.* Biochemistry, 2008. **47**(23): p. 6226-32.

359. Huang, S., J.H. Chua, W.S. Yew, J. Sivaraman, P.K. Moore, C.H. Tan, and L.W. Deng, *Site-directed mutagenesis on human cystathionine-gamma-lyase reveals insights into the modulation of H2S production.* J Mol Biol, 2010. **396**(3): p. 708-18.

360. Steegborn, C., T. Clausen, P. Sondermann, U. Jacob, M. Worbs, S. Marinkovic, R. Huber, and M.C. Wahl, *Kinetics and inhibition of recombinant human cystathionine gamma-lyase. Toward the rational control of transsulfuration.* J Biol Chem, 1999. **274**(18): p. 12675-84.

361. Lee, Z.W., Y.L. Low, S. Huang, T. Wang, and L.W. Deng, *The cystathionine gamma-lyase/hydrogen sulfide system maintains cellular glutathione status.* The Biochemical journal, 2014. **460**(3): p. 425-35.

362. Wang, L.P., T.J. Martinez, and V.S. Pande, *Building Force Fields: An Automatic, Systematic, and Reproducible Approach.* J Phys Chem Lett, 2014. **5**(11): p. 1885-91.

363. D.A. Case, K.B., I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F.Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R.Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman, *AMBER 2018*. 2018, University of California: San Francisco.

364. Wang, L.P., K.A. McKiernan, J. Gomes, K.A. Beauchamp, T. Head-Gordon, J.E. Rice, W.C. Swope, T.J. Martinez, and V.S. Pande, *Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15.* J Phys Chem B, 2017. **121**(16): p. 4023-4039.

365. Miyamoto, S. and P.A. Kollman, *Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models.* Journal of Computational Chemistry, 1992. **13**(8): p. 952-962.

366. Svensson, M., S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2+ H2Oxidative Addition.* The Journal of Physical Chemistry, 1996. **100**(50): p. 19357-19363.

367. Svensson, M., S. Humbel, and K. Morokuma, *Energetics using the single point IMOMO (integrated molecular orbital+molecular orbital) calculations: Choices of computational levels and model system.* The Journal of Chemical Physics, 1996. **105**(9): p. 3654-3661.

368. Dapprich, S., I. Komáromi, K.S. Byun, K. Morokuma, and M.J. Frisch, *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Journal of Molecular Structure: THEOCHEM, 1999. **461-462**: p. 1-21.

369. Fernandes, H.S., S.F. Sousa, and N. Cerqueira, *VMD Store-A VMD Plugin to Browse, Discover, and Install VMD Extensions.* J Chem Inf Model, 2019. **59**(11): p. 4519-4523.

370. Fukui, K., *The path of chemical reactions - the IRC approach.* Accounts of Chemical Research, 2002. **14**(12): p. 363-368.

371. Pauff, J.M. and R. Hille, *Inhibition studies of bovine xanthine oxidase by luteolin, silibinin, quercetin, and curcumin.* J Nat Prod, 2009. **72**(4): p. 725-31.

372. Sakamoto, N., T. Yamamoto, Y. Moriwaki, T. Teranishi, M. Toyoda, Y. Onishi, S. Kuroda, K. Sakaguchi, T. Fujisawa, M. Maeda, and T. Hada, *Identification of a new point mutation in the human xanthine dehydrogenase gene responsible for a case of classical type I xanthinuria.* Hum Genet, 2001. **108**(4): p. 279-83.

373. Ichida, K., Y. Amaya, N. Kamatani, T. Nishino, T. Hosoya, and O. Sakai, *Identification of two mutations in human xanthine dehydrogenase gene responsible for classical type I xanthinuria.* The Journal of clinical investigation, 1997. **99**(10): p. 2391-7.

374. Pacher, P., A. Nivorozhkin, and C. Szabo, *Therapeutic effects of xanthine oxidase inhibitors: renaissance half a century after the discovery of allopurinol.* Pharmacol Rev, 2006. **58**(1): p. 87-114.

375. Smalley, R.V., A. Guaspari, S. Haase-Statz, S.A. Anderson, D. Cederberg, and J.A. Hohneker, *Allopurinol: intravenous use for prevention and treatment of hyperuricemia.* J Clin Oncol, 2000. **18**(8): p. 1758-63.

376. Kela, U. and R. Vijayvargiya, *Studies on the mechanism of action of 6-mercaptopurine. Interaction with copper and xanthine oxidase.* The Biochemical journal, 1981. **193**(3): p. 799-803.

377. Gunduğdu, Ö., S.A.A. Noma, T. Taskin-Tok, B. Ateş, and N. Kishali, *Evaluation of xanthine oxidase inhibitor properties on isoindoline-1,3-dion derivatives and calculation of interaction mechanism.* Journal of Molecular Structure, 2020. **1204**: p. 127523.

378. Huneycutt, E., C. Board, and J.N. Clements, *Lesinurad, a Selective URAT-1 Inhibitor With a Novel Mechanism in Combination With a Xanthine Oxidase Inhibitor, for Hyperuricemia Associated With Gout.* Journal of Pharmacy Practice, 2017. **31**(6): p. 897190017734427.

379. Jordan, A. and U. Gresser, *Side Effects and Interactions of the Xanthine Oxidase Inhibitor Febuxostat.* Pharmaceuticals (Basel), 2018. **11**(2).

380. Kumar, R., G. Joshi, H. Kler, S. Kalra, M. Kaur, and R. Arya, *Toward an Understanding of Structural Insights of Xanthine and Aldehyde Oxidases: An Overview of their Inhibitors and Role in Various Diseases.* Med Res Rev, 2018. **38**(4): p. 1073-1125.

381. Luna, G., A.V. Dolzhenko, and R.L. Mancera, *Inhibitors of Xanthine Oxidase: Scaffold Diversity and Structure-Based Drug Design.* ChemMedChem, 2019. **14**(7): p. 714-743.

382. Malik, N., P. Dhiman, and A. Khatkar, *In Silico and 3D QSAR Studies of Natural Based Derivatives as Xanthine Oxidase Inhibitors.* Curr Top Med Chem, 2019. **19**(2): p. 123-138.

383. Masuoka, N. and I. Kubo, *Characterization of the xanthine oxidase inhibitory activity of alk(en)yl phenols and related compounds.* Phytochemistry, 2018. **155**: p. 100-106.

384. Mehmood, A., M. Ishaq, L. Zhao, B. Safdar, A.U. Rehman, M. Munir, A. Raza, M. Nadeem, W. Iqbal, and C. Wang, *Natural compounds with xanthine oxidase inhibitory activity: A review.* Chem Biol Drug Des, 2019. **93**(4): p. 387-418.

385. Ojha, R., J. Singh, A. Ojha, H. Singh, S. Sharma, and K. Nepali, *An updated patent review: xanthine oxidase inhibitors for the treatment of hyperuricemia and gout (2011-2015).* Expert Opin Ther Pat, 2017. **27**(3): p. 311-345.

386. Global Burden of Disease Study, C., *Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2015. **386**(9995): p. 743-800.

387. Okamoto, K., Y. Kawaguchi, B.T. Eger, E.F. Pai, and T. Nishino, *Crystal structures of urate bound form of xanthine oxidoreductase: substrate orientation and structure of the key reaction intermediate.* J Am Chem Soc, 2010. **132**(48): p. 17080-3.

388. Pearson, A.R., Godber, B.L.J., Eisenthal, R., Taylor, G.L., Harrison, R., *Human Milk Xanthine Dehydrogenase is Incompletely Converted to the Oxidase Form in the Absence of Proteolysis. A Structural Explanation.* 2006.

389. Guex, N., M.C. Peitsch, and T. Schwede, *Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective.* Electrophoresis, 2009. **30 Suppl 1**: p. S162-73.

390. Benkert, P., M. Biasini, and T. Schwede, *Toward the estimation of the absolute quality of individual protein structure models.* Bioinformatics, 2011. **27**(3): p. 343-50.

391. Bienert, S., A. Waterhouse, T.A. de Beer, G. Tauriello, G. Studer, L. Bordoli, and T. Schwede, *The SWISS-MODEL Repository-new features and functionality.* Nucleic acids research, 2017. **45**(D1): p. D313-D319.

392. Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede, *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic acids research, 2018. **46**(W1): p. W296-W303.

393. Ferreira, P., N. Cerqueira, N.F. Bras, P.A. Fernandes, and M.J. Ramos, *Parametrization of Molybdenum Cofactors for the AMBER Force Field.* J Chem Theory Comput, 2018. **14**(5): p. 2538-2548.

394. Zwanzig, R., *Nonlinear generalized Langevin equations.* Journal of Statistical Physics, 1973. **9**(3): p. 215-220.

395. Hay, P.J. and W.R. Wadt, *Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg.* The Journal of Chemical Physics, 1985. **82**(1): p. 270-283.

396. Ferreira, P., N.M.F.S.A. Cerqueira, P.A. Fernandes, M.J. Romao, and M.J. Ramos, *Catalytic Mechanism of Human Aldehyde Oxidase.* ACS Catalysis, 2020. **10**(16): p. 9276-9286.

397. Wu, F., S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E.C. Holmes, and Y.Z. Zhang, *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.

398. Anand, K., J. Ziebuhr, P. Wadhwani, J.R. Mesters, and R. Hilgenfeld, *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.* Science, 2003. **300**(5626): p. 1763-7.

399. Ullrich, S. and C. Nitsche, *The SARS-CoV-2 main protease as drug target.* Bioorg Med Chem Lett, 2020. **30**(17): p. 127377.

400. Zhang, L., D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld, *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors.* Science, 2020. **368**(6489): p. 409-412.

401. Das, S., S. Sarmah, S. Lyndem, and A. Singha Roy, *An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study.* J Biomol Struct Dyn, 2020: p. 1-11.

402. Ngo, S.T., N. Quynh Anh Pham, L. Thi Le, D.H. Pham, and V.V. Vu, *Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease.* J Chem Inf Model, 2020. **60**(12): p. 5771-5780.

403. Ton, A.T., F. Gentile, M. Hsing, F. Ban, and A. Cherkasov, *Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds.* Mol Inform, 2020. **39**(8): p. e2000028.

404. Eleftheriou, P., D. Amanatidou, A. Petrou, and A. Geronikaki, *In Silico Evaluation of the Effectivity of Approved Protease Inhibitors against the Main Protease of the Novel SARS-CoV-2 Virus.* Molecules, 2020. **25**(11).

405. Sacco, M.D., C. Ma, P. Lagarias, A. Gao, J.A. Townsend, X. Meng, P. Dube, X. Zhang, Y. Hu, N. Kitamura, B. Hurst, B. Tarbet, M.T. Marty, A. Kolocouris, Y. Xiang, Y. Chen, and J. Wang, *Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M(pro) and cathepsin L.* Sci Adv, 2020. **6**(50).

406. Lapaillerie, D., C. Charlier, H.S. Fernandes, S.F. Sousa, P. Lesbats, P. Weigel, A. Favereaux, V. Guyonnet-Duperat, and V. Parissi, *In Silico, In Vitro and In Cellulo Models for Monitoring SARS-CoV-2 Spike/Human ACE2 Complex, Viral Entry and Cell Fusion.* Viruses, 2021. **13**(3).

407. Kumar, Y., H. Singh, and C.N. Patel, *In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing.* J Infect Public Health, 2020. **13**(9): p. 1210-1223.

408. Li, Z., X. Li, Y.Y. Huang, Y. Wu, R. Liu, L. Zhou, Y. Lin, D. Wu, L. Zhang, H. Liu, X. Xu, K. Yu, Y. Zhang, J. Cui, C.G. Zhan, X. Wang, and H.B. Luo, *Identify potent*

SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. Proc Natl Acad Sci U S A, 2020. **117**(44): p. 27381-27387.

409. Lv, Z., Y. Chu, and Y. Wang, *HIV protease inhibitors: a review of molecular selectivity and toxicity.* HIV/AIDS (Auckland, N.Z.), 2015. **7**: p. 95-104.

410. Lurie, N., M. Saville, R. Hatchett, and J. Halton, *Developing Covid-19 Vaccines at Pandemic Speed.* N Engl J Med, 2020. **382**(21): p. 1969-1973.

411. Zhang, C., S. Huang, F. Zheng, and Y. Dai, *Controversial treatments: An updated understanding of the coronavirus disease 2019.* J Med Virol, 2020. **92**(9): p. 1441-1448.

412. Zhang, J., H. Zeng, J. Gu, H. Li, L. Zheng, and Q. Zou, *Progress and Prospects on Vaccine Development against SARS-CoV-2.* Vaccines (Basel), 2020. **8**(2).

413. Jin, Z., X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, and H. Yang, *Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors.* Nature, 2020. **582**(7811): p. 289-293.

414. Douangamath, A., D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C.D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Abranyi-Balogh, J. Brandao-Neto, A. Carbery, G. Davison, A. Dias, T.D. Downes, L. Dunnett, M. Fairhead, J.D. Firth, S.P. Jones, A. Keeley, G.M. Keseru, H.F. Klein, M.P. Martin, M.E.M. Noble, P. O'Brien, A. Powell, R.N. Reddi, R. Skyner, M. Snee, M.J. Waring, C. Wild, N. London, F. von Delft, and M.A. Walsh, *Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease.* Nat Commun, 2020. **11**(1): p. 5047.

415. Wang, H., S. He, W. Deng, Y. Zhang, G. Li, J. Sun, W. Zhao, Y. Guo, Z. Yin, D. Li, and L. Shang, *Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease.* ACS Catal, 2020. **10**(10): p. 5871-5890.

416. Su, H.X., S. Yao, W.F. Zhao, M.J. Li, J. Liu, W.J. Shang, H. Xie, C.Q. Ke, H.C. Hu, M.N. Gao, K.Q. Yu, H. Liu, J.S. Shen, W. Tang, L.K. Zhang, G.F. Xiao, L. Ni, D.W. Wang, J.P. Zuo, H.L. Jiang, F. Bai, Y. Wu, Y. Ye, and Y.C. Xu, *Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients.* Acta Pharmacol Sin, 2020. **41**(9): p. 1167-1177.

417. Dai, W., B. Zhang, X.M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.K. Zhang, Y. Xu, H. Yang, and H. Liu, *Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease.* Science, 2020. **368**(6497): p. 1331-1335.

418. Kneller, D.W., G. Phillips, H.M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, and A. Kovalevsky, *Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room temperature X-ray crystallography.* Nat Commun, 2020. **11**(1): p. 3202.

419. Ma, C., M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, and J. Wang, *Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease.* bioRxiv, 2020: p. 2020.04.20.051581.

420. Muramatsu, T., C. Takemoto, Y.T. Kim, H. Wang, W. Nishii, T. Terada, M. Shirouzu, and S. Yokoyama, *SARS-CoV 3CL protease cleaves its C-terminal*

autoprocessing site by novel subsite cooperativity. Proc Natl Acad Sci U S A, 2016. **113**(46): p. 12997-13002.

421. Huynh, T., H. Wang, and B. Luan, *In Silico Exploration of the Molecular Mechanism of Clinically Oriented Drugs for Possibly Inhibiting SARS-CoV-2's Main Protease.* J Phys Chem Lett, 2020. **11**(11): p. 4413-4420.

422. Suarez, D. and N. Diaz, *SARS-CoV-2 Main Protease: A Molecular Dynamics Study.* J Chem Inf Model, 2020. **60**(12): p. 5815-5831.

423. Frances-Monerris, A., C. Hognon, T. Miclot, C. Garcia-Iriepa, I. Iriepa, A. Terenzi, S. Grandemange, G. Barone, M. Marazzi, and A. Monari, *Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches.* J Proteome Res, 2020. **19**(11): p. 4291-4315.

424. Świderek, K. and V. Moliner, *Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods.* Chemical Science, 2020. **11**(39): p. 10626-10630.

425. Ramos-Guzmán, C.A., J.J. Ruiz-Pernía, and I. Tuñón, *Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods.* ACS Catalysis, 2020. **10**(21): p. 12544-12554.

426. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

427. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

428. Fan, K., P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen, and L. Lai, *Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase.* J Biol Chem, 2004. **279**(3): p. 1637-42.

429. Solowiej, J., J.A. Thomson, K. Ryan, C. Luo, M. He, J. Lou, and B.W. Murray, *Steady-state and pre-steady-state kinetic evaluation of severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro cysteine protease: development of an ion-pair model for catalysis.* Biochemistry, 2008. **47**(8): p. 2617-30.

430. Ramos-Guzmán, C.A., J.J. Ruiz-Pernía, and I. Tuñón, *A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design.* Chemical Science, 2021. **12**(10): p. 3489-3496.

431. Arafet, K., N. Serrano-Aparicio, A. Lodola, A.J. Mulholland, F.V. González, K. Świderek, and V. Moliner, *Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity.* Chemical Science, 2021. **12**(4): p. 1433-1444.

432. Lopez, E.D., J.P. Arcon, D.F. Gauto, A.A. Petruk, C.P. Modenutti, V.G. Dumas, M.A. Marti, and A.G. Turjanski, *WATCLUST: a tool for improving the design of drugs based on protein-water interactions.* Bioinformatics, 2015. **31**(22): p. 3697-9.

433. Laurent, B., M. Chavent, T. Cragnolini, A.C. Dahl, S. Pasquali, P. Derreumaux, M.S. Sansom, and M. Baaden, *Epock: rapid analysis of protein pocket dynamics.* Bioinformatics, 2015. **31**(9): p. 1478-80.

434. Scheurer, M., P. Rodenkirch, M. Siggel, R.C. Bernardi, K. Schulten, E. Tajkhorshid, and T. Rudack, *PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations.* Biophys J, 2018. **114**(3): p. 577-583.

435. Cerqueira, N.M.F.S.A., J. Ribeiro, P.A. Fernandes, and M.J. Ramos, *vsLab-An Implementation for Virtual High-Throughput Screening Using AutoDock and VMD.* International Journal of Quantum Chemistry, 2011. **111**(6): p. 1208-1212.

436. Contreras-Riquelme, S., J.A. Garate, T. Perez-Acle, and A.J.M. Martin, *RIP-MD: a tool to study residue interaction networks in protein molecular dynamics.* PeerJ, 2018. **6**(12): p. e5998.

437. Ribeiro, J.V., N.M. Cerqueira, P.A. Fernandes, and M.J. Ramos, *CHEM-PATH-TRACKER: An automated tool to analyze chemical motifs in molecular structures.* Chem Biol Drug Des, 2014. **84**(1): p. 44-53.

438. Giorgino, T., A. Laio, and A. Rodriguez, *METAGUI 3: A graphical user interface for choosing the collective variables in molecular dynamics simulations.* Computer Physics Communications, 2017. **217**: p. 204-209.

439. Cerqueira, N., P.A. Fernandes, and M.J. Ramos, *Visualizing the Microscopic World.* Interdiscip Sci, 2018. **10**(1): p. 105-110.

440. Liu, H. and T. Hou, *CaFE: a tool for binding affinity prediction using end-point free energy methods.* Bioinformatics, 2016. **32**(14): p. 2216-8.

441. Ribeiro, J.V., J.A. Tamames, N.M. Cerqueira, P.A. Fernandes, and M.J. Ramos, *Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems.* Chem Biol Drug Des, 2013. **82**(6): p. 743-55.

442. Vermaas, J.V., D.J. Hardy, J.E. Stone, E. Tajkhorshid, and A. Kohlmeyer, *TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD.* J Chem Inf Model, 2016. **56**(6): p. 1112-6.

443. Knapp, B., N. Lederer, U. Omasits, and W. Schreiner, *vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD.* J Comput Chem, 2010. **31**(16): p. 2868-73.

444. Machado, M.R. and S. Pantano, *SIRAH tools: mapping, backmapping and visualization of coarse-grained models.* Bioinformatics, 2016. **32**(10): p. 1568-70.

445. Li, L., Z. Jia, Y. Peng, A. Chakravorty, L. Sun, and E. Alexov, *DelPhiForce web server: electrostatic forces and energy calculations and visualization.* Bioinformatics, 2017. **33**(22): p. 3661-3663.

446. Balabin, I.A., X. Hu, and D.N. Beratan, *Exploring biological electron transfer pathway dynamics with the Pathways plugin for VMD.* J Comput Chem, 2012. **33**(8): p. 906-10.

447. Mota, C.S., M.G. Rivas, C.D. Brondino, I. Moura, J.J. Moura, P.J. Gonzalez, and N.M. Cerqueira, *The mechanism of formate oxidation by metal-dependent formate dehydrogenases.* J Biol Inorg Chem, 2011. **16**(8): p. 1255-68.

448. Gesto, D.S., N.M. Cerqueira, P.A. Fernandes, and M.J. Ramos, *Unraveling the enigmatic mechanism of L-asparaginase II with QM/QM calculations.* J Am Chem Soc, 2013. **135**(19): p. 7146-58.

449. Cerqueira, N.M.F.S.A., P.J. Gonzalez, P.A. Fernandes, J.J.G. Moura, and M.J. Ramos, *Periplasmic Nitrate Reductase and Formate Dehydrogenase: Similar Molecular Architectures with Very Different Enzymatic Activities.* Accounts of Chemical Research, 2015. **48**(11): p. 2875-2884.

450. Paiva, P., S.F. Sousa, P.A. Fernandes, and M.J. Ramos, *Human Fatty Acid Synthase: A Computational Study of the Transfer of the Acyl Moieties from MAT to the ACP Domain.* ChemCatChem, 2019. **11**(16): p. 3853-3864.

451. Saravanan, K., M. Sivanandam, G. Hunday, M.S. Pavan, and P. Kumaradhas, *Exploring the different environments effect of piperine via combined crystallographic, QM/MM and molecular dynamics simulation study.* Journal of Molecular Graphics and Modelling, 2019. **92**: p. 280-295.

452. Singh, W., D. Quinn, T.S. Moody, and M. Huang, *Reaction mechanism of histone demethylation in αKG-dependent non-heme iron enzymes.* The Journal of Physical Chemistry B, 2019. **123**(37): p. 7801-7811.

453. Zong, Z., Q. Li, Z. Hong, H. Fu, W. Cai, C. Chipot, H. Jiang, D. Zhang, S. Chen, and X. Shao, *Lysine mutation of the claw-arm-like loop accelerates catalysis by cellobiohydrolases.* Journal of the American Chemical Society, 2019. **141**(36): p. 14451-14459.

454. De Raffele, D., S. Martí, and V. Moliner, *QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design.* ACS Catalysis, 2019. **9**(3): p. 2482-2492.

455. Kokcu, Y., S. Kecel-Gunduz, Y. Budama-Kilinc, R. Cakir-Koc, B. Bicak, T. Zorlu, A.E. Ozel, and S. Akyuz, *Structural analysis, molecular dynamics and docking calculations of skin protective tripeptide and design, characterization, cytotoxicity studies of its PLGA nanoparticles.* Journal of Molecular Structure, 2020. **1200**: p. 127046.

456. Liang, J.-W., M.-Y. Wang, S. Wang, S.-L. Li, W.-Q. Li, and F.-H. Meng, *Identification of novel CDK2 inhibitors by a multistage virtual screening method based on SVM, pharmacophore and docking model.* Journal of enzyme inhibition and medicinal chemistry, 2020. **35**(1): p. 235-244.

457. Margreat, S.S., S. Ramalingam, S. Sebastian, S. Xavier, S. Periandy, J.C. Daniel, and M.M. Julie, *DFT, spectroscopic, DSC/TGA, electronic, biological and molecular docking investigation of 2, 5-thiophenedicarboxylic acid: A promising anticancer agent.* Journal of Molecular Structure, 2020. **1200**: p. 127099.

458. Schrödinger, L., *The PyMOL Molecular Graphics System*. 2015, Schrödinger, LLC.

459. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

460. Dias, J.M., M.E. Than, A. Humm, R. Huber, G.P. Bourenkov, H.D. Bartunik, S. Bursakov, J. Calvete, J. Caldeira, C. Carneiro, J.J. Moura, I. Moura, and M.J. Romao, *Crystal structure of the first dissimilatory nitrate reductase at 1.9 A solved by MAD methods.* Structure, 1999. **7**(1): p. 65-79.

461. Axel Kohlmeyer, J.V., Efrem Braun, *akohlmey/topotools: Release 1.8*. 2020, Zenodo.

462. Hanwell, M.D., D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, and G.R. Hutchison, *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform.* J Cheminform, 2012. **4**(1): p. 17.

463. Allouche, A.R., *Gabedit--a graphical user interface for computational chemistry softwares.* J Comput Chem, 2011. **32**(1): p. 174-82.

464. Schaftenaar, G. and J.H. Noordik, *Molden: a pre- and post-processing program for molecular and electronic structures.* J Comput Aided Mol Des, 2000. **14**(2): p. 123-34.

465. Gobert, J.D., L. O'Dwyer, P. Horwitz, B.C. Buckley, S.T. Levy, and U. Wilensky, *Examining the Relationship Between Students' Understanding of the Nature of Models and Conceptual Learning in Biology, Physics, and Chemistry.* International Journal of Science Education, 2010. **33**(5): p. 653-684.

466. Mendonça, P.C.C. and R. Justi, *An instrument for analyzing arguments produced in modeling-based chemistry lessons.* Journal of Research in Science Teaching, 2014. **51**(2): p. 192-218.

467. Hoffmann, R. and P. Laszlo, *Representation in Chemistry.* Angewandte Chemie-International Edition in English, 1991. **30**(1): p. 1-16.

468. Cooper, M.M. and R.L. Stowe, *Chemistry Education Research-From Personal Empiricism to Evidence, Theory, and Informed Practice.* Chem Rev, 2018. **118**(12): p. 6053-6087.

469. Graulich, N., *The tip of the iceberg in organic chemistry classes: how do students deal with the invisible?* Chemistry Education Research and Practice, 2015. **16**(1): p. 9-21.

470. McMurry, J.E., *Organic Chemistry*. 9th ed. 2015: Cengage Learning. 1518.

471. Ruiz-Chica, A.J., J.L. Urdiales, M.A. Medina, and F. Sanchez-Jimenez, *One century after Fischer: better tools for teaching the stereochemistry of carbohydrates.* Biochemical Education, 1999. **27**(1): p. 7-8.

472. Mistry, N., R. Singh, and J. Ridley, *A Web-Based Stereochemistry Tool to Improve Students' Ability to Draw Newman Projections and Chair Conformations and Assign R/S Labels.* Journal of Chemical Education, 2020. **97**(4): p. 1157-1161.

473. Nič, M., J. Jirát, B. Košata, A. Jenkins, and A. McNaught, *IUPAC Compendium of Chemical Terminology*. 2009.

474. Hutchison, J.M., *Improving Translational Accuracy between Dash-Wedge Diagrams and Newman Projections.* Journal of Chemical Education, 2017. **94**(7): p. 892-896.

475. Moreno, L.F., *Understanding Fischer Projection and Angular Line Representation Conversion.* Journal of Chemical Education, 2011. **89**(1): p. 175-176.

476. Eriksen, K., B.E. Nielsen, and M. Pittelkow, *Visualizing 3D Molecular Structures Using an Augmented Reality App.* Journal of Chemical Education, 2020. **97**(5): p. 1487-1490.

477. Mohamed-Salah, B. and D. Alain, *To what degree does handling concrete molecular models promote the ability to translate and coordinate between 2D and 3D molecular structure representations? A case study with Algerian students.* Chemistry Education Research and Practice, 2016. **17**(4): p. 862-877.

478. Wu, H.K. and P. Shah, *Exploring visuospatial thinking in chemistry learning.* Science Education, 2004. **88**(3): p. 465-492.

479. Habraken, C.L., *Integrating into Chemistry Teaching Today's Student's Visuospatial Talents and Skills, and the Teaching of Today's Chemistry's Graphical Language.* Journal of Science Education and Technology, 2004. **13**(1): p. 89-94.

480. Burrmann, N.J. and J.W. Moore, *Development of a Web-Based, Student-Centered Stereochemistry Tutorial.* Journal of Chemical Education, 2013. **90**(12): p. 1622-1625.

481. Kumi, B.C., J.T. Olimpo, F. Bartlett, and B.L. Dixon, *Evaluating the effectiveness of organic chemistry textbooks in promoting representational fluency and understanding of 2D-3D diagrammatic relationships.* Chemistry Education Research and Practice, 2013. **14**(2): p. 177-187.

482. Da Veiga Beltrame, E., J. Tyrwhitt-Drake, I. Roy, R. Shalaby, J. Suckale, and D. Pomeranz Krummel, *3D Printing of Biomolecular Models for Research and Pedagogy.* J Vis Exp, 2017(121).

483. Dickenson, C.E., R.A.R. Blackburn, and R.G. Britton, *3D Printing Workshop Activity That Aids Representation of Molecules and Student Comprehension of Shape and Chirality.* Journal of Chemical Education, 2020. **97**(10): p. 3714-3719.

484. *Avogadro: an open-source molecular builder and visualization tool*.

485. Varetto, U., *Molekel*.

486. Sayle, R.A. and E.J. Milner-White, *RASMOL: biomolecular graphics for all.* Trends Biochem Sci, 1995. **20**(9): p. 374.

487. Bernstein, H.J., *Recent changes to RasMol, recombining the variants.* Trends Biochem Sci, 2000. **25**(9): p. 453-5.

488. Schrödinger, L., *PyMOL*.

489. Niece, B.K., *Custom-Printed 3D Models for Teaching Molecular Symmetry*. Journal of Chemical Education, 2019. **96**(9): p. 2059-2062.

490. Savchenkov, A.V., *Designing Three-Dimensional Models That Can Be Printed on Demand and Used with Students to Facilitate Teaching Molecular Structure, Symmetry, and Related Topics.* Journal of Chemical Education, 2020. **97**(6): p. 1682-1687.

491. Pernaa, J. and S. Wiedmer, *A systematic review of 3D printing in chemistry education – analysis of earlier research and educational use through technological pedagogical content knowledge framework.* Chemistry Teacher International, 2019. **0**(0): p. 20190005.

492. *GVR: Global 3D Printing Market Size was $11.58 Billion in 2019*. 2020 [cited 2020 2020-12-16]; Available from: https://www.printedelectronicsnow.com/contents/view_breaking-news/2020-02-10/gvr-global-3d-printing-market-size-was-1158-billion-in-2019/.

493. *Gartner Says Worldwide Smartphone Sales Will Decline 2.5% in 2019*. 2019 [cited 2020 2020-12-16]; Available from: https://www.gartner.com/en/newsroom/press-releases/2019-08-01-gartner-says-worldwide-smartphone-sales-will-decline-.

494. Bennie, S.J., K.E. Ranaghan, H. Deeks, H.E. Goldsmith, M.B. O'Connor, A.J. Mulholland, and D.R. Glowacki, *Teaching Enzyme Catalysis Using Interactive Molecular Dynamics in Virtual Reality.* Journal of Chemical Education, 2019. **96**(11): p. 2488-2496.

495. O'Connor, M.B., S.J. Bennie, H.M. Deeks, A. Jamieson-Binnie, A.J. Jones, R.J. Shannon, R. Walters, T.J. Mitchell, A.J. Mulholland, and D.R. Glowacki, *Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework.* The Journal of Chemical Physics, 2019. **150**(22): p. 220901.

496. Jiménez, Z.A., *Teaching and Learning Chemistry via Augmented and Immersive Virtual Reality*, in *Technology Integration in Chemistry Education and Research (TICER)*. 2019, American Chemical Society. p. 31-52.

497. Wiebrands, M., C.J. Malajczuk, A.J. Woods, A.L. Rohl, and R.L. Mancera, *Molecular Dynamics Visualization (MDV): Stereoscopic 3D Display of Biomolecular Structure and Interactions Using the Unity Game Engine.* J Integr Bioinform, 2018. **15**(2): p. 20180010.

498. O'Connor, M., H.M. Deeks, E. Dawn, O. Metatla, A. Roudaut, M. Sutton, L.M. Thomas, B.R. Glowacki, R. Sage, and P. Tew, *Sampling molecular conformations and dynamics in a multiuser virtual reality framework.* Science Advances, 2018. **4**(6): p. eaat2731.

499. Yang, S.X., B. Mei, and X.Y. Yue, *Mobile Augmented Reality Assisted Chemical Education: Insights from Elements 4D.* Journal of Chemical Education, 2018. **95**(6): p. 1060-1062.

500. Chatzopoulos, D., C. Bermejo, Z.P. Huang, and P. Hui, *Mobile Augmented Reality Survey: From Where We Are to Where We Go.* IEEE Access, 2017. **5**: p. 6917-6950.

501. Berry, C. and J. Board, *A Protein in the palm of your hand through augmented reality.* Biochem Mol Biol Educ, 2014. **42**(5): p. 446-9.

502. Sung, R.-J., A.T. Wilson, S.M. Lo, L.M. Crowl, J. Nardi, K. St. Clair, and J.M. Liu, *BiochemAR: An Augmented Reality Educational Tool for Teaching Macromolecular Structure and Function.* Journal of Chemical Education, 2019. **97**(1): p. 147-153.

503. Maier, P. and G. Klinker. *Augmented chemical reactions: An augmented reality tool to support chemistry teaching*. in *2013 2nd Experiment@ International Conference (exp.at'13)*. 2013.

504. Seshadri, K., P. Liu, and D.R. Koes, *The 3Dmol.js Learning Environment: A Classroom Response System for 3D Chemical Structures.* Journal of Chemical Education, 2020. **97**(10): p. 3872-3876.

505. Hofmann, A., M.J. Coster, and P. Taylor, *Disseminating a Free, Practical Java Tool To Interactively Generate and Edit 2D Chemical Structures.* Journal of Chemical Education, 2019. **96**(6): p. 1262-1267.

506. *Browser Market Share Worldwide: Jan - Dec 2020*. 2020 [cited 2020 2020-12-16]; Available from: https://gs.statcounter.com/browser-market-share#monthly-202001-202012.

507. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.* Nucleic acids research, 2000. **28**(1): p. 235-42.

508. Community, B.O., *Blender - a 3D modelling and rendering package*. Blender Foundation: Blender Institute, Amsterdam.

509. Magalhaes, R.P., T.F. Vieira, H.S. Fernandes, A. Melo, M. Simoes, and S.F. Sousa, *The Biofilms Structural Database.* Trends Biotechnol, 2020. **38**(9): p. 937-940.

510. Flemming, H.C. and J. Wingender, *The biofilm matrix.* Nat Rev Microbiol, 2010. **8**(9): p. 623-33.

511. Jamal, M., W. Ahmad, S. Andleeb, F. Jalil, M. Imran, M.A. Nawaz, T. Hussain, M. Ali, M. Rafiq, and M.A. Kamil, *Bacterial biofilm and associated infections.* J Chin Med Assoc, 2018. **81**(1): p. 7-11.

512. Simões, M., L.C. Simões, and M.J. Vieira, *A review of current and emergent biofilm control strategies.* LWT - Food Science and Technology, 2010. **43**(4): p. 573-583.

513. Roy, R., M. Tiwari, G. Donelli, and V. Tiwari, *Strategies for combating bacterial biofilms: A focus on anti-biofilm agents and their mechanisms of action.* Virulence, 2018. **9**(1): p. 522-554.

514. Koo, H., R.N. Allan, R.P. Howlin, P. Stoodley, and L. Hall-Stoodley, *Targeting microbial biofilms: current and prospective therapeutic strategies.* Nat Rev Microbiol, 2017. **15**(12): p. 740-755.

515. Kamaruzzaman, N.F., L.P. Tan, K.A. Mat Yazid, S.I. Saeed, R.H. Hamdan, S.S. Choong, W.K. Wong, A. Chivu, and A.J. Gibson, *Targeting the Bacterial Protective Armour; Challenges and Novel Strategies in the Treatment of Microbial Biofilm.* Materials (Basel), 2018. **11**(9).

516. Guilhen, C., C. Forestier, and D. Balestrino, *Biofilm dispersal: multiple elaborate strategies for dissemination of bacteria with unique properties.* Mol Microbiol, 2017. **105**(2): p. 188-210.

517. Gaulton, A., A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magarinos, J.P. Overington, G. Papadatos, I. Smit, and A.R. Leach, *The ChEMBL database in 2017.* Nucleic acids research, 2017. **45**(D1): p. D945-D954.

518. Liu, T., Y. Lin, X. Wen, R.N. Jorissen, and M.K. Gilson, *BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities.* Nucleic acids research, 2007. **35**(Database issue): p. D198-201.

519. Artimo, P., M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, S. Duvaud, V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios, and H. Stockinger, *ExPASy: SIB bioinformatics resource portal.* Nucleic acids research, 2012. **40**(Web Server issue): p. W597-603.

520. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic acids research, 2000. **28**(1): p. 27-30.

521. Bateman, A. *Uniprot: a universal hub of protein knowledge*. in *Protein Science*. 2019. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.

522. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.

This page was intentionally left blank.

F. Appendix

This page was intentionally left blank.

1. Posters

This page was intentionally left blank.

1.1. Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase

Encontro De Jovens Investigadores Em Biologia Computacional Estrutural, 2015, Coimbra, Portugal

U.PORTO
FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase

Henrique S. Fernandes[1‡], Nuno M. F. S. A. Cerqueira[1]

[1] REQUIMTE/UCIBIO, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
[‡] henriquefer11@gmail.com

UCIBIO @ REQUIMTE

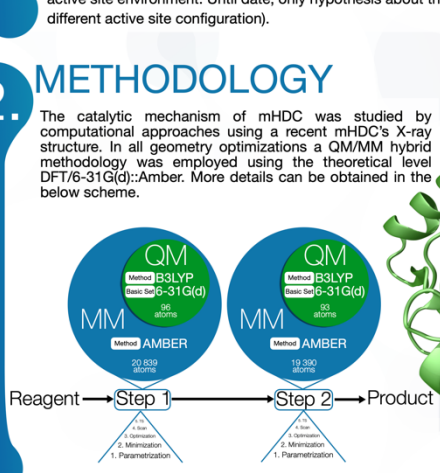
1. INTRODUCTION

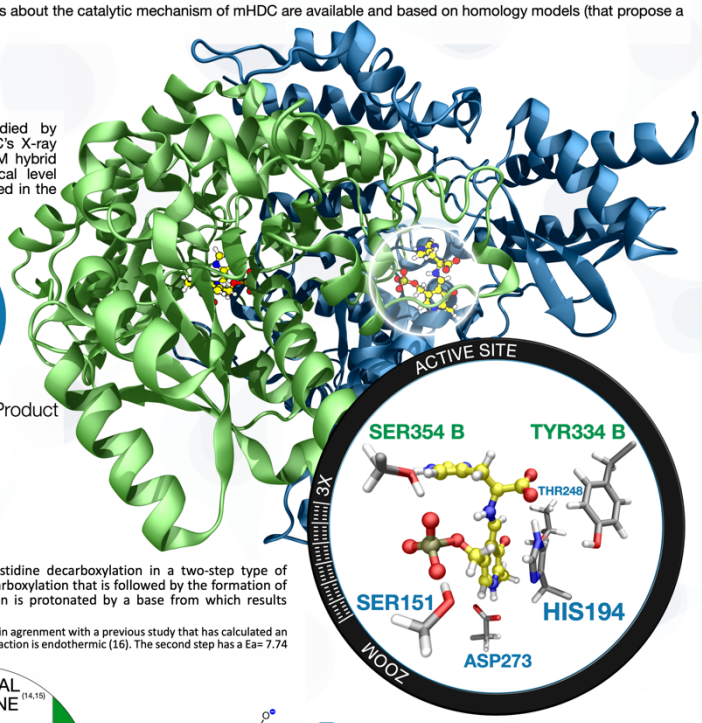
Mammalian histidine decarboxylase (mHDC) is an enzyme that requires pyridoxal-5'-phosphate (PLP) as a cofactor (1). mHDC belongs to group II of PLP-dependent decarboxylases (together with *L*-DOPA and glutamate decarboxylases) and catalyses the decarboxylation of *L*-histidine to histamine (2). Histamine plays a key role in several biological events such as immune response (3), gastric system modulation (4,5) and as a neurotransmitter in the nervous system (6). Several inhibitors for histamine action have been studied in order to treat some diseases such as atopic dermatitis (7), allergies (8), and cancer (9-12).

mHDC has been studied for a long time, but only in 2012 Komori's (13) group was able to determine the X-ray structure of the enzyme and revealed the active site environment. Until date, only hypothesis about the catalytic mechanism of mHDC are available and based on homology models (that propose a different active site configuration).

2. METHODOLOGY

The catalytic mechanism of mHDC was studied by computational approaches using a recent mHDC's X-ray structure. In all geometry optimizations a QM/MM hybrid methodology was employed using the theoretical level DFT/6-31G(d)::Amber. More details can be obtained in the below scheme.





3. RESULTS

The results have shown that mHDC catalysis the histidine decarboxylation in a two-step type of mechanism. The **first step** involves a histidine (R) decarboxylation that is followed by the formation of a carbanion. Then, in the **second step**, the carbanion is protonated by a base from which results histamine (P).

Our early results indicate that the first step is the limiting one. This is in agrenment with a previous study that has calculated an activation energy of 22.6 kcal/mol for this step and showed that the reaction is endothermic (16). The second step has a Ea= 7.74 kcal/mol and it is exothermic by 26.8 kcal/mol.

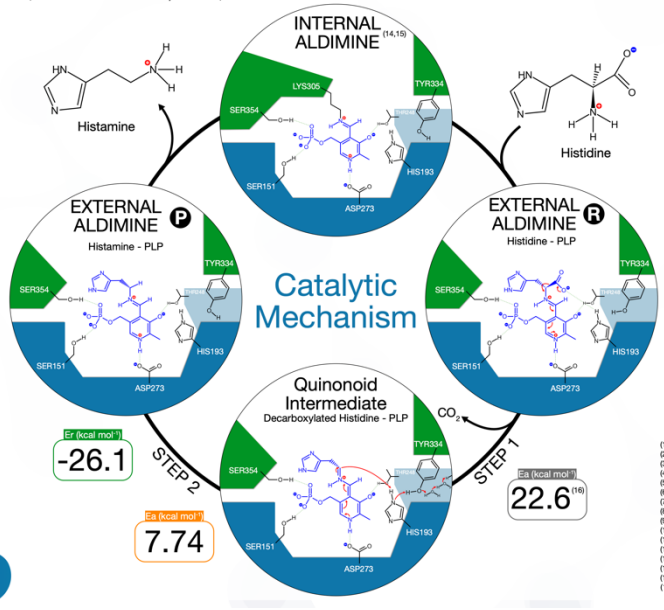


FIG. 1 Structure of the *Homo Sapiens L*-Histidine Decarboxylase dimer (NewCartoon representation; blue subunit A; green subunit B). Natural substrate (*L*-Histidine) linked to the PLP co-factor (CPK representation). Highlight of the active site; Histidine linked to PLP (CPK); active site residues that should be involved in catalysis (Licorice representation). (PDB code: 4E1O adapted) (13)

4. FUTURE

Determination of the activation and reaction energies for the first step.

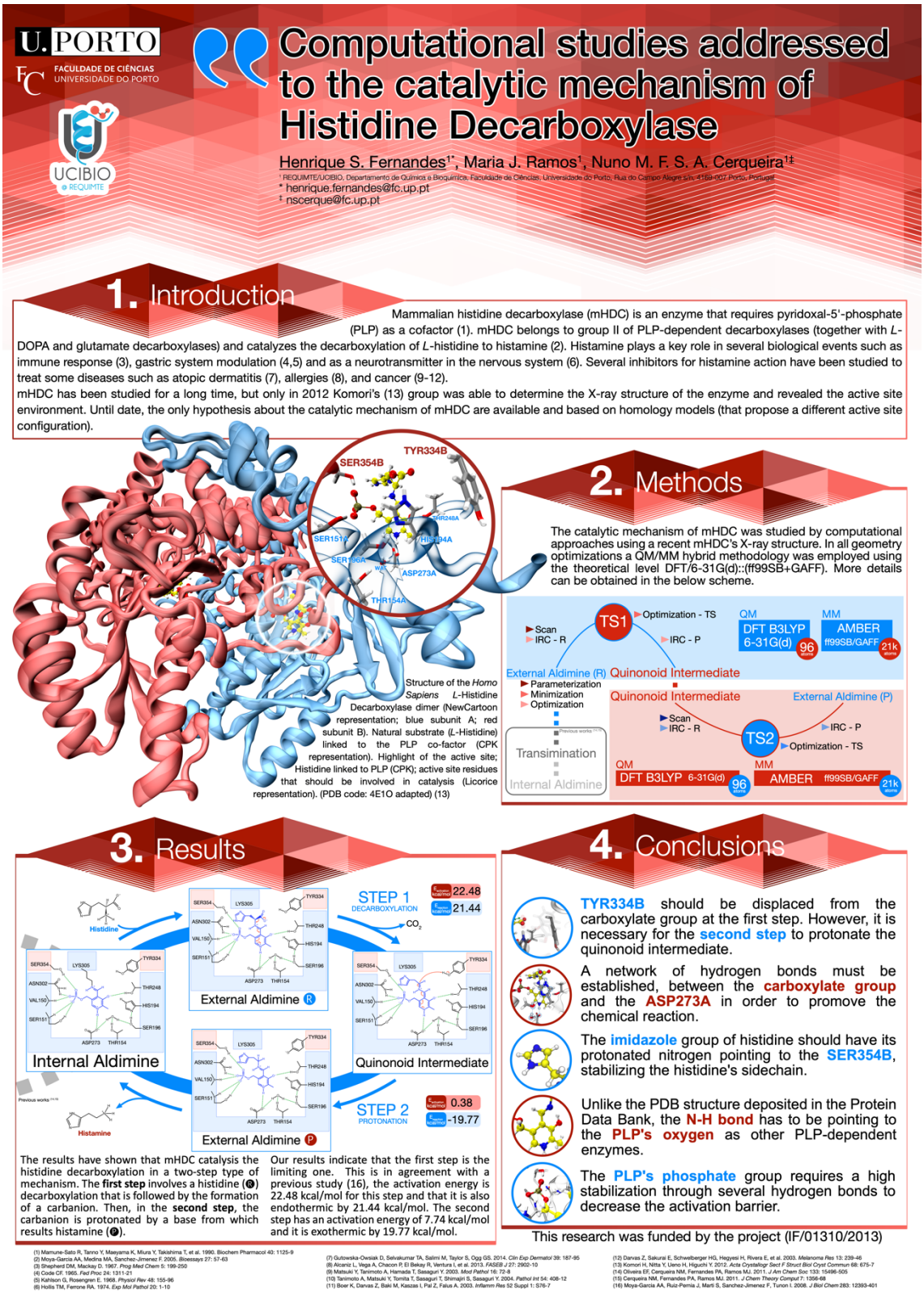
Improve the level of theory of the final energies through single-point energy calculations.

Computational Alanine Scanning Mutagenesis (CompASM) to search for relevant protein-protein interactions between the two subunits of the mHDC.

(1) Mamune-Sato R, Tanno Y, Maeyama K, Miura Y, Takishima T, et al. 1990. Biochem Pharmacol 40: 1125-9
(2) Moya-Garcia AA, Medina MA, Sanchez-Jimenez F. 2005. Biosessays 27: 57-63
(3) Shepherd DM, Mackay D. 1967. *Prog Med Chem* 5: 199-250
(4) Code CF. 1965. *Fed Proc* 24: 1311-21
(5) Kahlson G, Rosengren E. 1968. *Physiol Rev* 48: 155-96
(6) Hollis TM, Ferrone RA. 1974. *Exp Mol Pathol* 20: 1-10
(7) Gutowska-Owsiak D, Selvakumar TA, Salimi M, Taylor S, Ogg GS. 2014. *Clin Exp Dermatol* 39: 187-95
(8) Alcaniz L, Vega A, Chacon P, El Bekay R, Ventura I, *et al.* 2013. *FASEB J* 27: 2902-10
(9) Matsuki Y, Tanimoto A, Hamada T, Sasaguri Y. 2003. *Mod Pathol* 16: 72-8
(10) Tanimoto A, Matsuki Y, Tomita T, Sasaguri T, Shimajiri S, Sasaguri Y. 2004. *Pathol Int* 54: 408-12
(11) Boer K, Darvas Z, Baki M, Kaszas I, Pal Z, Falus A. 2003. *Inflamm Res* 52 Suppl 1: S76-7
(12) Darvas Z, Sakurai E, Schwelberger HG, Hegyesi H, Rivera E, et al. 2003. *Melanoma Res* 13: 239-46
(13) Komori H, Nitta Y, Ueno H, Higuchi Y. 2012. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68: 675-7
(14) Oliveira EF, Cerqueira NM, Fernandes PA, Ramos MJ. 2011, *J Am Chem Soc* 133: 15496-505
(15) Cerqueira NM, Fernandes PA, Ramos MJ. 2011. *J Chem Theory Comput* 7: 1356-68
(16) Moya-Garcia AA, Ruiz-Pernia J, Marti S, Sanchez-Jimenez F, Tunon I. 2008. *J Biol Chem* 283: 12393-401

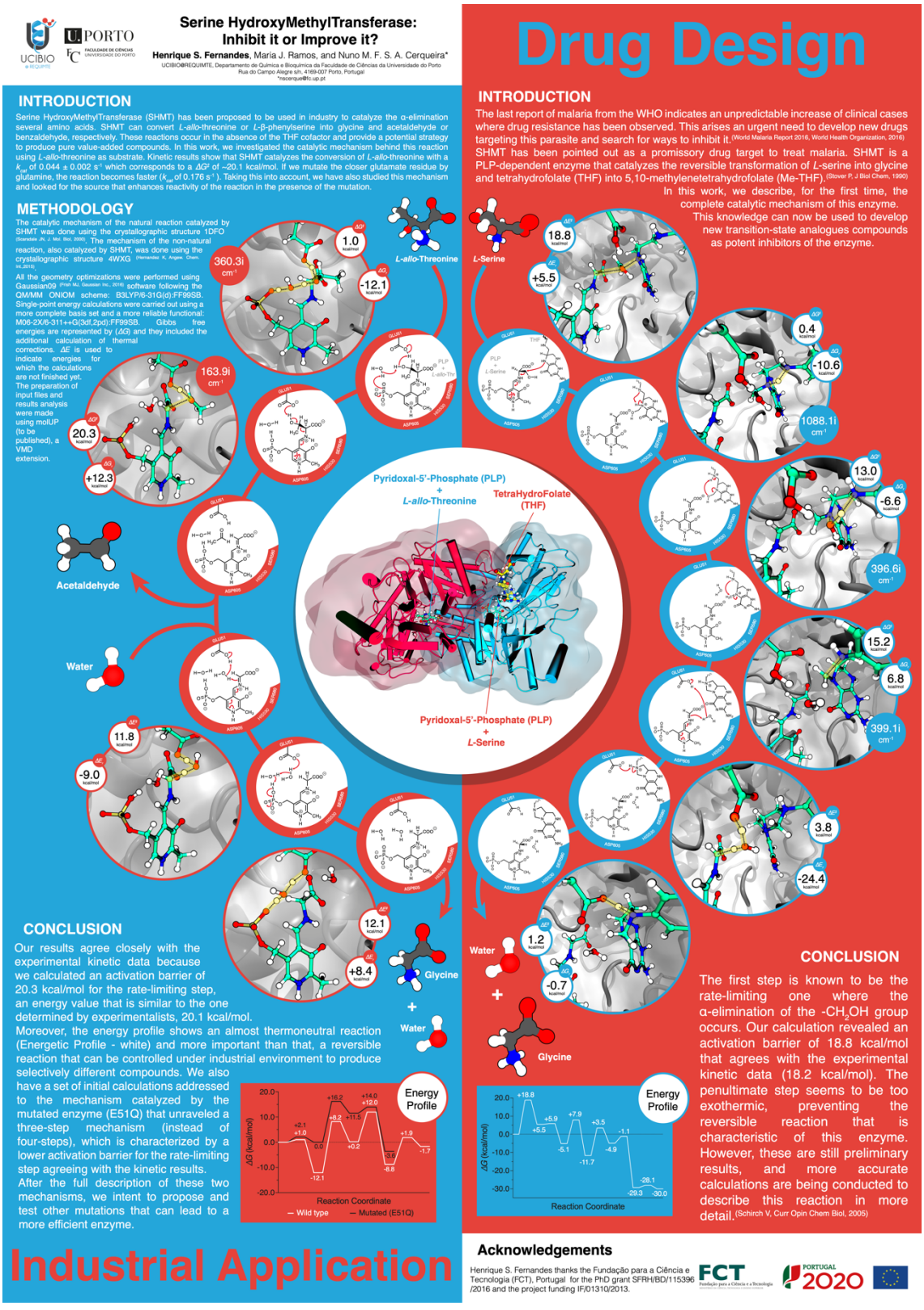
1.2. Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase

1st European Young Chemists Meeting and 5º Encontro Português de Jovens Químicos, 2016, Guimarães, Portugal



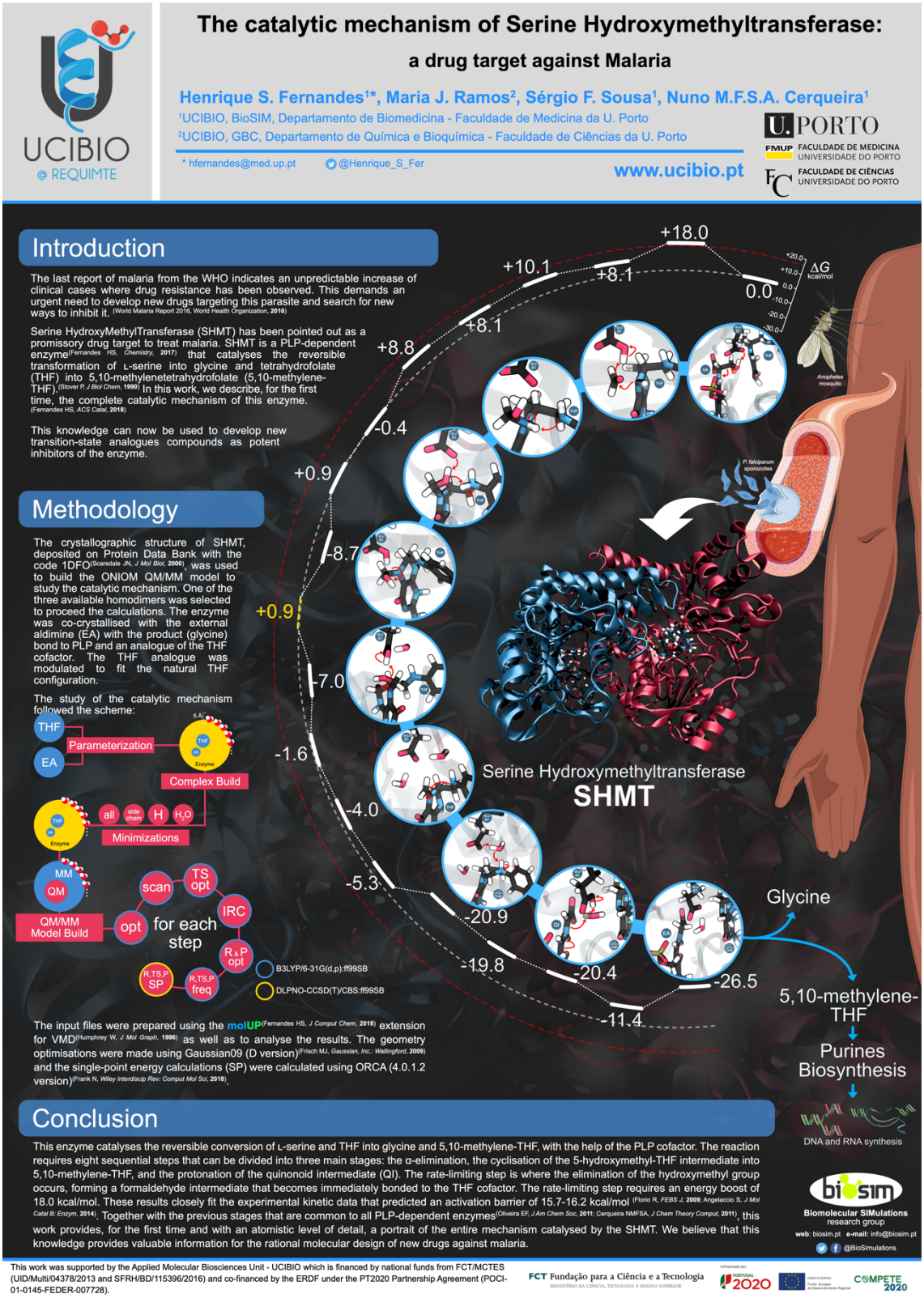
1.3. Serine HydroxyMethylTransferase: Inhibit it or Improve it?

III Symposium on Medicinal Chemistry of University of Minho, 2017, Braga, Portugal



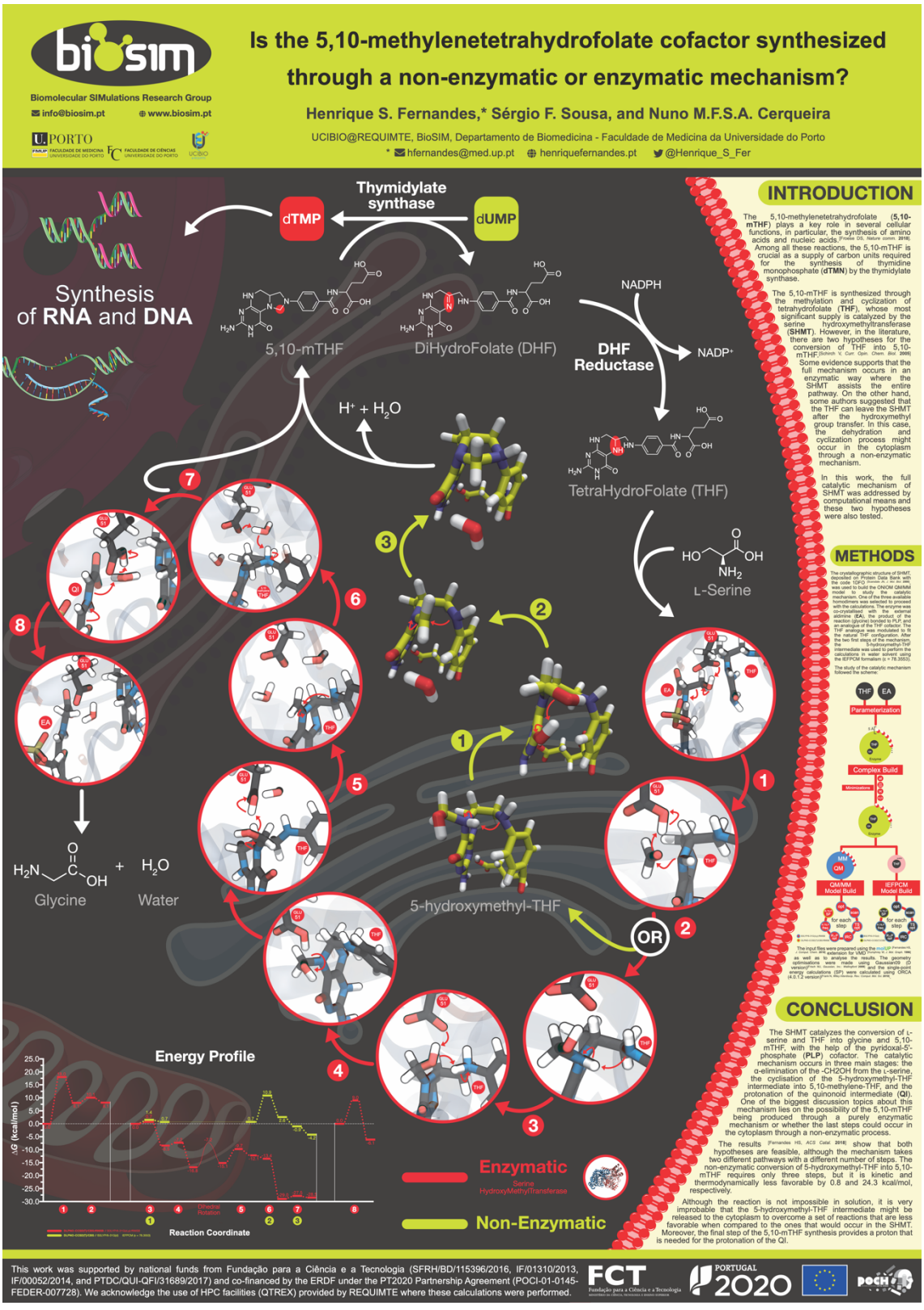
1.4. The catalytic mechanism of Serine Hydroxymethyltransferase: a drug target against Malaria

UCIBIO Annual Meeting, 2018, Lisboa, Portugal



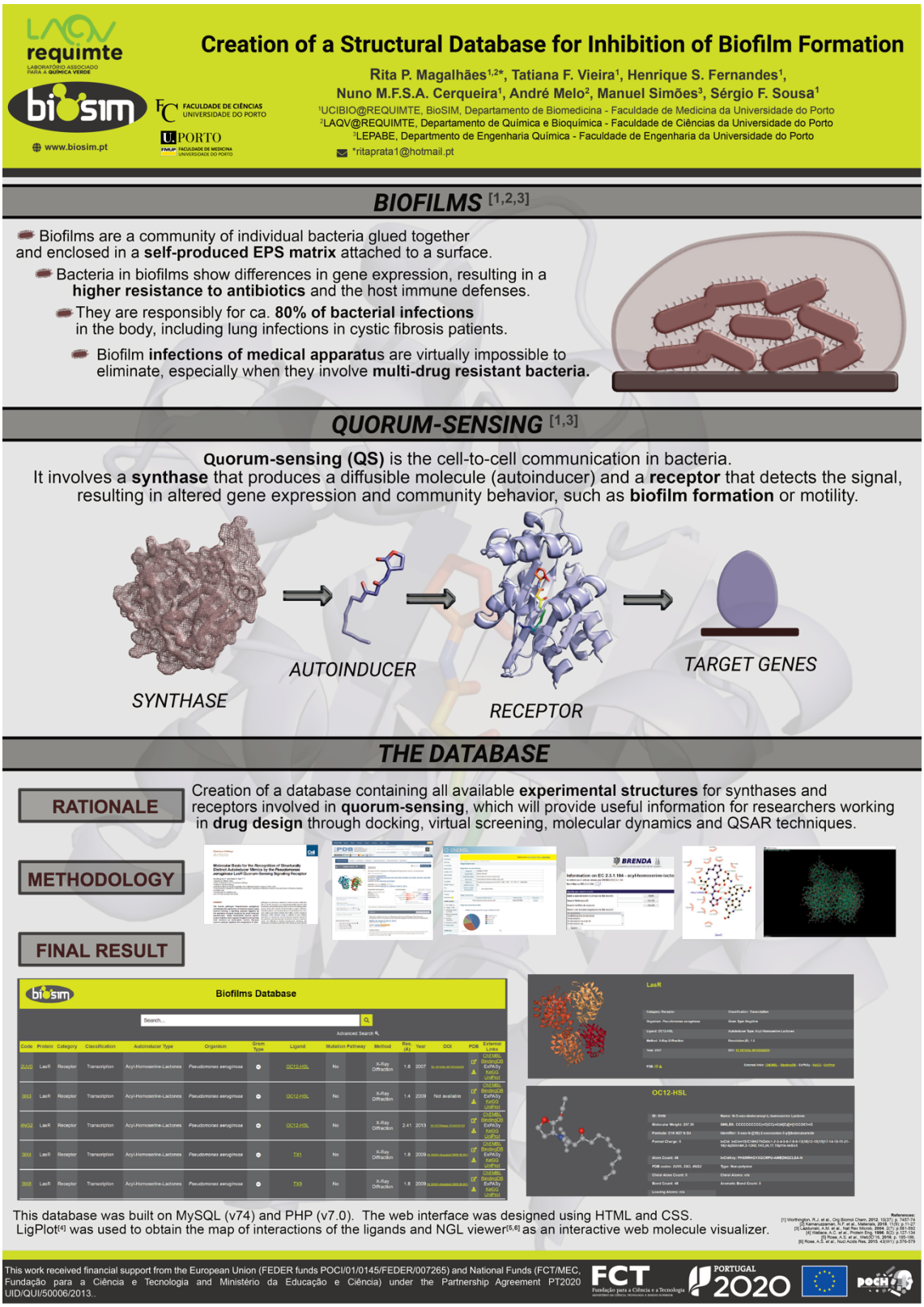
1.5. Is the 5,10-methylenetetrahydrofolate cofactor synthesized through a non-enzymatic or enzymatic mechanism?

6º Encontro de Jovens Investigadores de Biologia Computacional Estrutural, 2018, Porto, Portugal



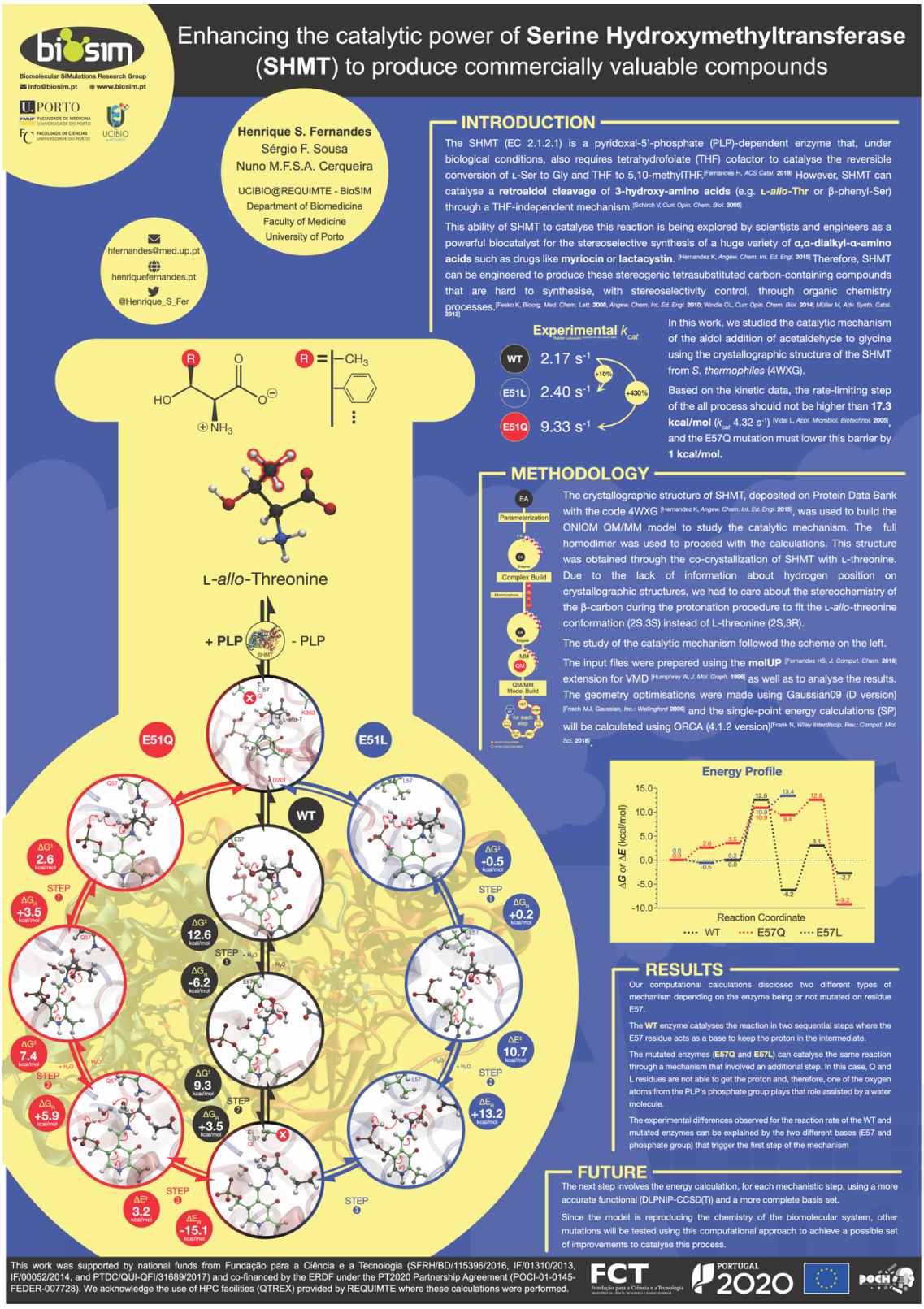
1.6. Creation of a Structural Database for Inhibition of Biofilm Formation

6º Encontro De Jovens Investigadores Em Biologia Computacional Estrutural, 2018, Porto, Portugal



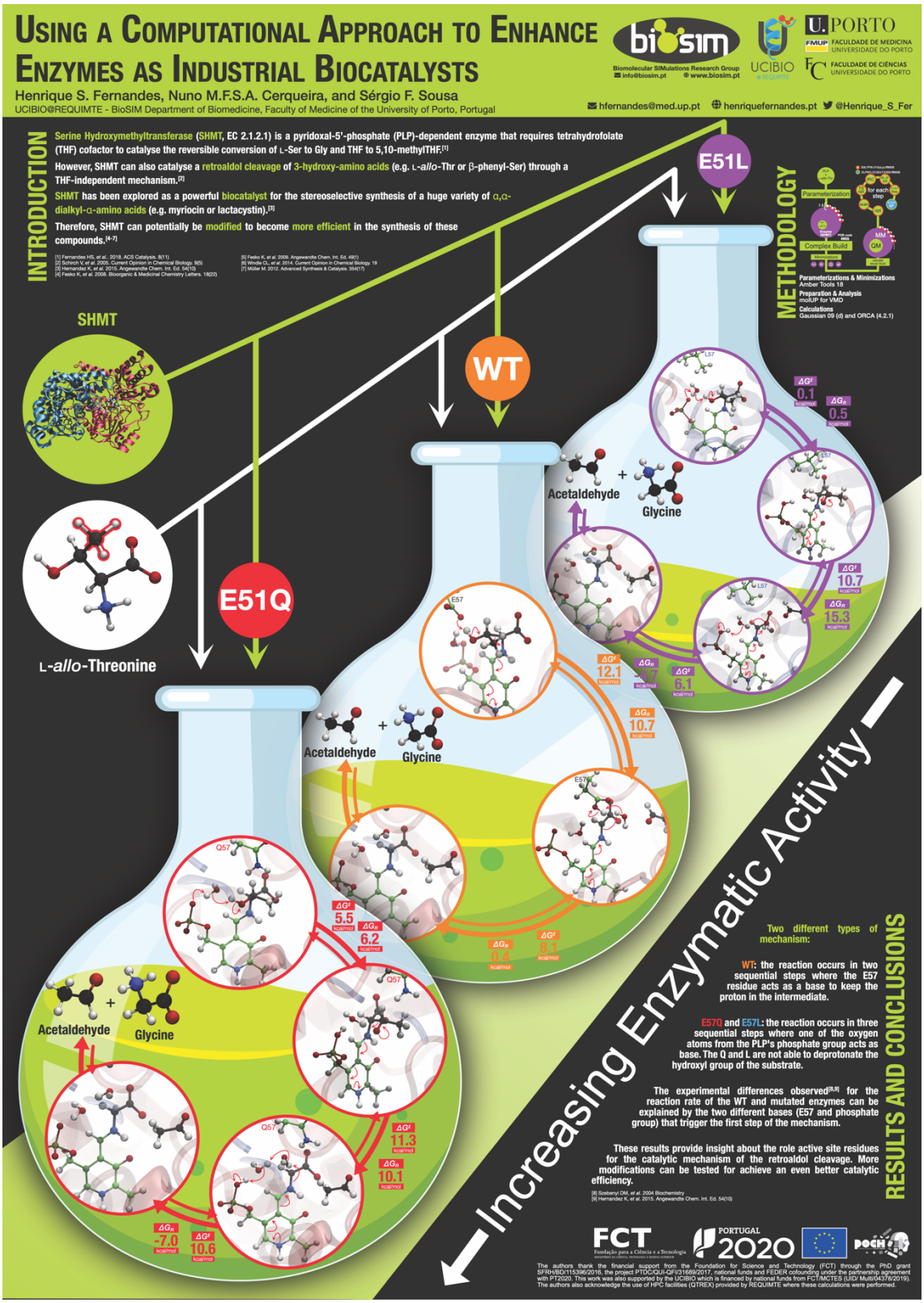
1.7. Enhancing the catalytic power of SHMT to produce commercially valuable compounds

19th FEBS Young Scientists' Forum and *The 44th FEBS Congress*, 2019, Krakow's Jagiellonian University Campus, Krakow, Poland



1.8. Using a Computational Approach to Enhance Enzymes as Industrial Biocatalysts

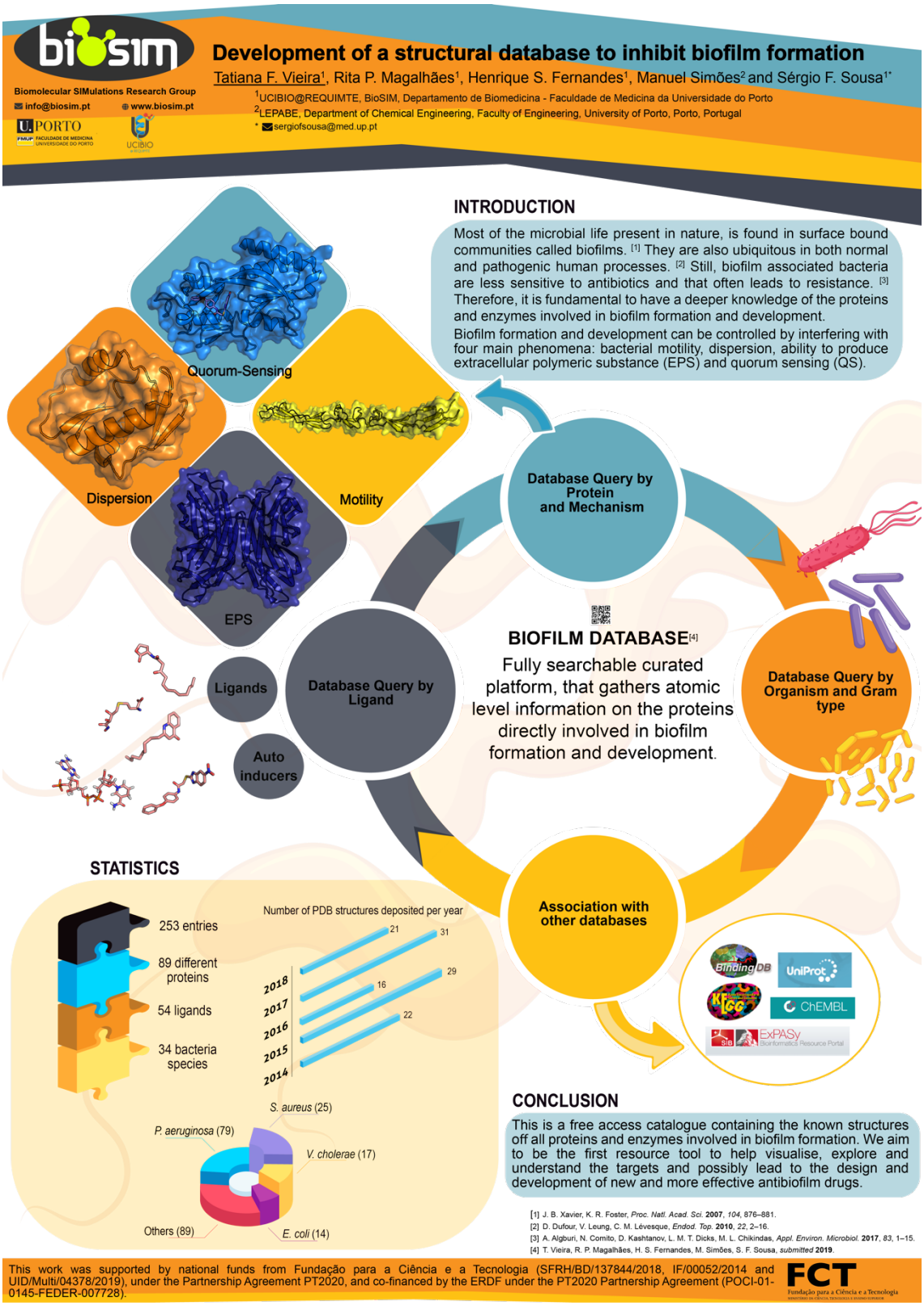
7º Encontro De Jovens Investigadores Em Biologia Computacional Estrutural 2019, Lisbon, Portugal



1.9. Development of a structural database to inhibit biofilm formation

7º Encontro De Jovens Investigadores Em Biologia Computacional Estrutural

2019, Lisbon, Portugal



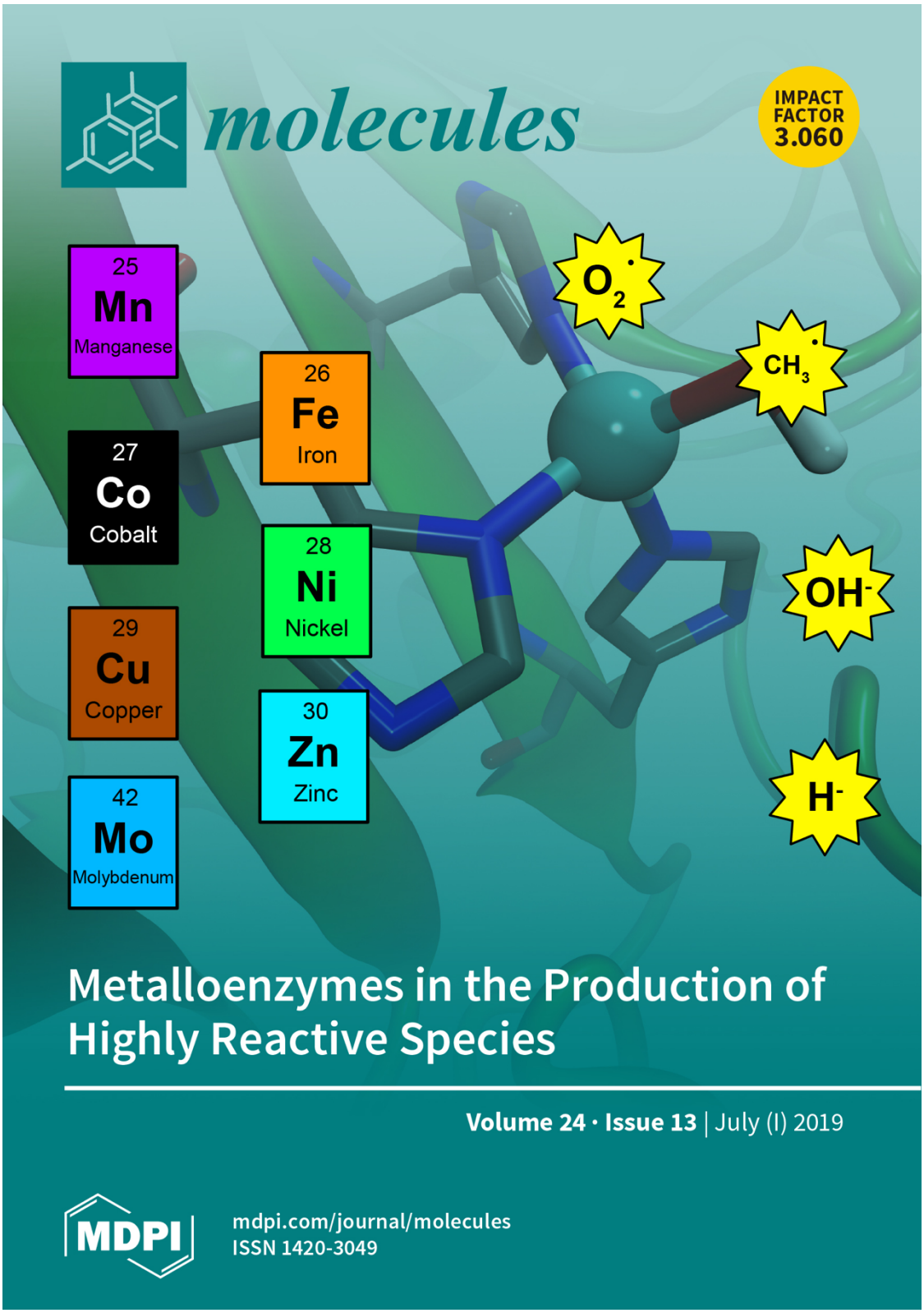
This page was intentionally left blank.

2. Covers

This page was intentionally left blank.

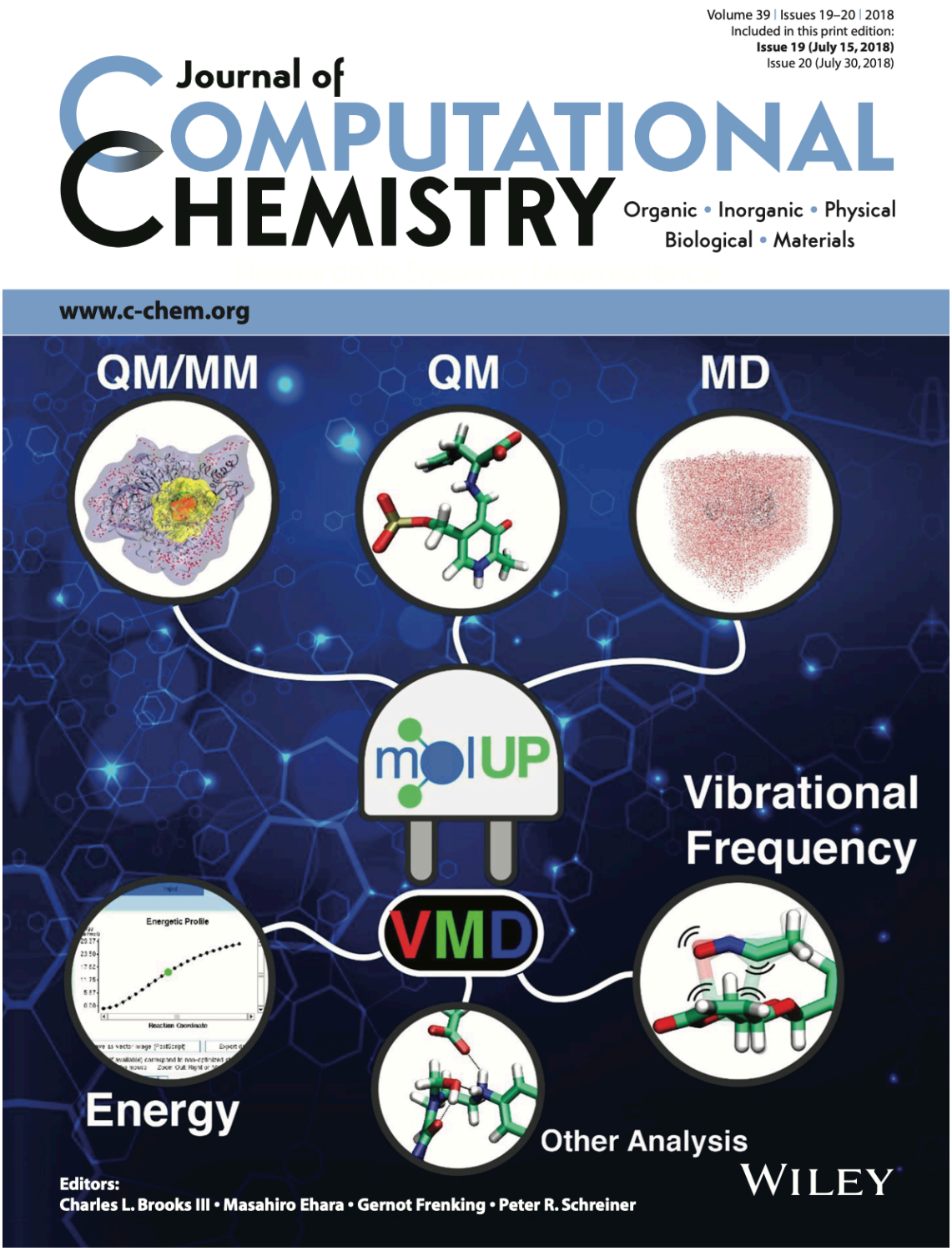
2.1. Cover of the Molecules

Volume 24, Issue 13, 2019



2.2. Cover of the Journal of Computational Chemistry

Volume 39, Issues 19-20, 2018



3. Reviews and Book Chapters

This page was intentionally left blank.

molecules

MDPI

Review

Formation of Unstable and very Reactive Chemical Species Catalyzed by Metalloenzymes: A Mechanistic Overview

Henrique S. Fernandes, Carla S. Silva Teixeira, Sérgio F. Sousa and Nuno M. F. S. A. Cerqueira *

UCIBIO@REQUIMTE, BioSIM, Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Alameda Professor Hernâni Monteiro, 4200-319 Porto, Portugal

* Correspondence: nunoscerqueira@med.up.pt

Academic Editors: Eugene Permyakov and Sergey Permyakov

Received: 10 May 2019; Accepted: 3 July 2019; Published: 4 July 2019

Abstract: Nature has tailored a wide range of metalloenzymes that play a vast array of functions in all living organisms and from which their survival and evolution depends on. These enzymes catalyze some of the most important biological processes in nature, such as photosynthesis, respiration, water oxidation, molecular oxygen reduction, and nitrogen fixation. They are also among the most proficient catalysts in terms of their activity, selectivity, and ability to operate at mild conditions of temperature, pH, and pressure. In the absence of these enzymes, these reactions would proceed very slowly, if at all, suggesting that these enzymes made the way for the emergence of life as we know today. In this review, the structure and catalytic mechanism of a selection of diverse metalloenzymes that are involved in the production of highly reactive and unstable species, such as hydroxide anions, hydrides, radical species, and superoxide molecules are analyzed. The formation of such reaction intermediates is very difficult to occur under biological conditions and only a rationalized selection of a particular metal ion, coordinated to a very specific group of ligands, and immersed in specific proteins allows these reactions to proceed. Interestingly, different metal coordination spheres can be used to produce the same reactive and unstable species, although through a different chemistry. A selection of hand-picked examples of different metalloenzymes illustrating this diversity is provided and the participation of different metal ions in similar reactions (but involving different mechanism) is discussed.

Keywords: metalloenzymes; transition metals; catalytic mechanism; iron; zinc; molybdenum

1. Introduction

Nature has tailored a wide range of proteins that play a vast array of functions in all living organisms and from which their survival and evolution depends on [1]. Some of these proteins require the presence of metal cofactors that can be as small as a metal ion or as large as a heterocyclic organometallic complex [2–5]. These proteins are called metalloproteins and they are estimated to account for almost one-third of the full proteomic world [6,7].

Metalloproteins can play many different roles in nature, such as catalytic mediators, transport, storage, and signal transduction [6–10]. Among the most important ones are those with enzymatic activity, commonly named metalloenzymes [11–14]. These enzymes catalyze some of the most important biological processes in nature, such as photosynthesis [15,16], respiration [17,18], water oxidation [19], molecular oxygen reduction [20,21], and nitrogen fixation [22]. They are also among the most proficient catalysts in terms of their activity, selectivity, and ability to operate at mild conditions of temperature,

pH, and pressure [23]. In the absence of these enzymes, these reactions would proceed very slowly, if at all, suggesting that these enzymes made the way for the emergence of life as we know today.

Over the years, a wealth of knowledge on the metal-dependent enzymes has been accumulated, including data from X-ray crystallography, kinetic and biochemical properties, and reaction mechanisms. These studies revealed a variety of metal cofactors [24] that catalyze very specialized chemical reactions [25]. The majority of these cofactors contain first-row transition-metal ions, such as manganese [26], iron [27], cobalt [28], copper [29], or zinc [30]; the alkaline earth metal magnesium [31] or even second-row and third-row transition metal ions, such as molybdenum [32] and tungsten [32,33], respectively. Most of these metal ions are usually coordinated to the proteins by imidazole substituents in histidine residues [34], thiolate substituents in cysteine residues [35], and carboxylate groups provided by aspartate or glutamate residues. Other metal ions can also coordinate to organic cofactors, such as the tetradentate N4 macrocyclic ligands present in hemes, pyranopterin, and among others. The nature and the charge of the ligands in the coordination sphere of the metal ion dictates its coordination numbers and/or geometries and ultimately defines its role in the biological systems. For example, in the case of Zn^{2+}, the ion can play either a catalytic, co-catalytic, or structural role in the enzyme [36,37].

The role played by the protein in catalysis of the metalloenzymes is often disregarded. However, this is far from being correct, since proteins also have an important role in the reactivity of the metal cofactor [23]. For example, they can hamper a symmetrical binding of the substrate to a metal ion favoring by this way a stereoselective reaction; they can prevent the direct interaction between a substrate and the metal ion, through the coordination of several amino acid residues to the metal ion, etc. All of these facts together indicate that the chemistry played by metalloenzymes is not exclusively dependent on the metal ion, but instead there is a reciprocal interaction between the metal and the protein.

The chemistry catalyzed by metalloenzymes is very specific from enzyme to enzyme and commonly involves the formation and/or sequestration of extremely unstable chemical species that are ultimately transferred to a given substrate [9,38]. In the majority of the cases, once the product of the reaction is released to the solvent, the metal cofactor needs to be recycled back to its original reactive form and only afterwards the enzymatic turnover takes place.

In this review, the catalytic mechanism of a diverse selection of metalloenzymes will be reviewed. All the included examples describe several enzymes in which the cofactors contain one or more metal ions that can play a catalytic or co-catalytic function and involve the formation of unstable and very reactive reaction intermediates such as hydroxide anions (OH^-), hydrides (H^-), radicals, and superoxide molecules.

2. Enzymes that Catalyze the Formation of Hydroxide Anions

Hydroxide is a diatomic anion, which consists of an oxygen and a hydrogen atom held together by a covalent bond that carries a negative charge [39]. It is a strong nucleophile that is employed by a variety of metalloenzymes to catalyze hydrolysis or hydration reactions. The formation of such a highly reactive and unstable anion is not a straightforward process (especially under physiological conditions) and can only be achieved by a reduced number of enzymes that requires the presence of special metal cofactors. These often include zinc in their structure, but occasionally, other transition metals, such as manganese and nickel, can also be found.

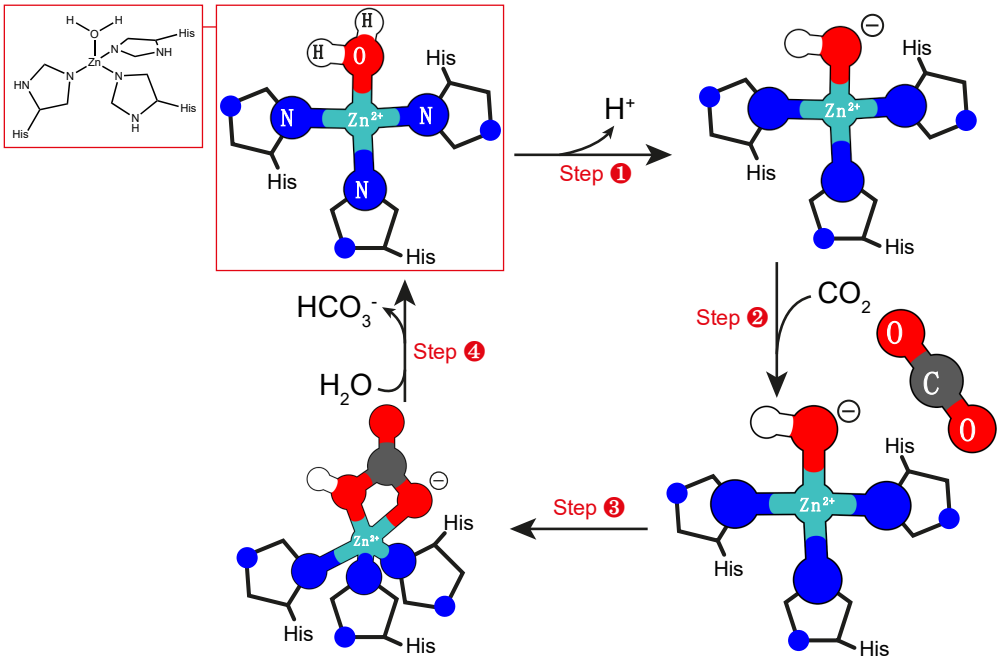
In these reactions, the metal cofactor always acts as an electron pair acceptor (as a metal-based Lewis acid) that stabilizes the highly reactive hydroxide anion, thereby ensuring that an activated nucleophile is available for catalysis at physiological pH. Although these metalloenzymes seem to share a similar function in nature, the chemistry that they catalyze is different. In this section, we compare different metal coordination spheres present on enzymes that catalyze the formation of hydroxide anions, their particularities, and the mechanism that allows these enzymes with such different metals and coordination spheres to carry out similar reactions, while employing a diverse chemistry.

2.1. Mono-nuclear Zinc Cofactor

α-Carbonic anhydrase (CA, EC 4.2.1.1) belongs to a family of enzymes that catalyze the inter-conversion of carbon dioxide (CO_2) to carbonic acid (H_2CO_3) though the formation of a hydroxide ion (bound to zinc) at neutral pH [40–42]. In this particular case, a tetracoordinate Zn ion is responsible for the formation of hydroxide intermediate that triggers the conversion of carbon dioxide into carbonic acid.

The mechanism starts with a water molecule bonded directly to the Zn(II) cofactor that is coordinated to three histidine residues [43]. This complex polarizes the bonds between the oxygen and the two hydrogen atoms from the water molecule, making these chemical bonds weaker, and the oxygen atom slightly more positive. This effect favors the proton transfer from the water molecule to a neighboring histidine residue originating from a hydroxide anion that remains coordinated to the Zn(II) ion (Scheme 1—Step ❶). When carbon dioxide becomes available in the active site (Scheme 1—Step ❷), the nucleophilic oxygen from the hydroxide ion attacks the carbon from carbon dioxide, at the same time that one of the oxygen atoms from the same molecule binds directly to the Zn(II) cofactor (Scheme 1—Step ❸). These chemical reactions induce a conformational rearrangement on the metal cofactor that goes from a distorted tetrahedral conformation to a trigonal bipyramidal configuration. This new rearrangement turns carbon dioxide very close to the hydroxide ion (both of them coordinated directly to the Zn(II) ion) favoring by this way the reaction between both ligands. At the end of the reaction, a water molecule becomes coordinated to the Zn ion forming a pentacoordinate complex that triggers the release of the bicarbonate molecule, making the enzyme ready for a new catalytic cycle (Scheme 1—Step ❹) [41,44–49].

CA has been the focus of many experimental [50–60] and theoretical [61–77] investigations, many of them devoted to the study of its catalytic mechanism. Those studies show values of k_{cat} in the range of 10^4 to 10^6 s^{-1} [78] and suggests that the rate-limiting step of the reaction mechanism is the deprotonation of metal bound water molecules [79], although this hypothesis is not consensual [71].



Scheme 1. General description of the catalytic mechanism of α-Carbonic anhydrase (CA). Only the atoms involved in the first coordination shell of the metal ion were included to simplify the representation.

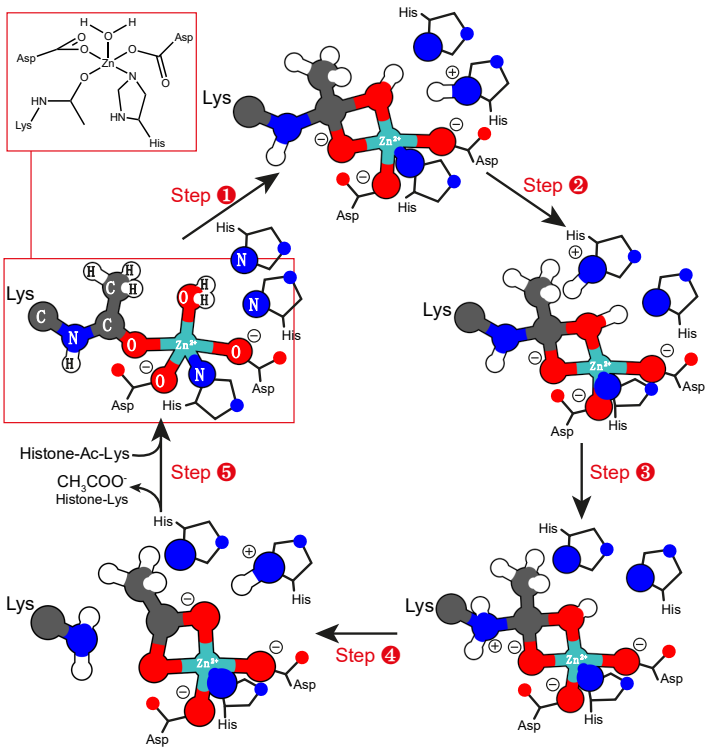
Histone Deacetylase 8 (HDAC8, EC 3.5.1.98) catalyzes the deacetylation of lysine residues on the N-terminal part of the core histones [80,81]. Similar to CA, the reaction catalyzed by HDAC8 involves a zinc cofactor that reacts with a water molecule and from which results a hydroxide anion. This reaction

intermediate is a sturdy nucleophile that attacks the substrate directly from the metal ion. In spite of the similarities between the mechanisms of HDAC8 and CA, these enzymes have different amino acid residues coordinated to the Zn(II) ion and follow different mechanisms.

In CA, the Zn(II) ion is coordinated with three histidine residues, whereas in HDAC8, the Zn(II) ion is coordinated to one histidine and two aspartate residues.

In HDAC8, the catalytic mechanism begins with the binding of a carbonyl group of an acetylated lysine to the Zn(II) ion that is tetra-coordinated with three amino acid residues and a water molecule. At the end of this process, a penta-coordinated complex is obtained. Only afterwards, the formation of the hydroxide anion takes place. This reaction occurs by the abstraction of a proton from the water molecule by a neighboring histidine residue, which is concomitant with the formation of a covalent bond between the oxygen from the hydroxide anion and the sp2 carbon from the acetyl group (Scheme 2—Step ❶). The next step involves a proton exchange between the two active site histidine residues, through the oxygen from the hydroxide anion (Scheme 2—Step ❷). The protonated histidine then donates a proton to the nitrogen atom of the substrate (Scheme 2—Step ❸). The last step of the mechanism involves the formation of an acetyl group and the release of the lysine residue from histones (Scheme 2—Step ❹). The enzymatic turnover is obtained when the Zn(II) cofactor reacts again with a new water molecule, making the enzyme ready for a new cycle (Scheme 2—Step ❺) [82,83].

Computational studies have shown that the limiting step of the catalytic mechanism is the first one with a calculated overall activation free energy barrier of 18.3 ± 0.4 kcal/mol [84], which is consistent with the experimental value of 17.7 kcal/mol that can be estimated from the experimental k_{cat} value of 0.90 ± 0.03 s^{-1} for HDAC8 [85].



Scheme 2. General description of the catalytic mechanism of Histone Deacetylase 8 (HDAC8). Only the atoms involved in the first coordination shell of the metal ion were represented to simplify the representation.

The differences that are observed between the catalytic mechanisms of CA and HDAC8 are considered to be determined by the nucleophilic strength of the hydroxide anion that is bonded to the Zn(II) cofactor. This is modulated by the capacity of the Zn(II) ion to behave as an electron pair

acceptor (metal-based Lewis acid) that, in the case of these two enzymes, is enhanced by the enzyme through the coordination of different amino acid residues to the zinc ion. HDAC8 has two aspartate residues and one histidine residue coordinated to Zn(II) instead of the three histidine residues that are bonded to Zn(II) in CA. This means that HDAC8 has a higher capacity to shelter the negative charge from the hydroxide anion than CA, turning by this way its hydroxide anion into a worse nucleophile [86–88]. Although this would suggest that the hydroxide anion in HDAC8 would be less reactive and therefore the enzyme less efficient than CA, this is not observed. In fact, this feature seems to be endorsed by the enzyme due to the different substrates that both enzymes catalyze. In the case of HDAC8, the protonation of the substrate is required, which is only possible because the proton from the hydroxide anion can be easily interchanged with the active site histidine residues. This step would not be feasible if the hydroxide anion was very reactive, as it occurs in the mechanism of the other enzyme. In the case of CA, this step is not required and, therefore, the performance of the enzyme is improved with a more reactive hydroxide anion.

2.2. Binuclear Manganese Cofactor

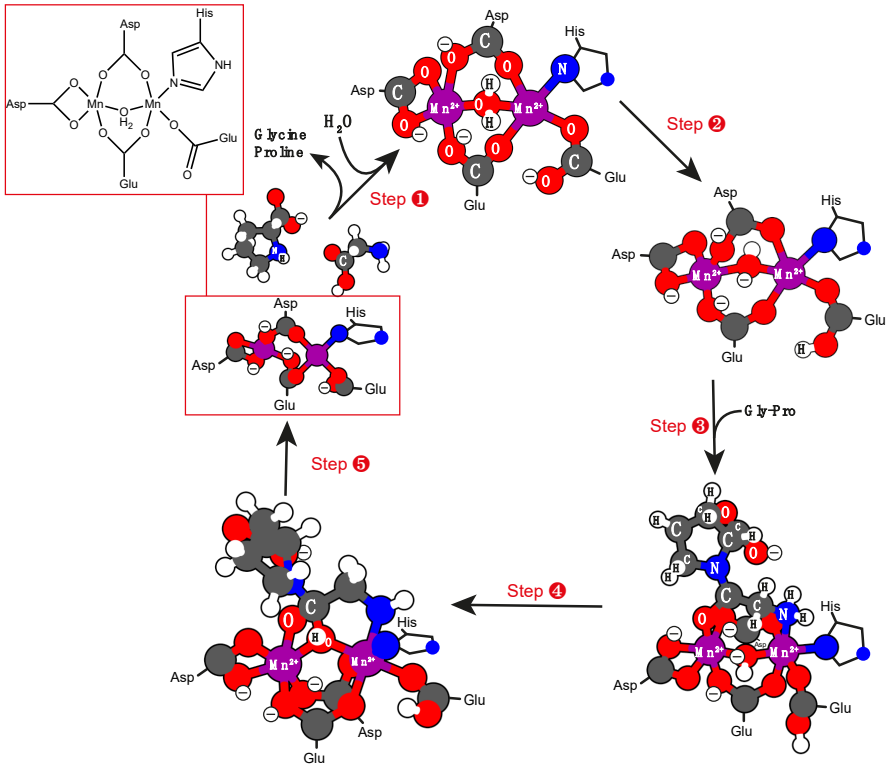
Prolidase or proline dipeptidase (EC 3.4.13.9) is a binuclear metallopeptidase that hydrolyzes dipeptides with proline or hydroxyproline at the carboxyl terminus [89,90]. These binuclear metalloenzymes hydrolyze mainly Gly-Pro dipeptides but they are also active against Ala-Pro, Met-Pro, Phe-Pro, Leu-Pro, and Val-Pro [89,91,92].

The X-ray structures of human prolidase indicates the presence of two catalytic Mn(II) ions in the active site that are hexa-coordinated to a histidine, two glutamate, two aspartate residues, and a water molecule [93,94]. Both Mn(II) ions are bridged between themselves by a hydroxide anion and two *syn*-bidentate carboxylate groups from the aspartate and glutamate residues. Due to this disposition, the Mn1 ion has a distorted octahedral configuration, whereas the Mn2 ion adopts a trigonal bipyramidal configuration.

The catalytic mechanism starts with the activation of the Mn1–Mn2 metal ions bridging one water molecule (Scheme 3—Step ❶). This is catalyzed by the carboxylate group of one of the neighbor glutamate residues, which abstracts a proton from the water molecule, and from which results a hydroxide anion that becomes coordinated to both metal ions (Scheme 3—Step ❷).

When the substrate is available in the active site (for example, Gly-Pro), it becomes coordinated to both Mn(II) ions and it is stabilized by a network of hydrogen bonds created by one neighbor arginine, two histidine residues, and two aspartate residues that are coordinated to both metal ions (Scheme 3—Step ❸). The formation of this complex endorses the polarization of the peptide bond of the substrate that favors the nucleophilic attack of the hydroxide anion to the partially positively charged (δ^+) carbonyl group of the substrate (Scheme 3—Step ❹). Once this step is completed, a tetrahedral reaction intermediate is obtained. The next step of the catalytic mechanism is proposed to occur in two simultaneous reactions: (i) The proton transfer from the hydroxide anion to the nitrogen atom enclosed on the proline ring with the help of a neighbor glutamate residue and (ii) the cleavage of the peptide bond of the substrate (in this case, Pro-Gly) (Scheme 3—Step ❺). [89] Once this step is finished, Gly and Pro leave the active site and a water molecule from the solvent binds to the bimetallic center. There, it is activated by a glutamate residue that assists a proton abstraction, regenerating the enzyme for a new catalytic cycle (Scheme 3—Step ❶) [93,95].

Experimental data show a k_{cat} of 683 s^{-1} for the human prolidase catalyzing the substrate Gly-Pro [96].



Scheme 3. General description of the catalytic mechanism of prolidase. Only the atoms involved in the first coordination shell of the metal ion were represented to simplify the representation.

2.3. Binuclear Nickel Cofactor

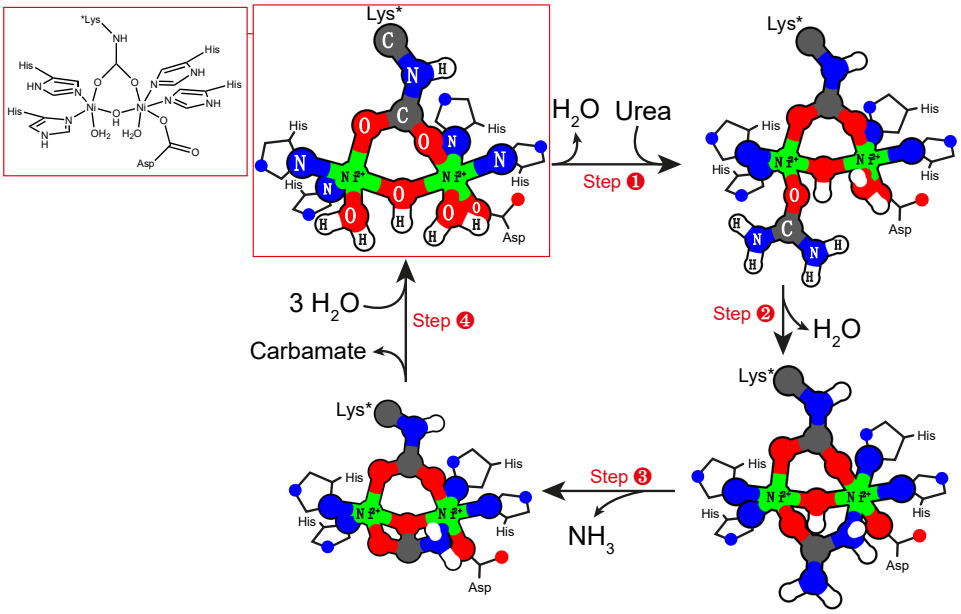
Urease (EC 3.5.1.5), in terms of function, belongs to a family amidohydrolases and phosphotriesterases [97] that employs two asymmetric Ni(II) metal ions in the active site to degrade urea into carbon dioxide and ammonia [98,99]. The active site also has an unusual carbamylated lysine (Lys*) residue that establishes a bridge between both catalytic Ni(II) ions.

One of the Ni(II) cations, Ni1, is penta-coordinated to an oxygen from the lysine residue, a hydroxide anion, a water molecule, and two histidine residues. The second Ni(II), Ni2, is hexa-coordinated to the other oxygen atom of the lysine, to the same hydroxide anion, a water molecule, two histidine residues, and an additional aspartate residue.

Similar to what is observed in all the other enzymes discussed in this section, the hydroxide anion is rapidly formed through the reaction of a water molecule with the metal cofactor that loses a proton to an active site residue.

Once the substrate, urea, is available in the active site, the water molecules coordinated to the Ni1 ion dissociate from the metal cofactor and urea becomes coordinated directly to the Ni1 ion through the oxygen atom (Scheme 4—Step ❶). Afterwards, the water molecule coordinated to Ni2 ion also dissociates from the active site and the amino group from the bonded substrate also becomes coordinated to the Ni2 ion (Scheme 4—Step ❷). This configuration enhances the proximity between urea and the hydroxide ion that is also bidentate to both Ni(II) ions. In the next step, the nucleophilic attack of the hydroxide anion to the carbon atom of urea occurs, at the same time that one proton is transferred from the first molecule to the amino group of the second one. From this reaction results a carbamate molecule that remains coordinated to the metal cofactor and ammonia that is released to the solvent (Scheme 4—Step ❸). The generation of the active site occurs after the dissociation of carbamate from the metal cofactor and, the subsequent coordination of three water molecules directly to the Ni(II) ions (Scheme 4—Step ❹). In solution, the carbamate molecule is spontaneously degraded into carbon dioxide and ammonia [100,101].

Experimental studies indicates a k_{cat} for the substrate urease that can vary between 1250 and 3500 s^{-1} depending on the organism and the temperature and pressure conditions [78,102–104]. Based on a kinetic analysis, it has been predicted that urease is the most proficient enzyme identified to date to convert urea into carbon dioxide [104].



Scheme 4. General description of the catalytic mechanism of urease. Only the atoms involved in the first coordination shell of the metal ion were represented to simplify the representation.

2.4. Binuclear Zn Cofactor

So far, we have discussed the catalytic mechanism of metalloenzymes that catalyze the formation of hydroxide anions and employ a monodentate Zn cofactor (CA and HDAC8) or a binuclear Mn (prolidase) or binuclear Ni (urease) cofactors. The main difference between the mechanisms employed by these enzymes is the way that each substrate reacts with the hydroxide anion. In the case of Ni(II)- and Mn(II)-dependent enzymes, the substrate become bidentate to two metal ions becoming stereo-chemically positioned to undertake the nucleophilic attack of the coordinated hydroxide anion. In the case of the monodentate Zn enzymes, this is not observed since the substrate becomes bidentate to the same Zn ion that also shelters the same hydroxide anion. This happens because Zn is an exceptionally good Lewis acid for biological catalysis and allows a fast interconversion among its four-, five-, and six-coordinate states.

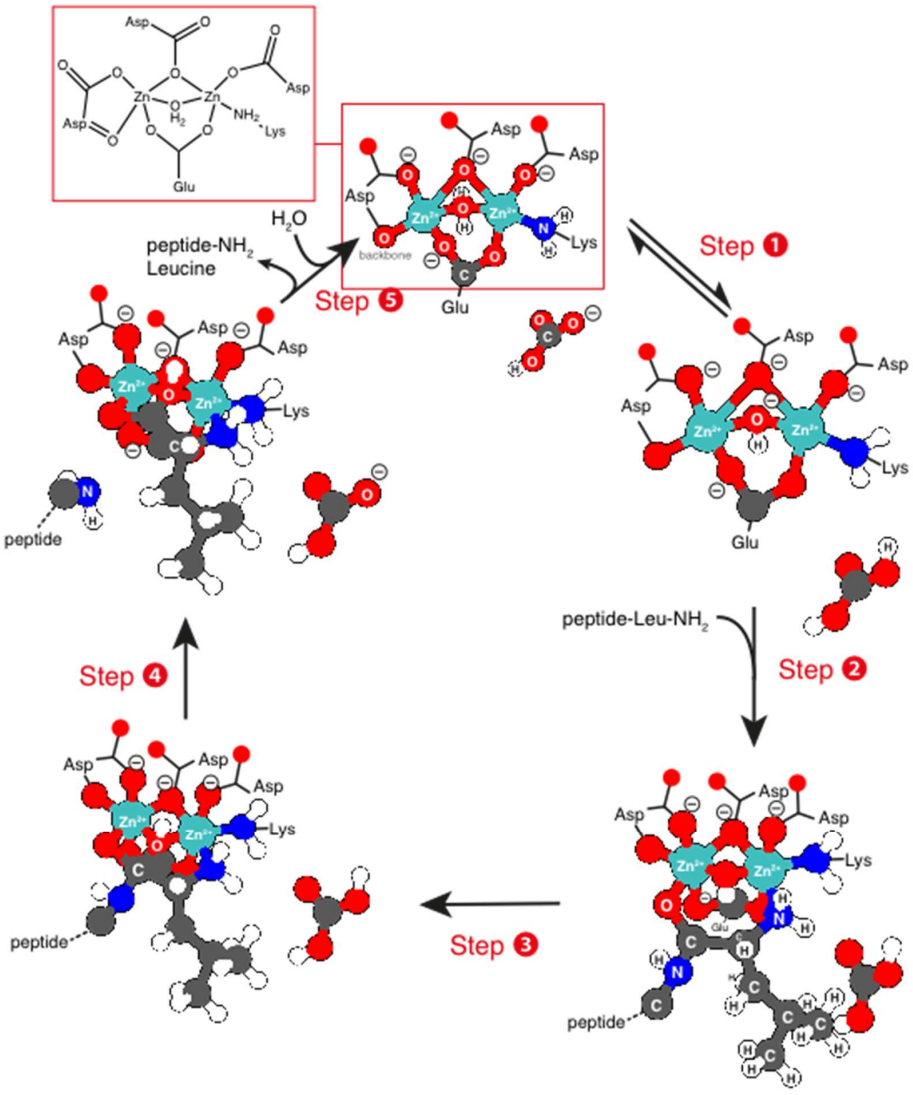
Although these findings suggested that the Zn-dependent metalloenzymes that are involved in the formation of hydroxide anions are all monodentate, one enzyme was found that produces the same unstable molecule but lodges two catalytic Zn ions in the active site. This enzyme is leucine aminopeptidase (LeuAP; EC 3.4.11.1) and catalyzes the hydrolysis of leucine residues from the *N*-terminus of a wide variety of di- and tripeptide substrates [105–110].

LeuAP, together with other aminopeptidases, carboxypeptidases, phosphatases, β-lactamase, and others, belongs to the hydrolase enzyme class [111,112]. The zinc hydrolase superfamily includes both mono zinc enzymes, such as carboxypeptidase or thermolysin, as enzymes that harbors co-catalytic zinc sites, such as, aminopeptidase, β-lactamase, or alkaline phosphatase [113].

The active site of this enzyme, LeuAP, contains two non-equivalent Zn ions placed at about 3.02 Å away from each other [107]. Each metal ion has a different coordination sphere that endorses different characteristics to each metal ion. The Zn(II) ion at site one (Zn1) is penta-coordinated to two aspartates and one glutamate residue. The Zn(II) ion at site two (Zn2) is also penta-coordinated

with two aspartates, one glutamate, and one lysine. The glutamate and one of the aspartates are also bridging the two metal ions [114].

The currently accepted mechanism for this enzyme was proposed based on computational means and agrees very well with the available experimental data [114,115]. The first step of the catalytic mechanism starts by the activation of the Zn1–Zn2 metal ions through the abstraction of a proton from the water molecule (that is coordinated to both ions) by carbonate and from which results a hydroxide anion (Scheme 5—Step ❶). Then, the peptide (substrate) becomes coordinated to the metal ion through the carbonyl group of the backbone and the terminal amine group (Scheme 5—Step ❷). Afterwards, the nucleophilic attack of the hydroxide anion to the α-carbon of the scissile peptide bond of the substrate to form a tetrahedral gem-diolate reaction intermediate occurs (Scheme 5—Step ❸). Subsequently, the carbonic acid donates its previously acquired proton to the nitrogen atom of the scissile peptide bond of the substrate, at the same time that Asp273 accepts a proton from hydroxyl group attached to α-carbon of the substrate (Scheme 5—Step ❹) [108,114,115]. These concerted proton transfers lead to the cleavage of the peptide bond at the *N*-terminus of the peptide (preferentially) releasing a leucine molecule (Scheme 5—Step ❺).



Scheme 5. General description of the catalytic mechanism of leucine aminopeptidase (LeuAP). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

A computational study has shown that the rate-limiting step for the reaction with L-leucine-*p*-nitroanilide is the cleavage of the peptide bond with a barrier of 17.8 kcal/mol [114]. This result is in agreement with the experimentally measured barrier of 18.7 kcal/mol estimated from the k_{cat} value of $0.117 s^{-1}$ [116].

Comparing the mechanism of the mononuclear Zn-dependent enzymes with the binuclear ones, they present one big difference. In the first type of enzymes, the substrate becomes coordinated to the same Zn ion that also shelters the same hydroxide anion. In the binuclear Zn(II) cofactors, the substrate acts as bidentate ligand, coordinating both metal ions and promoting a close and very specific access of the hydroxide anion (also coordinated to both metal ions) to the electrophilic atom of the substrate.

The metal site of LeuAP is commonly termed as a readily exchangeable site in which the Zn1 ion can be replaced by Mg^{2+}, Co^{2+}, and Mn^{2+} ions without destroying the activity of the enzyme. Site two, however, is referred to act as a tight binding site, where the Zn2 ion can only be replaced by a Co^{2+} ion [117]. Computational studies have shown that the electronic nature of the metal ion at site one influences only the generation of the gem-diolate intermediate, suggesting that different metals can be used to turn the reaction specific for different types of substrate [114].

3. Enzymes that Catalyze the Hydride Transfer Reactions

Hydride transfer is another elementary process catalyzed by several metalloenzymes. Unlike hydroxide anions (OH^-), hydride ions do not exist in aqueous solutions as free ions. For this reason, they can only be transferred directly from one organic moiety to another one through the simultaneous breakage and formation of covalent bonds. The metal cofactor present in the enzymes that catalyze this type of reaction often provides the binding site of the substrate and are directly involved in the hydride transfer. The metalloenzymes that house molybdenum, tungsten, and zinc ions often catalyze hydride transfer reactions. The catalytic mechanism of these enzymes will be reviewed in the next sections.

3.1. Molybdenum and Tungsten Cofactors

The molybdenum is found in a variety of enzymes that catalyze hydride transfers that are coupled with redox reactions. Examples of these enzymes are present in the xanthine oxidase (EC 1.17.3.2, XO) [118], DMSO reductase (EC 1.8.5.3) [119], and sulfite oxidase (EC 1.8.3.1, SO) [120] families [121]. In these enzymes, the Mo ion is coordinated to one or two pyranopterin dithiolene ligands (PDT), forming what is known as a molybdenopterin cofactor. The enzymes from DMSO reductase are coordinated to two PDT ligands, while the enzymes from the XO and SO families have only one PDT ligand. Depending on the enzyme, the coordination sphere around the Mo ion is completed by a wide range of ligands that can range from oxygen or sulfur anions, or side chains from amino acid residues such as cysteine, selenocysteine, glutamates, or serine [122]. In prokaryotes, tungsten can also be found in the place of Mo, but they share similar functions [123].

The essential role played by Mo (or W) during catalysis is to promote a controlled *oxo*-transfer between the substrate and the metal cofactor that involves, at some point, a hydride transfer that is coupled to an electron transfer between the substrate and the external oxidant, reductant species.

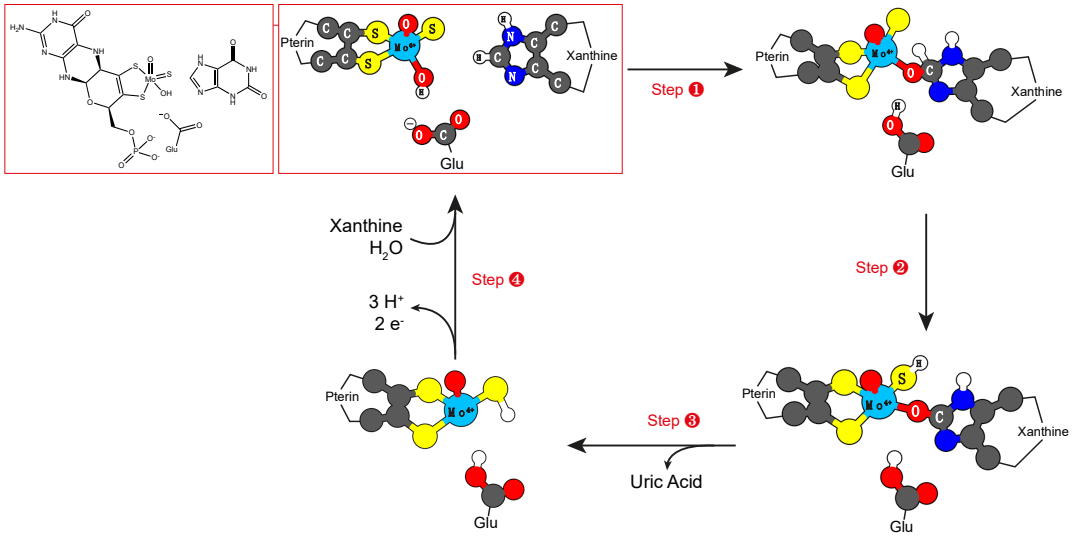
The representative reactions that are catalyzed by these enzymes are generally well understood and can be summarized based on the following chemical equations:

- Xanthine oxidase: $RCHO + H_2O \rightarrow RCO_2H + 2\,H^+ + 2\,e^-$;
- Sulfite oxidase: $SO_3{}^{-2} + H_2O \rightarrow O_4{}^{-2} + 2\,H^+ + 2\,e^-$;
- DMSO reductase: $HCO_2{}^- \rightarrow O_2 + 2\,H^+ + 2\,e^-$.

In the case of the enzymes from the XO family, the first step of the catalytic mechanism involves a proton abstraction from the equatorial Mo–OH, by a glutamate residue [95,124]. Subsequently, the nucleophilic attack of the oxygen that is coordinated to the Mo ion to a carbon of the substrate occurs (Scheme 6—Step ❶). This leads to the formation of a penta-coordinated complex around the Mo

ion. Afterwards, the hydride transfer from the substrate to the Mo = S group occurs (Scheme 6—Step ❷), as does the cleavage of the bond between the oxygen and the Mo ion (Scheme 6—Step ❸). After the oxidation of the cofactor (electron transfer from the Mo to the other redox-active centers present on the enzyme), a water molecule occupies the free coordination position around the Mo ion. The enzymatic turnover only occurs after two deprotonation steps and the enzyme is ready for a new cycle (Scheme 6—Step ❹).

The k_{cat} for the conversion of xanthine to uric acid by the human XO is of 18.3 s^{-1} [125] but it can vary between 0.41 and 108 depending on the species and the experimental conditions [78].



Scheme 6. General description of the catalytic mechanism of xanthine oxidase (XO). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

The enzymes from the SO family also use a water molecule to catalyze a simple *oxo*-transfer from sulfite to form sulfate and the mechanism also involves a hydride transfer from the substrate to the metal cofactor. These enzymes thus employ a similar mechanism to the one that was described for the enzymes of XO family [126].

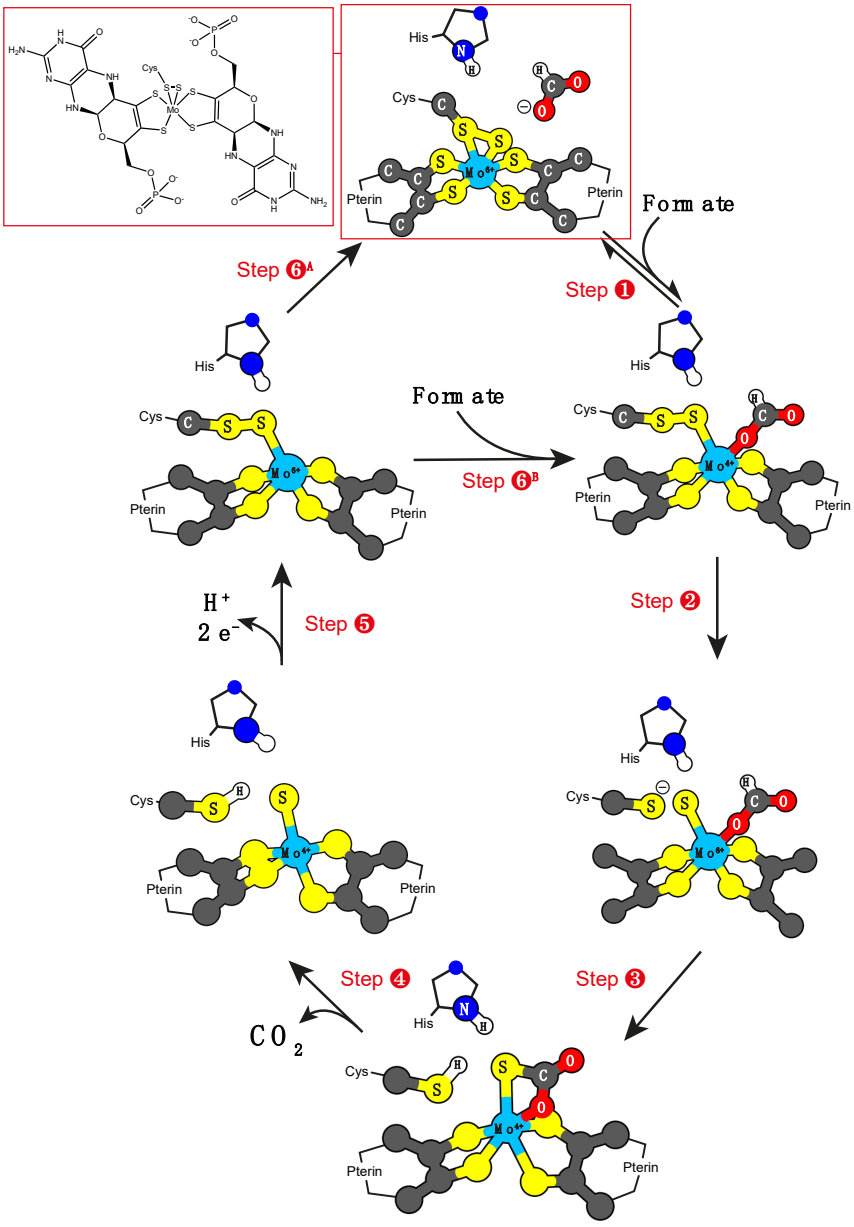
The catalytic mechanism of the enzymes from the DMSO family is quite different from the ones present on the SO and XO families. Part of these differences may be attributed to the presence of two PDT ligands in the cofactor instead of one, and of course by the different substrates that they catalyze (they are negatively charged) [127,128].

The catalytic mechanism of formate dehydrogenase (EC 1.2.1.2, FdH) is described below to exemplify how hydride atom transfer reactions can be catalyzed by the enzymes from the DMSO family [129,130].

FdH catalyzes the oxidation of the formate anion to carbon dioxide in a redox reaction that involves the transfer of a hydride and two electrons from the substrate to the external oxidizing species [131].

When substrate is not present at the active site of FdH, the Mo (VI) ion is hexa-coordinated and no free position is available to bind the substrate (formate) (Scheme 7—Step ❻^{A}). In such state, the cofactor has a total charge of −1 and this led several authors to propose, in the past, that the catalytic mechanism should occur through the second shell of the metal ion, as the approach and binding of the negative charged substrate to the also negatively charged cofactor would be extremely unfavorable. Recently, it was found that the mechanism takes place through a first-shell type of mechanism, in which the substrate binds and reacts directly with the Mo ion. This is only possible due to a rearrangement on the coordination sphere around the Mo ion that takes place when the substrate comes close to it. This rearrangement was named "sulfur-shift mechanism" and is defined as a change in the Mo ion coordination sphere, which involves a first-to-second shell displacement (shift) of the SeCys ligand of

the Mo ion (Scheme 7—Step ❻^{B}), resulting in a free coordination position that is used by the enzyme to coordinate an oxygen atom from formate directly to the Mo ion (Scheme 7—Step ❻^{B} or ❶) [132]. The mechanism was first discovered by theoretical means and later on validated by experimental means [132–135]. Interestingly, the same type of activation mechanism was found in another enzyme from the DMSO family (nitrate reductase) but catalyzes a different reaction.



Scheme 7. General description of the catalytic mechanism of formate dehydrogenase (FdH). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

The next step of the catalytic mechanism in FdH involves the cleavage of the bond that connects SeCys to the inorganic sulfur (Scheme 7—Step ❷). This cleavage is possible thanks to the side chain of an active site histidine residue, which forms a hydrogen bond with the resulting selenol anion and stabilizes it. Subsequently, the active site histidine residue catalyzes the proton transfer from formate to SeCys by forming a hydrogen bond with the selenol (Scheme 7—Step ❸). Once this reaction is complete, the SeCys moves away from the active site, resembling what has been observed in the

re-interpretation of the X-ray data of reduced from Fdh-H [136]. At the same time that the proton transfer between formate and the SeCys occurs, a pair of electrons is transferred from the carbon to the cofactor that becomes reduced.

A great deal of discussion has been held on this step, since this step has always been regarded as a hydride transfer and not a proton transfer. A classical hydride transfer involves the transfer of a hydrogen atom and a pair of electrons from the carbon of the substrate to another molecule. In the case of FdH, a similar mechanism occurs. Indeed, there is a proton transfer between the substrate and the SeCys, and a pair of electrons are transferred from the carbon of formate to the metal cofactor.

Once the "hydride transfer" takes place, the resulting carbon dioxide molecule does not dissociate from the cofactor and remains bound to the Mo ion. The next step of the catalytic cycle involves the release of carbon dioxide (Scheme 7—Step ❹) and the oxidation of the cofactor (Scheme 7—Step ❺). The enzymatic turnover only takes place once the proton bound to the SeCys side chain is transferred to the solvent or to another residue in the catalytic pocket. This event forces a conformational rearrangement of the loop where SeCys is located, to occupy a position near the Mo ion. Once this is accomplished, the SeCys forms a covalent bond with the sulfide ligand that is coordinated to the Mo ion. From this point, two pathways can be followed. If no substrate is available at the active site, the sulfur-shift mechanism is reversed, and the metal site adopts the original resting conformation (Scheme 7—Step ❻$^{\mathbf{A}}$). If formate is available, it can occupy the free coordinating position at the Mo ion and the catalytic process continues, without further delay (Scheme 7—Step ❻$^{\mathbf{B}}$) [137].

Computational studies addressing FdH have shown that no differences are observed if Mo or W are used to model the reactions that were described, accentuating the similar chemical role played by both metal ions.

A study combining computational and experimental data showed that the rate-limiting step of the FdH catalyzed reaction involves the transfer of the proton from the substrate to the selenium with an activation barrier of 21.2 kcal/mol [131].

The experimental k_{cat} for some molybdenum-containing formate dehydrogenases varies between 1.4 and 347 s^{-1} [131,138].

3.2. Zinc Cofactors

Zinc cofactors can also catalyze hydride transfers. This type of reaction can, for example, be found in alcohol dehydrogenases (EC 1.1.1.1, ADH), a group of dehydrogenase enzymes that facilitates the interconversion between alcohols, ketones, and aldehydes through the reduction of nicotinamide derivatives NAD^+ or $NADP^+$ to the corresponding 1,4-dihydronicotinamides [139–142]. This role is particularly important in humans and other animals, to degrade alcohols that otherwise would be toxic for the cells.

Each monomer of ADH houses two Zn ions in the active site: Zn1 and Zn2, which act as structural and catalytic metal ions, respectively. One of the Zn ions (Zn1) is coordinated to four cysteine residues and acts as a structural metal, whereas the other Zn ion (Zn2) is coordinated to two cysteine residues, one histidine residue, and a water molecule. The water molecule also establishes a hydrogen bond with a catalytic dyad formed by a serine and histidine residues from the active site.

The catalytic mechanism starts with the binding of NAD^+ to the active site. [143] Once substrate (for example ethanol) is available, it becomes coordinated to the Zn2 ion and forces the water molecule to dissociate from the active site (Scheme 8—Step ❶). At the end of this process, the serine from the catalytic dyad becomes hydrogen bonded to the hydroxyl group of the substrate.

The next step involves the proton transfer from the hydroxyl group that is coordinated to the Zn(II) ion to an active site residue that behaves as a base. This occurs through a sequential proton transfer involving the serine and histidine residues from the catalytic dyad (Scheme 8—Step ❷). At the end of this step, an alkoxyde intermediate is generated that remains coordinated to the Zn2 ion. The third step of mechanism involves the hydride transfer, from the reaction intermediate to NAD^+ (Scheme 8—Step ❸) [144]. From this reaction results the formation of an aldehyde that

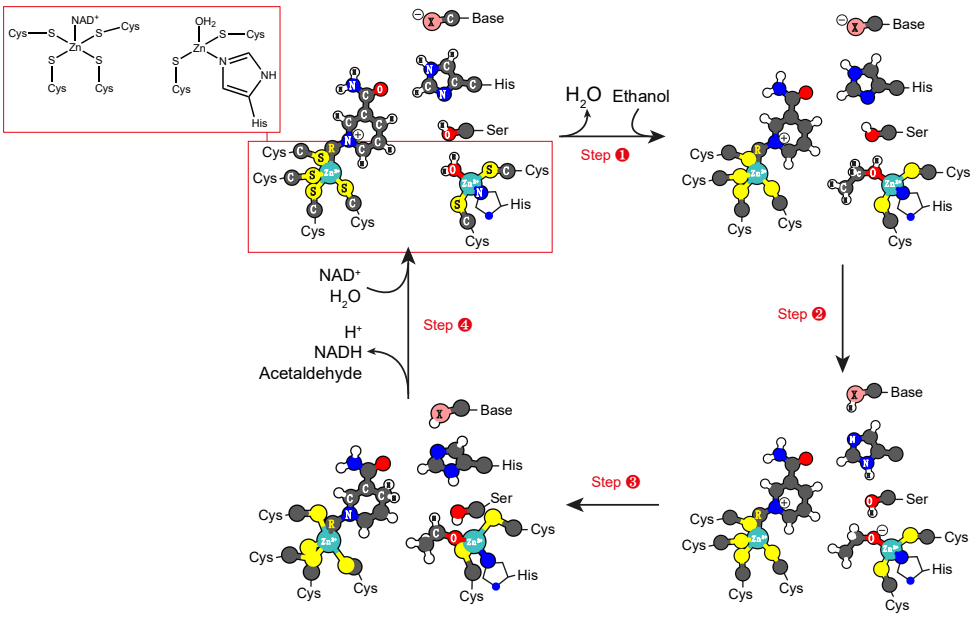
remains coordinated to the Zn(II) ion and NADH. The last step of the catalytic mechanism involves the dissociation of the aldehyde and the regeneration of the active site (Scheme 8—Step ❹). This is accomplished by the coordination of a water molecule to the Zn(II) ion, the proton transfer from the amino acid residue that behaves as a base with the catalytic dyad, and the loss of a proton [145].

In the full catalytic mechanism, only the Zn2 ion has a catalytic function, where it is important to polarize the C–O bond of the substrate, which is important to help the hydride migration to the NAD^+ molecule. However, the importance of Zn1 ion in the mechanism cannot be underestimated. In fact, both metal ions have a preponderant role in the initial steps of the mechanism, where they help the correct orientation of the substrate in such an orientation that the reactive CH_2 group of the nicotinamide ring of NAD^+ points toward the hydroxyl group of the substrate.

Currently, it is still not clear which residue from the active site plays the function of the base and if it is in close contact with the catalytic dyad.

Computational studies predict an activation barrier of 21.9 kcal/mol for the rate-limiting step for the oxidation mechanism of benzyl alcohol by ADH [144].

Another recent theoretical study focused on the study of the reaction catalyzed by a secondary ADH showed, among others, a calculated overall barrier of 13.4 kcal/mol for (R)-2-butanol [143], which is consistent with the experimental value of 15 kcal/mol estimated from the experimental k_{cat} value of 0.35 s^{-1} [146].



Scheme 8. General description of the catalytic mechanism of alcohol dehydrogenases (ADH). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

4. Enzymes that Catalyze the Formation of Radicals

The production of radical species by enzymes are of great importance. They produce very reactive and unstable chemical species that have a preponderant role in every living organism. Iron, manganese, and cobalt are the most common metal ions acting as cofactors present on these enzymes.

One of the best examples of enzymes that master the synthesis of radicals in nature are the ribonucleotide reductases (RNR). These enzymes catalyze the only pathway for de novo synthesis of the monomeric building blocks for both DNA replication and DNA repair [147–149].

Arguably the most difficult step during RNR-catalyzed ribonucleotide reduction is the initial activation of a chemically unreactive C–H bond [150,151]. The only pathway known to date involves

the formation of a radical in a conserved cysteine residue that is capable of abstracting the 3'H-atom from the ribose moiety. This initial radical is, in turn, always generated via a metal-containing cofactor.

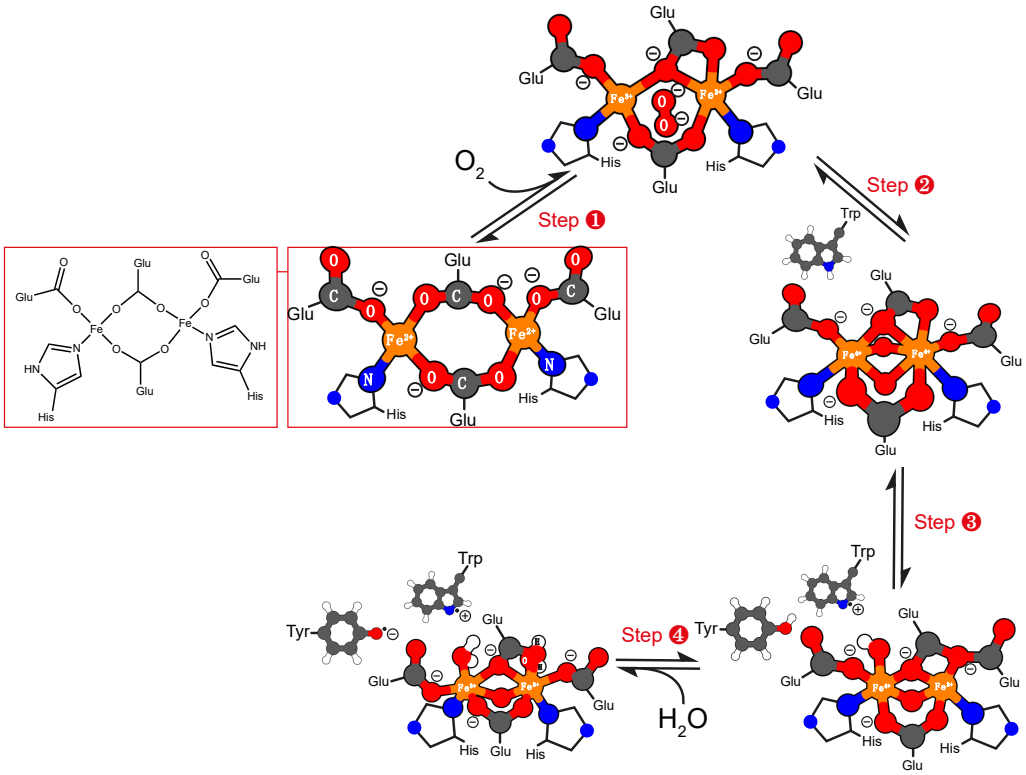
There are different RNRs that were divided in different classes based on the type of cofactor that they generate and require for the catalytic process [152]. The most studied ones are from class I and class II. Class I RNRs are oxygen-dependent enzymes and are present in both eukaryotes and eubacteria and contain a binuclear metal cluster formed by Fe or Mn [153–155]. Class II RNRs, found in both aerobic and anaerobic microbes, are oxygen-independent and use a cobalt-containing cobalamin (vitamin B12) cofactor [156,157].

4.1. Iron Cofactor

In the RNRs from class I, the tyrosyl radical is generated by a bi-iron metal cofactor that is channeled through 35 Å to a strictly conserved cysteine residue that is located in the active site very close to the substrate [158]. The active site and the metal cofactor are located in different subunits of the enzyme that need to dimerize in order to turn the enzyme active.

In the case of RNR from class Ia, the metal cofactor is composed by two Fe(II) ions that are coordinated to three glutamate and a histidine residue. Two of these glutamate residues are bi-dentated to both catalytic Fe(II) ions. Depending if one of these glutamates is bi-dentated or mono-dentated to the Fe(II) ions, the cofactor can adopt a five-coordinated or a distorted tetrahedral configuration [156,159].

The radical begins being produced when a di-oxygen molecule reacts with the Fe(II)–Fe(II) metal cofactor, from which results a peroxide intermediate and an oxidized Fe(III)–Fe(III) metal cofactor (Scheme 9—Step ❶) [160]. During this process, one of the glutamate residues that was bi-dentated to both Fe ions forms a single oxo-bridge between the ions and only one of the Fe ions becomes bi-dentated.



Scheme 9. Schematic representation of the catalytic mechanism of ribonucleotide reductases (RNR) from class Ia (Fe–Fe). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

Afterwards, the cleavage of the peroxide occurs, as does the formation of two oxo-bridges between the two metal ions, from which results the oxidation of Fe(III)–Fe(III) to Fe(IV)–Fe(IV)

(Scheme 9—Step ❷). Subsequently, there is a homolytic cleavage of a hydrogen atom from a tryptophan residue to one of the oxygen atoms that is coordinated to the Fe(IV)–Fe(IV) metal cluster (Scheme 9—Step ❸). The resulting hydroxyl molecule dissociates from the cofactor, reducing one of the Fe(IV) to Fe(III). The free position in the coordination sphere around the iron ion is promptly filled by a water molecule that is available in the active site.

The final step of the mechanism involves the formation of the tyrosyl radical (Scheme 9—Step ❹). This involves the loss of a hydrogen atom from the other hydroxyl molecule. During this process, the hydroxyl is reduced to water and the Fe ion is reduced from Fe(IV) to Fe(III).

At the end of this reaction, a tyrosyl radical is generated, and the radical is transferred over 35 Å along the protein to the active site of the enzyme where it is finally accepted by a cysteine residue [156,161].

The experimental k_{cat} obtained for RNR from class Ia varies depending on the species, the concentration of an allosteric effector (ATP), and the quaternary structure of the enzyme used in the assays. They are, for example, in the range between 0.047 and 0.30 s^{-1} [162] or 2 and 10 s^{-1} [163] for the substrate cytidine diphosphate (CDP) and have values of 0.25 s^{-1} for uridine diphosphate (UDP), 0.18 s^{-1} for adenosine diphosphate (ADP), and 0.28 s^{-1} for guanosine diphosphate (GDP) [162].

4.2. Manganese Cofactor

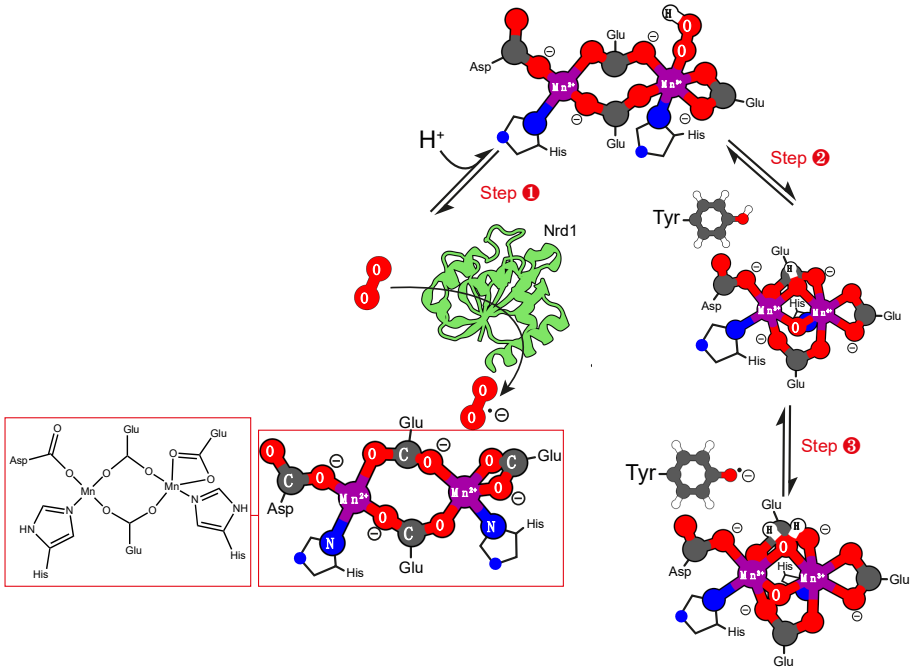
In RNRs from class Ib, the tyrosyl radical is generated by a Mn–Mn cofactor, where both metal ions are catalytic. These enzymes also require the presence of an additional protein Nrd1, otherwise no activity is observed for the Mn(II)–Mn(II) center [154,164].

Little is known about the mechanism from which the radical is formed in this class of RNR enzymes. It is known that the process starts by the action of protein Nrd1 that reduces an O_2 molecule to O_2^- and tunnels it towards the Mn(II)–Mn(II) center of the enzyme. [164] The presence of protein Nrd1 is justified due to the limitation of Mn(II)–Mn(II) center that cannot reduce the oxygen, as it is observed in the Fe(II)–Fe(II) cluster of RNRs from class Ia [154].

The O_2^- molecule then binds to one of the Mn(II) atom, oxidizing it to Mn(III) (Scheme 10—Step ❶). At the same time, one of the oxygens is protonated and an oxo-bridge and a hydroxyl-bridge between the two metal ions is established (Scheme 10—Step ❷). At the end of this process, the metal ions Mn(II)–Mn(II) are oxidized to Mn(III)–Mn(IV). The resulting reaction intermediate is very similar to the one that was observed in the mechanism of class Ia (Scheme 9) [164,165].

In the last step of the catalytic mechanism, a nearby tyrosine residue loses one hydride for the hydroxyl group that is coordinated to the two Mn ions, generating the desired tyrosyl radical (Scheme 10—Step ❸). When the hydride transfer occurs, the Mn(II) is reduced by the hydroxyl molecule, and a water molecule is formed [164,165]. Once these steps are complete, the enzyme is ready for a new turnover.

In contrast to the Fe–Fe center of class Ia, no tryptophan residue is involved in the formation/stabilization of the tyrosyl radical in class Ib [164].



Scheme 10. Schematic representation of the catalytic mechanism of RNR from class Ib (Mn–Mn). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

4.3. Cobalt Cofactor

The RNRs from class II are oxygen-independent and use a cobalt-containing cobalamin (vitamin B_{12}) cofactor. Although many studies have been devoted to understanding the mechanism for the radical formation, it continues to be poorly understood. However, other enzymes employ the same cofactor to generate radicals, from which more information is available and can provide more insights to understand more clearly how these radicals are generated. One of these enzymes is methionine synthase (EC 2.1.1.13, MetH), which catalyzes the regeneration of methionine from homocysteine. This enzyme is a modular protein that contains different domains, which are implicated in the binding and activation of different substrates and cofactors involved in the catalytic cycle.

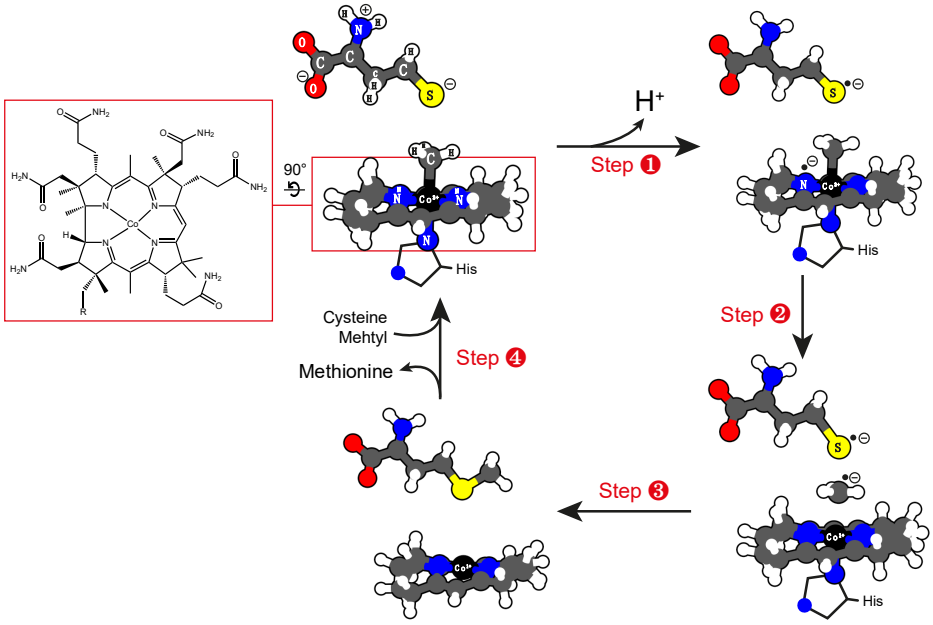
One of these domains contains a Zn ion that activates the substrate of the enzyme, homocysteine, to a reactive thiolate. The other domain of MetH contains a cobalt-containing cobalamin (vitamin B12) cofactor and is where the catalytic reaction takes place.

Before the cobalt-containing cobalamin (vitamin B_{12}) cofactor can react with the activated substrate, it needs to be activated by 5-methyl-tetrahydrofolate (5-methyl-THF) [166,167]. In this process, the methyl group from the 5-methyl-THF becomes coordinated directly to the metal ion and THF is released to the solvent. During this reaction, the metal cofactor is oxidized from Co(I) to Co(III), an active site histidine residue becomes coordinated to the Co(III) ion, and the metal cofactor changes from a tetra-coordinated to a hexa-coordinated configuration [166].

The reaction of the substrate with the Co cofactor has been a subject of great debate in the last decade [168–172]. Some authors suggest that it goes through a S_N2 type of mechanism and involves the nucleophilic attack of the anionic sulfur of the substrate to the carbon of the methyl group that is coordinated to Co ion. In this hypothesis, no radical species are involved in the reaction, and at the end of this process, the Co(III) becomes reduced to Co(I) [168–173]. A new hypothesis, however, has been advanced that involves a radical reaction [174] that is in line with recent studies devoted to this enzyme [175]. In this new version of the mechanism, the negatively charged substrate acts as a reductive specie transferring one electron to the Co cofactor, generating two radical species inside the active site (Scheme 11—Step ❶). Subsequently, the electron from the cofactor is transferred to

the methyl group, from which results a methyl radical (Scheme 11—Step ❷). The group then binds to the thiol radical of the homocysteine and forms methionine (Scheme 11—Step ❸). At the end of this process, the Co(III) ion is reduced again to Co(I) and the enzyme is ready for a new turnover (Scheme 11—Step ❹) [175,176].

Computational studies estimate an energy barrier for the described reaction of ~8.5 kcal/mol and is comparable to the also theoretically obtained barrier for the S_N2 reaction pathway hypothesis, which is of 10.5 kcal/mol [173,175,176]. An experimental k_{cat} of 27.1 s^{-1} was obtained under specific experimental conditions, in line with the computational studies [177].



Scheme 11. Schematic representation of the catalytic mechanism of methionine synthase (MetH). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

The cobalamin cofactor, and in particular the Co ion, has a preponderant role in this mechanism, first promoting the stabilization of the radical specie, and secondly, allowing the stabilization of the methyl radical, which otherwise could be toxic to the cells.

5. Enzymes that Catalyze the Formation of Superoxide

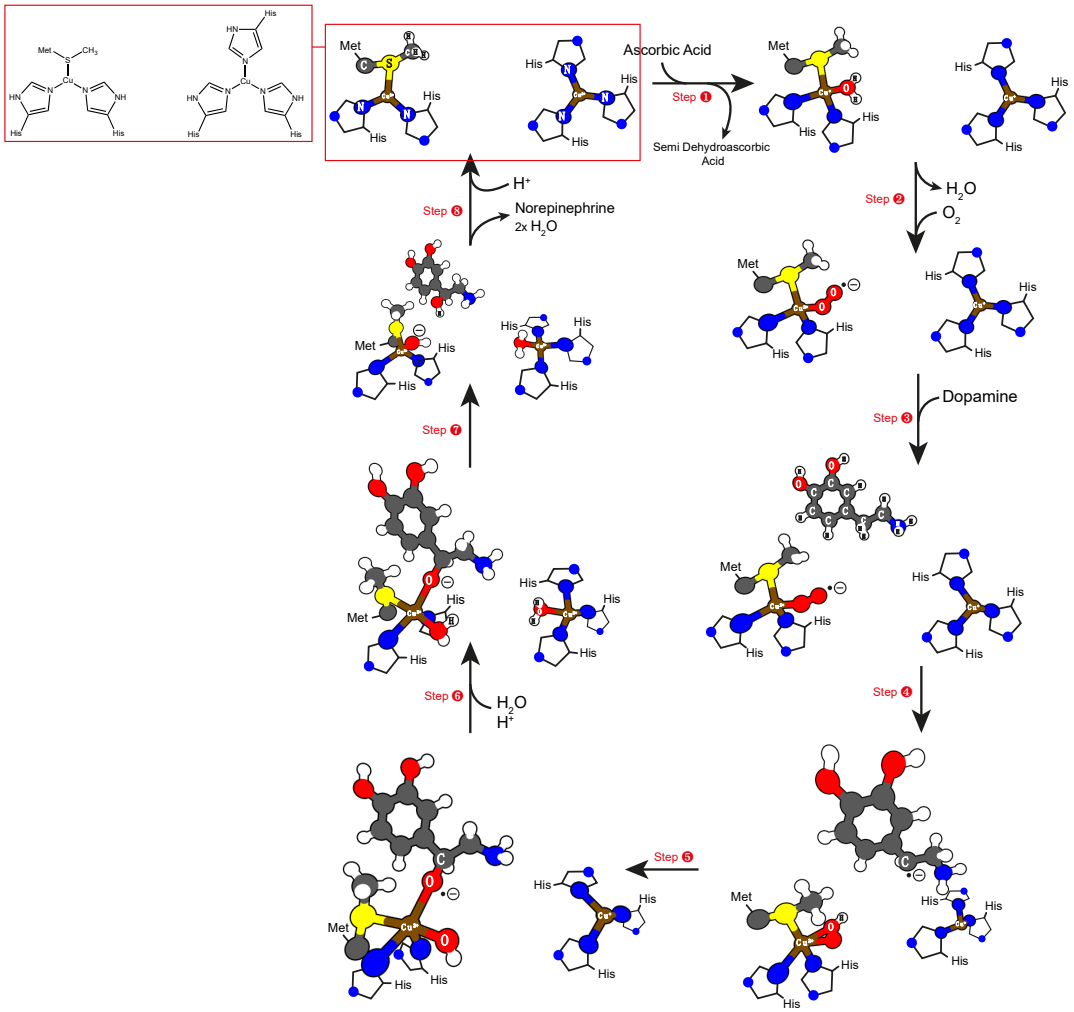
A superoxide is a compound that contains the superoxide anion with the chemical formula O_2^-. Superoxide is formed by some enzymes to catalyze some specific reactions. Because superoxide can be very toxic to the cells, these metalloenzymes keep it coordinated to the metal cofactors and therefore preclude their dissociation to the solvent, a condition that could put the cells in danger. Two examples that employ copper and iron cofactors are described in the following sections.

5.1. Copper Cofactor

Dopamine β-hydroxylase (EC 1.14.17.1, DBH), also known as dopamine β-monooxygenase, is an enzyme that has an uncommon non-coupled binuclear Cu–Cu cofactor and catalyzes the production of norepinephrine and water through the hydroxylation of the benzyl carbon of dopamine and di-oxygen through a radical mechanism [178].

The crystallographic structure of this enzyme [179] shows one copper atom (Cu1) coordinated to two histidine residues and a methionine residue. The second copper metal ion (Cu2) is coordinated by three histidine residues [180]. Both Cu1 and Cu2 metal ions play a catalytic role in the mechanism.

The first step of the catalytic mechanism starts with the reduction of the two Cu(II) atoms to Cu(I) by two ascorbic acid molecules that become oxidized to semi-dehydroascorbate (Scheme 12—Step ❶) [181]. Afterwards, a di-oxygen molecule binds to Cu1(I), where it is rapidly reduced to O_2^- and Cu1(I) is oxidized back to Cu(II) (Scheme 12—Step ❷).



Scheme 12. General description of the catalytic mechanism of dopamine β-hydroxylase (DBH). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

When substrate is available in the active site (dopamine), it becomes enclosed by a network of hydrogen bonds provided by a neighbor tyrosine and two glutamate residues that ensures the stereospecific of the chemical reaction and moves it close to both cooper metal ions (Scheme 12—Step ❸). Afterwards, the homolytic cleavage of the substrate occurs, and by the action of the activated di-oxygen anion (O_2^-), an OOH species is formed that becomes coordinated by both oxygens to Cu1 (Scheme 12—Step ❹). The generated radical located in the carbon of the substrate then attacks the non-protonated oxygen of the OOH species, cleaving the O–O bond and generating an oxo-radical (Scheme 12—Step ❺). At this moment, an electron-transfer between the Cu2 and Cu1 sites, that is not completely understood, takes place. The Cu2 becomes oxidized (Cu(II)) and becomes coordinated to an additional water molecule (Scheme 12—Step ❻). At the same time, the oxo-radical of the substrate undergoes a one-electron reduction, and a proton donor (probably a water molecule) provides a proton that converts the hydroxyl species into a water molecule. The last step of the catalytic mechanism involves the abstraction of a proton from the water molecule coordinated to the Cu1 atom by the

negatively charged oxygen atom of the substrate (Scheme 12—Step ❼). Subsequently, the substrate becomes hydroxylated, and it is quickly released from the active site, leaving the enzyme ready for a new catalytic cycle (Scheme 12—Step ❽) [180].

The role played by both Cu ions in the catalytic mechanism of DBH is very distinct but both of them have an active role in the chemical processes that takes place in the active site. The Cu1 is responsible for the initial activation of the dioxygen molecule and the Cu2 is required for the second one-electron reduction. It is worth mentioning that the hydroxylation function of DBH is only possible due to the uncommon non-coupled binuclear Cu–Cu cofactor. By displacing the two metals, the enzyme enables the two-electron oxidation of the substrate through two sequential one-electron steps [180]. This highly controlled oxidation is necessary to prevent side reactions that could oxidize the hydroxyl and amine groups of reaction intermediates.

The experimental k_{cat} found for the substrate dopamine varies between 13 and 110 s^{-1} depending on the experimental conditions [182,183].

5.2. Binuclear Iron Cofactor

The generation of reactive and unstable oxygen species can also be accomplished by enzymes containing iron cofactors. One of these examples is found in the enzyme myo-inosnitol oxygenase (EC 1.13.99.1, MIOX) that catalyzes the oxidation of myo-inositol (MI) to glucuronic acid (GA) [182–184].

MIOX employs two catalytic iron metals, (Fe1(II) and Fe2(II)), that are kept together through the simultaneously coordination to a hydroxide molecule. Both iron ions are also coordinated by two histidine residues. The Fe1(II) is also coordinated to a water molecule and an oxygen atom from the carboxylate group of an aspartate residue from the active site.

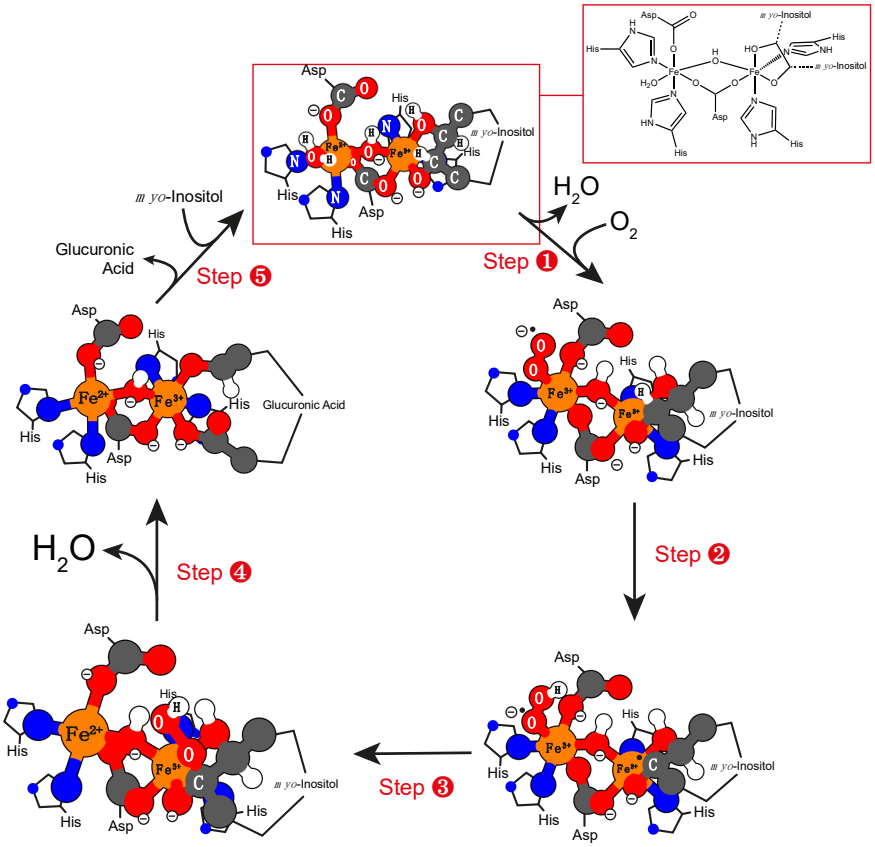
When substrate is available in the active site (MI), it becomes coordinated to the Fe2 ion through two hydroxyl groups. In the first step of the catalytic mechanism (Scheme 13—Step ❶), the water molecule that is coordinated to the Fe1(II) ion is replaced by an oxygen molecule. This molecule is then rapidly reduced to superoxide at the same time that Fe1(II) is oxidized to Fe1(III).

The superoxide molecule remains coordinated to the Fe1(III) ion and in the next step, the hydride transfer from the carbon atom of the MI (Scheme 13—Step ❷) to the superoxide molecule takes place. This step originates a radical intermediate that is stabilized by the Fe2(III) and a hydroperoxide that remains bounded to the Fe1(II) ion. These two species have a short half-life time, reacting quickly with each other in the next step of the catalytic mechanism. In this reaction, the hydroperoxide becomes covalently bonded to the radical intermediate and the generated compound remains coordinated to the Fe2(III) (Scheme 13—Step ❸).

In the next step of the catalytic mechanism (Scheme 13—Step ❹), the hydroxyl group that is coordinated to Fe2(III) is deprotonated by a base, at the same time that the cleavage of the hydroperoxo group and the concomitant cleavage of the C–C bond that originates the final product, GA, takes place [185]. Currently, there are no data that confirms what the base that abstracts the proton during this step is, but Morokuma and co-workers suggested that the hydroxide molecule placed between the two metal ions could play that role [185].

Finally, the GA is released from the active site and the enzymatic turnover takes place with the entrance of a new molecule of MI (Scheme 13—Step ❺) [184–186].

According to theoretical data, the rate-limiting step of this mechanism is the breakage of the O–O bond that occurs in step 4, requiring an activation energy of 26.9 kcal/mol [185]. Experimental data show a k_{cat} for the substrate myo-inositol of 0.183 s^{-1} [187].



Scheme 13. General description of the catalytic mechanism of myo-inosnitol oxygenase (MIOX). Regarding amino acid residues belonging to the protein, only the atoms involved in the coordination of the metal were represented to simplify the representation.

6. Conclusions

Nature has developed highly efficient metalloenzymes bearing one or more metal cofactors that can carry out an outstanding range of chemical reactions. Some of the most interesting ones produce highly reactive and unstable species, such as hydroxide anions, hydrides, radical species, and superoxide molecules. The importance of these enzymes is enormous since they have paved the way to the emergence of the life as we know today. The formation of such reaction intermediates does not occur easily under biological conditions and only a rationalized selection of a particular metal ion, coordinated to a very specific group of ligands, allows these reactions to proceed.

The review of the catalytic mechanism of these enzymes highlighted that Zn, Mn, and Ni are involved in the formation of hydroxide as reaction intermediates. Zn is able to produce it through mononuclear metal centers in CA and HDAC8 enzymes, but also through a binuclear Zn–Zn center in the case of LeuAP. This last binuclear center likewise serves as a template to the Mn–Mn center of prolidase enzyme, which catalyzes an analogous reaction through a similar mechanism. Additionally, urease uses a binuclear metal center of Ni to produce a hydroxide intermediate involved in the urea degradation. (Table 1).

Enzymes like ADH can catalyze the production of hydride intermediates using binuclear Zn–Zn metal centers where one of the metal ions has a structural role and the other one has a catalytic character. The formation of hydride intermediates is also mediated by enzymes as XO and FdH that use Mo as a metal cofactor complexed with one or two pterin molecules. (Table 1).

In terms of radical formation, RNR enzymes can produce radical species through binuclear Fe–Fe or Mn–Mn centers accordingly if they belong to the Ia or Ib families, respectively. The MetH is an

enzyme that uses a Co ion coordinated to a cobalamin molecule and a histidine residue to help the production of a methyl radical intermediate (Table 1).

Finally, MIOX and DBH enzymes use binuclear Fe–Fe and Cu–Cu centers, respectively, to promote the generation of superoxide molecules inside the active site that act as crucial intermediates in the general catalytic mechanism (Table 1).

Table 1. Summary of the reviewed enzymes organized by their metal cofactor and respective produced intermediate.

	Catalysed Reaction			
Metal	**Hydroxide**	**Hydride**	**Radical**	**Superoxide**
Mn	Prolidase (Asp)**Mn-**(Asp)(Glu)(OH_2)-**Mn**(Glu)(His)		RNR-Ib (His)(Asp)**Mn**-*(Glu)*$_2$-**Mn**(His)(Glu)	
Fe			RNR-Ia (His)(Glu)**Fe**-*(Glu)*$_2$-**Fe**(His)(Glu)	MIOX (His)$_2$$(OH_2)$(Asp)**Fe**-*(OH)(Asp)*-**Fe**(His)$_2$
Co			MetH **Co**(Cobalamin)(His)	
Ni	Urease (His)$_2$$(OH_2)$**Ni-**(Lys*)$(OH_2)$-**Ni**(His)$_2$$(OH_2)$(Asp) *carbamylated			
Cu				DBH (His)$_2$(Met)**Cu**–**Cu**(His)$_3$
Zn	CA **Zn**(His)$_3$$(OH_2)$			
	HDAC8 **Zn**(Asp)$_2$(His)(OH_2)			
	LeuAP (Asp)**Zn-**(Asp)(Glu) (OH_2)-**Zn**(Lys)(Asp)	ADH (Cys)$_4$**Zn**✱–**Zn**(His)(Cys)$_2$$(OH_2)$		
Mo		XO **Mo**OS(OH)(Pterin)		
		FdH **Mo**S(Pterin)$_2$(Cys)		

✱ structural function.

A very important feature of these metalloenzymes is that the unstable and reactive reaction intermediates that are produced during catalysis remain coordinated to the metal ion until it reacts with another molecule and generates the final product of the reaction. This not only reduces the concentration of these species in the cells, which otherwise can be lethal due to their high toxicity, but also turns the catalytic process very efficient.

Several experimental and computational studies have already unraveled the outstanding biochemistry behind the processes that are catalyzed by these enzymes. In some cases, the same highly reactive and unstable species can be produced by enzymes containing very different metal coordination spheres, different in terms of the identity of the transition metal, number and nature of the ligands involved, and even number of metal atoms. This diversity ensures alternative strategies to generate the same chemical species through rather different mechanisms, illustrating the rich chemistry that nature can offer. Together, these studies have become very important because understanding the enzyme structure and the reaction pathways is a direct way to address the questions about the nature of the enzymatic power and enzyme evolution. This is very important for the development of new catalysts but also to understand how their function can be improved or inactivated. In the latter case, this knowledge can be used to develop new inhibitors for metalloenzymes that are currently drug targets for the treatment of several diseases.

Funding: This research was funded by European Union (FEDER funds POCI/01/0145/FEDER/007728) and National Funds (FCT/MEC, Fundação para a Ciência e Tecnologia and Ministério da Educação e Ciência) under the Partnership Agreement PT2020 UID/MULTI/04378/2013.UID/MULTI/04378/2013; NORTE-01-0145-FEDER-000024, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF); Fundação para a Ciência e a Tecnologia (FCT) through project IF/01310/2013 and IF/00052/2014. This work was supported by the Applied Molecular Biosciences Unit-UCIBIO which is financed by national funds from FCT/MCTES (UID/Multi/04378/2019). Henrique S. Fernandes and Carla S. Silva Teixeira were funded by Fundação para a Ciência e a Tecnologia (FCT) through PhD grants SFRH/BD/115396/2016 and SFRH/BD/114886/2016, respectively.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kozlowski, L.P. Proteome-pi: Proteome isoelectric point database. *Nucleic Acids Res.* **2016**, *45*, D1112–D1116. [CrossRef] [PubMed]
2. Zhang, Y.; Ying, H.; Xu, Y. Comparative genomics and metagenomics of the metallomes. *Metallomics* **2019**. [CrossRef] [PubMed]
3. Pernil, R.; Schleiff, E. Metalloproteins in the biology of heterocysts. *Life* **2019**, *9*. [CrossRef] [PubMed]
4. Hausinger, R.P. New metal cofactors and recent metallocofactor insights. *Curr. Opin. Struct. Biol.* **2019**, *59*, 1–8. [CrossRef] [PubMed]
5. Mendel, R.R.; Smith, A.G.; Marquet, A.; Warren, M.J. Metal and cofactor insertion. *Nat. Prod. Rep.* **2007**, *24*, 963–971. [CrossRef] [PubMed]
6. Ascone, I.; Fourme, R.; Hasnain, S.S. Introductory overview: X-ray absorption spectroscopy and structural genomics. *J. Synchrotron Radiat.* **2003**, *10*, 1–3. [CrossRef] [PubMed]
7. Cook, J.D.; Penner-Hahn, J.E.; Stemmler, T.L. Structure and dynamics of metalloproteins in live cells. *Methods Cell Biol.* **2008**, *90*, 199–216. [PubMed]
8. Malmstrom, B.G.; Neilands, J.B. Metalloproteins. *Annu. Rev. Biochem* **1964**, *33*, 331–354. [CrossRef] [PubMed]
9. Finkelstein, J. Metalloproteins. *Nature* **2009**, *460*, 813. [CrossRef] [PubMed]
10. Lombardi, A. Metalloproteins: Simple structure, complex function. *Nat. Chem. Biol.* **2015**, *11*, 760. [CrossRef] [PubMed]
11. Vallee, B.L.; Williams, R. Metalloenzymes: The entatic nature of their active sites. *Proc. Natl. Acad. Sci. USA* **1968**, *59*, 498. [CrossRef] [PubMed]
12. Karlin, K.D. Metalloenzymes, structural motifs, and inorganic models. *Science* **1993**, *261*, 701–708. [CrossRef] [PubMed]
13. Tan, X. *Metalloproteins and Metalloenzymes: Roles and Mechanisms of Metals in Functional Proteins*, 1st ed.; World Scientific Publishing: Shanghai, China, 2020.
14. Yu, F.; Cangelosi, V.M.; Zastrow, M.L.; Tegoni, M.; Plegaria, J.S.; Tebo, A.G.; Mocny, C.S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V.L. Protein design: Toward functional metalloenzymes. *Chem. Rev.* **2014**, *114*, 3495–3578. [CrossRef] [PubMed]
15. Yruela, I. Transition metals in plant photosynthesis. *Metallomics* **2013**, *5*, 1090–1109. [CrossRef] [PubMed]
16. Yano, J.; Yachandra, V. Mn4ca cluster in photosynthesis: Where and how water is oxidized to dioxygen. *Chem. Rev.* **2014**, *114*, 4175–4205. [CrossRef] [PubMed]
17. de Vries, S.; Dorner, K.; Strampraad, M.J.F.; Friedrich, T. Electron tunneling rates in respiratory complex i are tuned for efficient energy conversion. *Angew. Chem. (International ed in English)* **2015**, *54*, 2844–2848. [CrossRef] [PubMed]
18. Al-Attar, S.; de Vries, S. Energy transduction by respiratory metallo-enzymes: From molecular mechanism to cell physiology. *Coord. Chem. Rev.* **2013**, *257*, 64–80. [CrossRef]
19. Yamaguchi, K.; Shoji, M.; Isobe, H.; Yamanaka, S.; Kawakami, T.; Yamada, S.; Katouda, M.; Nakajima, T. Theory of chemical bonds in metalloenzymes xxi. Possible mechanisms of water oxidation in oxygen evolving complex of photosystem ii. *Mol. Phys.* **2018**, *116*, 717–745. [CrossRef]
20. Augustine, A.J.; Kjaergaard, C.; Qayyum, M.; Ziegler, L.; Kosman, D.J.; Hodgson, K.O.; Hedman, B.; Solomon, E.I. Systematic perturbation of the trinuclear copper cluster in the multicopper oxidases: The role of active site asymmetry in its reduction of O_2 to H_2O. *J. Am. Chem. Soc.* **2010**, *132*, 6057–6067. [CrossRef]

21. Solomon, E.I.; Augustine, A.J.; Yoon, J. O_2 reduction to H_2O by the multicopper oxidases. *Dalton Trans.* **2008**, 3921–3932. [CrossRef]
22. Rees, D.C.; Akif Tezcan, F.; Haynes, C.A.; Walton, M.Y.; Andrade, S.; Einsle, O.; Howard, J.B. Structural basis of biological nitrogen fixation. *Philos. Trans. A Math. Phys. Eng. Sci.* **2005**, *363*, 971–984; discussion 1035–1040. [CrossRef] [PubMed]
23. Valdez, C.E.; Smith, Q.A.; Nechay, M.R.; Alexandrova, A.N. Mysteries of metals in metalloenzymes. *Acc. Chem. Res.* **2014**, *47*, 3110–3117. [CrossRef] [PubMed]
24. Foster, A.W.; Osman, D.; Robinson, N.J. Metal preferences and metallation. *J. Biol. Chem.* **2014**, *289*, 28095–28103. [CrossRef] [PubMed]
25. Wei, W.J.; Qian, H.X.; Wang, W.J.; Liao, R.Z. Computational understanding of the selectivities in metalloenzymes. *Front. Chem.* **2018**, *6*, 638. [CrossRef] [PubMed]
26. Christianson, D.W. Structural chemistry and biology of manganese metalloenzymes. *Prog. Biophys. Mol. Biol.* **1997**, *67*, 217–252. [CrossRef]
27. Pain, D.; Dancis, A. Roles of Fe–S proteins: From cofactor synthesis to iron homeostasis to protein synthesis. *Curr. Opin. Genet. Dev.* **2016**, *38*, 45–51. [CrossRef] [PubMed]
28. Kobayashi, M.; Shimizu, S. Cobalt proteins. *Eur. J. Biochem.* **1999**, *261*, 1–9. [CrossRef] [PubMed]
29. Rubino, J.T.; Franz, K.J. Coordination chemistry of copper proteins: How nature handles a toxic cargo for essential function. *J. Inorg. Biochem.* **2012**, *107*, 129–143. [CrossRef] [PubMed]
30. McCall, K.A.; Huang, C.; Fierke, C.A. Function and mechanism of zinc metalloenzymes. *J. Nutr.* **2000**, *130*, 1437S–1446S. [CrossRef] [PubMed]
31. Piovesan, D.; Profiti, G.; Martelli, P.L.; Casadio, R. The human "magnesome": Detecting magnesium binding sites on human proteins. *BMC Bioinformatics* **2012**, *13* Suppl. 14, S10. [CrossRef]
32. Mendel, R.R. The molybdenum cofactor. *J. Biol. Chem.* **2013**, *288*, 13165–13172. [CrossRef] [PubMed]
33. Hille, R.; Schulzke, C.; Kirk, M.L. *Molybdenum and tungsten enzymes: Bioinorganic chemistry*, 1st ed.; The Royal Society of Chemistry: Cambridge, UK, 2016; Vol. 6.
34. Sundberg, R.J.; Martin, R.B. Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. *Chem. Rev.* **1974**, *74*, 471–517. [CrossRef]
35. Giles, N.M.; Watts, A.B.; Giles, G.I.; Fry, F.H.; Littlechild, J.A.; Jacob, C. Metal and redox modulation of cysteine protein function. *Chem. Biol.* **2003**, *10*, 677–693. [CrossRef]
36. Lamine, W.; Boughdiri, S.; Christ, L.; Morell, C.; Chermette, H. Coordination chemistry of Zn^{2+} with sal(ph)en ligands: Tetrahedral coordination or penta-coordination? A dft analysis. *J. Comput. Chem.* **2019**, *40*, 717–725.
37. Lamine, W.; Boughdiri, S.; Jeanneau, E.; Sanglar, C.; Morell, C.; Christ, L.; Chermette, H. Unexpected structure of a helical N_4-schiff-base Zn(ii) complex and its demetallation: Experimental and theoretical studies. *Chemphyschem* **2018**, *19*, 2938–2946. [CrossRef] [PubMed]
38. Ryan, K.S.; Drennan, C.L. Metalloenzymes in natural product biosynthetic pathways. *Nat. Prod. Rep.* **2018**, *35*, 612–614. [CrossRef] [PubMed]
39. Geissler, P.L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in liquid water. *Science* **2001**, *291*, 2121. [CrossRef] [PubMed]
40. Crichton, R.R. Chapter 12—zinc—lewis acid and gene regulator. In *Biological Inorganic Chemistry (Second Edition)*; Crichton, R.R., Ed.; Elsevier: Oxford, UK, 2012; pp. 229–246.
41. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **1997**, *74*, 1–20. [CrossRef]
42. Kannan, K.K.; Petef, M.; Fridborg, K.; Cid-Dresdner, H.; Lovgren, S. Structure and function of carbonic anhydrases. Imidazole binding to human carbonic anhydrase b and the mechanism of action of carbonic anhydrases. *FEBS Lett.* **1977**, *73*, 115–119. [PubMed]
43. Wang, W.; Zhang, Y.; Wang, L.; Jing, Q.; Wang, X.; Xi, X.; Zhao, X.; Wang, H. Molecular structure of thermostable and zinc-ion-binding gamma-class carbonic anhydrases. *BioMetals* **2019**, *32*, 317–328. [CrossRef] [PubMed]
44. Supuran, C.T.; Scozzafava, A.; Casini, A. Carbonic anhydrase inhibitors. *Med. Res. Rev.* **2003**, *23*, 146–189. [CrossRef] [PubMed]
45. Piazzetta, P.; Marino, T.; Russo, N. Theoretical investigation on the restoring step of the carbonic anhydrase catalytic cycle for natural and promiscuous substrates. *Arch. Biochem. Biophys.* **2015**, *582*, 101–106. [CrossRef] [PubMed]
46. Miscione, G.P.; Stenta, M.; Spinelli, D.; Anders, E.; Bottoni, A. New computational evidence for the catalytic mechanism of carbonic anhydrase. *Theor. Chem. Acc.* **2007**, *118*, 193–201. [CrossRef]

47. Liang, J.Y.; Lipscomb, W.N. Theoretical study of carbonic anhydrase-catalyzed hydration of CO_2: A brief review. *Int. J. Quantum Chem* **1989**, *36*, 299–312. [CrossRef]
48. Liang, J.Y.; Lipscomb, W.N. Binding of substrate CO_2 to the active site of human carbonic anhydrase ii: A molecular dynamics study. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3675–3679. [CrossRef] [PubMed]
49. Fisher, Z.; Hernandez Prada, J.A.; Tu, C.; Duda, D.; Yoshioka, C.; An, H.; Govindasamy, L.; Silverman, D.N.; McKenna, R. Structural and kinetic characterization of active-site histidine as a proton shuttle in catalysis by human carbonic anhydrase ii. *Biochemistry* **2005**, *44*, 1097–1105. [CrossRef] [PubMed]
50. Christianson, D.W.; Fierke, C.A. Carbonic anhydrase: Evolution of the zinc binding site by nature and by design. *Acc. Chem. Res.* **1996**, *29*, 331–339. [CrossRef]
51. Silverman, D.N.; Lindskog, S. The catalytic mechanism of carbonic anhydrase: Implications of a rate-limiting protolysis of water. *Acc. Chem. Res.* **2002**, *21*, 30–36. [CrossRef]
52. Liljas, A.; Kannan, K.K.; Bergsten, P.C.; Waara, I.; Fridborg, K.; Strandberg, B.; Carlbom, U.; Jarup, L.; Lovgren, S.; Petef, M. Crystal structure of human carbonic anhydrase c. *Nat. New Biol.* **1972**, *235*, 131–137. [CrossRef]
53. Kannan, K.K.; Liljas, A.; Waara, I.; Bergsten, P.C.; Lovgren, S.; Strandberg, B.; Bengtsson, U.; Carlbom, U.; Fridborg, K.; Jarup, L.; et al. Crystal structure of human erythrocyte carbonic anhydrase c. Vi. The three-dimensional structure at high resolution in relation to other mammalian carbonic anhydrases. *Cold Spring Harb. Symp. Quant. Biol.* **1972**, *36*, 221–231. [CrossRef]
54. Kimura, E. Model studies for molecular recognition of carbonic anhydrase and carboxypeptidase. *Acc. Chem. Res.* **2001**, *34*, 171–179. [CrossRef] [PubMed]
55. Huang, C.C.; Lesburg, C.A.; Kiefer, L.L.; Fierke, C.A.; Christianson, D.W. Reversal of the hydrogen bond to zinc ligand histidine-119 dramatically diminishes catalysis and enhances metal equilibration kinetics in carbonic anhydrase ii. *Biochemistry* **1996**, *35*, 3439–3446. [CrossRef] [PubMed]
56. Lesburg, C.A.; Huang, C.; Christianson, D.W.; Fierke, C.A. Histidine –> carboxamide ligand substitutions in the zinc binding site of carbonic anhydrase ii alter metal coordination geometry but retain catalytic activity. *Biochemistry* **1997**, *36*, 15780–15791. [CrossRef] [PubMed]
57. Zhang, X.; Hubbard, C.D.; van Eldik, R. Carbonic anhydrase catalysis: A volume profile analysis. *J. Phys. Chem.* **1996**, *100*, 9161–9171. [CrossRef]
58. Tu, C.; Tripp, B.C.; Ferry, J.G.; Silverman, D.N. Bicarbonate as a proton donor in catalysis by Zn(ii)- and Co(ii)-containing carbonic anhydrases. *J. Am. Chem. Soc.* **2001**, *123*, 5861–5866. [CrossRef]
59. Thoms, S. Hydrogen bonds and the catalytic mechanism of human carbonic anhydrase ii. *J. Theor. Biol.* **2002**, *215*, 399–404. [CrossRef] [PubMed]
60. Schröder, D.; Schwarz, H.; Schenk, S.; Anders, E. A gas-phase reaction as a functional model for the activation of carbon dioxide by carbonic anhydrase. *Angew. Chem. (International ed in English)* **2003**, *42*, 5087–5090. [CrossRef] [PubMed]
61. Liang, J.Y.; Lipscomb, W.N. Hydration of carbon dioxide by carbonic anhydrase: Internal protein transfer of Zinc$^{(2+)}$-bound bicarbonate. *Biochemistry* **2002**, *26*, 5293–5301. [CrossRef]
62. Merz, K.M.; Hoffmann, R.; Dewar, M.J.S. The mode of action of carbonic anhydrase. *J. Am. Chem. Soc.* **1989**, *111*, 5636–5649. [CrossRef]
63. Krauss, M.; Garmer, D.R. Active site ionicity and the mechanism of carbonic anhydrase. *J. Am. Chem. Soc.* **1991**, *113*, 6426–6435. [CrossRef]
64. Zheng, Y.J.; Merz, K.M. Mechanism of the human carbonic anhydrase ii-catalyzed hydration of carbon dioxide. *J. Am. Chem. Soc.* **1992**, *114*, 10498–10507. [CrossRef]
65. Sola, M.; Lledos, A.; Duran, M.; Bertran, J. Ab initio study of the hydration of carbon dioxide by carbonic anhydrase. A comparison between the lipscomb and lindskog mechanisms. *J. Am. Chem. Soc.* **1992**, *114*, 869–877. [CrossRef]
66. Sakurai, M.; Furuki, T.; Inoue, Y. The pka of the zinc-bound water in carbonic anhydrase and its model compounds as studied by the am1 calculation coupled with a reaction field theory. *J. Phys. Chem.* **1995**, *99*, 17789–17794. [CrossRef]
67. Merz, K.M.; Banci, L. Binding of bicarbonate to human carbonic anhydrase ii: A continuum of binding states. *J. Am. Chem. Soc.* **1997**, *119*, 863–871. [CrossRef]
68. Lu, D.; Voth, G.A. Proton transfer in the enzyme carbonic anhydrase: Anab initiostudy. *J. Am. Chem. Soc.* **1998**, *120*, 4006–4014. [CrossRef]

69. Toba, S.; Colombo, G.; Merz, K.M. Solvent dynamics and mechanism of proton transfer in human carbonic anhydrase ii. *J. Am. Chem. Soc.* **1999**, *121*, 2290–2302. [CrossRef]
70. Denisov, V.P.; Jonsson, B.-H.; Halle, B. Dynamics of functional water in the active site of native carbonic anhydrase from17o magnetic relaxation dispersion. *J. Am. Chem. Soc.* **1999**, *121*, 2327–2328. [CrossRef]
71. Mauksch, M.; Bräuer, M.; Weston, J.; Anders, E. New insights into the mechanistic details of the carbonic anhydrase cycle as derived from the model system $[(NH_3)_3Zn(OH)]^+/CO_2$: How does the H_2O/HCO_3^- replacement step occur? *ChemBioChem* **2001**, *2*, 190–198. [CrossRef]
72. Bräuer, M.; Pérez-Lustres, J.L.; Weston, J.; Anders, E. Quantitative reactivity model for the hydration of carbon dioxide by biomimetic zinc complexes. *Inorg. Chem.* **2002**, *41*, 1454–1463. [CrossRef]
73. Smedarchina, Z.; Siebrand, W.; Fernández-Ramos, A.; Cui, Q. Kinetic isotope effects for concerted multiple proton transfer: A direct dynamics study of an active-site model of carbonic anhydrase ii. *J. Am. Chem. Soc.* **2003**, *125*, 243–251. [CrossRef]
74. Garmer, D.R. Carbonic anhydrase reactivity, mutation, and inhibition probed with a model of ab initioquantum chemistry within a protein. *J. Phys. Chem. B* **1997**, *101*, 2945–2953. [CrossRef]
75. Cui, Q.; Karplus, M. Is a "proton wire" concerted or stepwise? A model study of proton transfer in carbonic anhydrase. *J. Phys. Chem. B* **2003**, *107*, 1071–1078. [CrossRef]
76. Muguruma, C. Ab initio Mo study on the catalytic mechanism in the active site of carbonic anhydrase. *J. Mol. Struct. THEOCHEM* **1999**, *461-462*, 439–452. [CrossRef]
77. Bottoni, A.; Lanza, C.Z.; Miscione, G.P.; Spinelli, D. New model for a theoretical density functional theory investigation of the mechanism of the carbonic anhydrase: How does the internal bicarbonate rearrangement occur? *J. Am. Chem. Soc.* **2004**, *126*, 1542–1550. [CrossRef] [PubMed]
78. Jeske, L.; Placzek, S.; Schomburg, I.; Chang, A.; Schomburg, D. Brenda in 2019: A european elixir core data resource. *Nucleic Acids Res.* **2019**, *47*, D542–D549. [CrossRef] [PubMed]
79. Amata, O.; Marino, T.; Russo, N.; Toscano, M. Catalytic activity of a zeta-class zinc and cadmium containing carbonic anhydrase. Compared work mechanisms. *PCCP* **2011**, *13*, 3468–3477. [CrossRef] [PubMed]
80. Nechay, M.R.; Gallup, N.M.; Morgenstern, A.; Smith, Q.A.; Eberhart, M.E.; Alexandrova, A.N. Histone deacetylase 8: Characterization of physiological divalent metal catalysis. *J. Phys. Chem. B* **2016**, *120*, 5884–5895. [CrossRef]
81. Banerjee, S.; Adhikari, N.; Amin, S.A.; Jha, T. Histone deacetylase 8 (HDAC8) and its inhibitors with selectivity to other isoforms: An overview. *Eur. J. Med. Chem.* **2019**, *164*, 214–240. [CrossRef]
82. Chen, K.; Zhang, X.X.; Wu, Y.D.; Wiest, O. Inhibition and mechanism of HDAC8 revisited. *J. Am. Chem. Soc.* **2014**, *136*, 11636–11643. [CrossRef]
83. Gantt, S.M.; Decroos, C.; Lee, M.S.; Gullett, L.E.; Bowman, C.M.; Christianson, D.W.; Fierke, C.A. General base-general acid catalysis in human histone deacetylase 8. *Biochemistry* **2016**, *55*, 820–832. [CrossRef]
84. Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y. A proton-shuttle reaction mechanism for histone deacetylase 8 and the catalytic role of metal ions. *J. Am. Chem. Soc.* **2010**, *132*, 9471–9479. [CrossRef] [PubMed]
85. Gantt, S.L.; Gattis, S.G.; Fierke, C.A. Catalytic activity and inhibition of human histone deacetylase 8 is dependent on the identity of the active site metal ion. *Biochemistry* **2006**, *45*, 6170–6178. [CrossRef] [PubMed]
86. Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc coordination spheres in protein structures. *Inorg. Chem.* **2013**, *52*, 10983–10991. [CrossRef] [PubMed]
87. Maret, W. New perspectives of zinc coordination environments in proteins. *J. Inorg. Biochem.* **2012**, *111*, 110–116. [CrossRef] [PubMed]
88. Maret, W.; Li, Y. Coordination dynamics of zinc in proteins. *Chem. Rev.* **2009**, *109*, 4682–4707. [CrossRef] [PubMed]
89. Alberto, M.E.; Leopoldini, M.; Russo, N. Can human prolidase enzyme use different metals for full catalytic activity? *Inorg. Chem.* **2011**, *50*, 3394–3403. [CrossRef] [PubMed]
90. Kitchener, R.L.; Grunden, A.M. Prolidase function in proline metabolism and its medical and biotechnological applications. *J. Appl. Microbiol.* **2012**, *113*, 233–247. [CrossRef]
91. Vanhoof, G.; Goossens, F.; De Meester, I.; Hendriks, D.; Scharpe, S. Proline motifs in peptides and their biological processing. *FASEB J.* **1995**, *9*, 736–744. [CrossRef]
92. Fujii, M.; Nagaoka, Y.; Imamura, S.; Shimizu, T. Purification and characterization of a prolidase from aureobacterium esteraromaticum. *Biosci. Biotechnol. Biochem.* **1996**, *60*, 1118–1122.

93. Wilk, P.; Uehlein, M.; Kalms, J.; Dobbek, H.; Mueller, U.; Weiss, M.S. Substrate specificity and reaction mechanism of human prolidase. *FEBS J.* **2017**, *284*, 2870–2885. [CrossRef]
94. Wilk, P.; Uehlein, M.; Piwowarczyk, R.; Dobbek, H.; Mueller, U.; Weiss, M.S. Structural basis for prolidase deficiency disease mechanisms. *FEBS J.* **2018**, *285*, 3422–3441. [CrossRef] [PubMed]
95. Metz, S.; Thiel, W. A combined qm/mm study on the reductive half-reaction of xanthine oxidase: Substrate orientation and mechanism. *J. Am. Chem. Soc.* **2009**, *131*, 14885–14902. [CrossRef] [PubMed]
96. Richter, A.M.; Lancaster, G.L.; Choy, F.Y.; Hechtman, P. Purification and characterization of activated human erythrocyte prolidase. *Biochem. Cell Biol.* **1989**, *67*, 34–41. [CrossRef] [PubMed]
97. Roberts, B.P.; Miller, B.R.; Roitberg, A.E.; Merz, K.M. Wide-open flaps are key to urease activity. *J. Am. Chem. Soc.* **2012**, *134*, 9934–9937. [CrossRef] [PubMed]
98. Krajewska, B.; van Eldik, R.; Brindell, M. Temperature- and pressure-dependent stopped-flow kinetic studies of jack bean urease. Implications for the catalytic mechanism. *J. Biol. Inorg. Chem.* **2012**, *17*, 1123–1134. [CrossRef] [PubMed]
99. Hameed, A.; Al-Rashida, M.; Uroos, M.; Qazi, S.U.; Naz, S.; Ishtiaq, M.; Khan, K.M. A patent update on therapeutic applications of urease inhibitors (2012–2018). *Expert Opin. Ther. Pat.* **2019**, *29*, 181–189. [CrossRef] [PubMed]
100. Carlsson, H.; Nordlander, E. Computational modeling of the mechanism of urease. *Bioinorg. Chem. Appl.* **2010**. [CrossRef]
101. Mazzei, L.; Cianci, M.; Benini, S.; Ciurli, S. The structure of the elusive urease-urea complex unveils the mechanism of a paradigmatic nickel-dependent enzyme. *Angew. Chem. (International ed. in English)* **2019**.
102. Pearson, M.A.; Park, I.S.; Schaller, R.A.; Michel, L.O.; Karplus, P.A.; Hausinger, R.P. Kinetic and structural characterization of urease active site variants. *Biochemistry* **2000**, *39*, 8575–8584. [CrossRef]
103. Suarez, D.; Diaz, N.; Merz, K.M., Jr. Ureases: Quantum chemical calculations on cluster models. *J. Am. Chem. Soc.* **2003**, *125*, 15324–15337. [CrossRef]
104. Estiu, G.; Merz, K.M., Jr. The hydrolysis of urea and the proficiency of urease. *J. Am. Chem. Soc.* **2004**, *126*, 6932–6944. [CrossRef]
105. Sanderink, G.J.; Artur, Y.; Siest, G. Human aminopeptidases: A review of the literature. *J. Clin. Chem. Clin. Biochem.* **1988**, *26*, 795–807. [CrossRef] [PubMed]
106. Matsui, M.; Fowler, J.H.; Walling, L.L. Leucine aminopeptidases: Diversity in structure and function. *Biol. Chem.* **2006**, *387*, 1535–1544. [CrossRef] [PubMed]
107. Taylor, A. Aminopeptidases: Structure and function. *FASEB J.* **1993**, *7*, 290–298. [CrossRef] [PubMed]
108. Lowther, W.T.; Matthews, B.W. Metalloaminopeptidases: Common functional themes in disparate structural surroundings. *Chem. Rev.* **2002**, *102*, 4581–4608. [CrossRef] [PubMed]
109. Liew, S.M.; Tay, S.T.; Puthucheary, S.D. Enzymatic and molecular characterisation of leucine aminopeptidase of burkholderia pseudomallei. *BMC Microbiol.* **2013**, *13*, 110.
110. Sträter, N.; Sun, L.; Kantrowitz, E.R.; Lipscomb, W.N. A bicarbonate ion as a general base in the mechanism of peptide hydrolysis by dizinc leucine aminopeptidase. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11151–11155. [CrossRef]
111. Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Counting the zinc-proteins encoded in the human genome. *J. Proteome Res.* **2006**, *5*, 196–201. [CrossRef]
112. Sousa, S.F.; Lopes, A.B.; Fernandes, P.A.; Ramos, M.J. The zinc proteome: A tale of stability and functionality. *Dalton Trans.* **2009**, 7946–7956. [CrossRef]
113. Zastrow, M.L.; Pecoraro, V.L. Designing hydrolytic zinc metalloenzymes. *Biochemistry* **2014**, *53*, 957–978. [CrossRef]
114. Zhu, X.; Barman, A.; Ozbil, M.; Zhang, T.; Li, S.; Prabhakar, R. Mechanism of peptide hydrolysis by co-catalytic metal centers containing leucine aminopeptidase enzyme: A DFT approach. *JBIC J. Biol. Inorg. Chem.* **2012**, *17*, 209–222. [CrossRef]
115. Zhang, T.; Ozbil, M.; Barman, A.; Paul, T.J.; Bora, R.P.; Prabhakar, R. Theoretical insights into the functioning of metallopeptidases and their synthetic analogues. *Acc. Chem. Res.* **2015**, *48*, 192–200. [CrossRef] [PubMed]
116. Allen, M.P.; Yamada, A.H.; Carpenter, F.H. Kinetic parameters of metal-substituted leucine aminopeptidase from bovine lens. *Biochemistry* **1983**, *22*, 3778–3783. [CrossRef] [PubMed]

117. Cappiello, M.; Alterio, V.; Amodeo, P.; Del Corso, A.; Scaloni, A.; Pedone, C.; Moschini, R.; De Donatis, G.M.; De Simone, G.; Mura, U. Metal ion substitution in the catalytic site greatly affects the binding of sulfhydryl-containing compounds to leucyl aminopeptidase. *Biochemistry* **2006**, *45*, 3226–3234. [CrossRef] [PubMed]
118. Kostić, D.A.; Dimitrijević, D.S.; Stojanović, G.S.; Palić, I.R.; Đorđević, A.S.; Ickovski, J.D. Xanthine oxidase: Isolation, assays of activity, and inhibition. *J. Chem.* **2015**, *2015*, 1–8. [CrossRef]
119. Cherak, S.J.; Turner, R.J. Assembly pathway of a bacterial complex iron sulfur molybdoenzyme. *Biomol. Concepts* **2017**, *8*, 155–167. [CrossRef] [PubMed]
120. Rudolph, M.J.; Johnson, J.L.; Rajagopalan, K.V.; Kisker, C. The 1.2 Å structure of the human sulfite oxidase cytochrome b5 domain. *Acta crystallographica. Section D* **2003**, *59*, 1183–1191. [CrossRef] [PubMed]
121. Hille, R.; Hall, J.; Basu, P. The mononuclear molybdenum enzymes. *Chem. Rev.* **2014**, *114*, 3963–4038. [CrossRef]
122. Gonzalez, P.J.; Rivas, M.G.; Mota, C.S.; Brondino, C.D.; Moura, I.; Moura, J.J.G. Periplasmic nitrate reductases and formate dehydrogenases: Biological control of the chemical properties of mo and w for fine tuning of reactivity, substrate specificity and metabolic role. *Coord. Chem. Rev.* **2013**, *257*, 315–331. [CrossRef]
123. Maia, L.B.; Moura, J.J.; Moura, I. Molybdenum and tungsten-dependent formate dehydrogenases. *J. Biol. Inorg. Chem.* **2015**, *20*, 287–309. [CrossRef]
124. Metz, S.; Thiel, W. Qm/mm studies of xanthine oxidase: Variations of cofactor, substrate, and active-site glu802. *J. Phys. Chem. B* **2010**, *114*, 1506–1517. [CrossRef]
125. Yamaguchi, Y.; Matsumura, T.; Ichida, K.; Okamoto, K.; Nishino, T. Human xanthine oxidase changes its substrate specificity to aldehyde oxidase type upon mutation of amino acid residues in the active site: Roles of active site residues in binding and activation of purine substrate. *J. Biochem.* **2007**, *141*, 513–524. [CrossRef] [PubMed]
126. Johnson-Winters, K.; Tollin, G.; Enemark, J.H. Elucidating the catalytic mechanism of sulfite oxidizing enzymes using structural, spectroscopic, and kinetic analyses. *Biochemistry* **2010**, *49*, 7242–7254. [CrossRef]
127. Cerqueira, N.M.F.S.A.; Pakhira, B.; Sarkar, S. Theoretical studies on mechanisms of some mo enzymes. *J. Biol. Inorg. Chem.* **2015**, *20*, 323–335. [CrossRef]
128. Cerqueira, N.M.F.S.A.; Gonzalez, P.J.; Fernandes, P.A.; Moura, J.J.G.; Ramos, M.J. Periplasmic nitrate reductase and formate dehydrogenase: Similar molecular architectures with very different enzymatic activities. *Acc. Chem. Res.* **2015**, *48*, 2875–2884. [CrossRef]
129. Hartmann, T.; Schrapers, P.; Utesch, T.; Nimtz, M.; Rippers, Y.; Dau, H.; Mroginski, M.A.; Haumann, M.; Leimkuhler, S. The molybdenum active site of formate dehydrogenase is capable of catalyzing c-h bond cleavage and oxygen atom transfer reactions. *Biochemistry* **2016**, *55*, 2381–2389. [CrossRef] [PubMed]
130. Niks, D.; Hille, R. Molybdenum- and tungsten-containing formate dehydrogenases and formylmethanofuran dehydrogenases: Structure, mechanism, and cofactor insertion. *Protein Sci.* **2019**, *28*, 111–122. [CrossRef] [PubMed]
131. Mota, C.S.; Rivas, M.G.; Brondino, C.D.; Moura, I.; Moura, J.J.G.; Gonzalez, P.J.; Cerqueira, N.M.F.S.A. The mechanism of formate oxidation by metal-dependent formate dehydrogenases. *J. Biol. Inorg. Chem.* **2011**, *16*, 1255–1268. [CrossRef]
132. Cerqueira, N.M.F.S.A.; Fernandes, P.A.; Gonzalez, P.J.; Moura, J.J.G.; Ramos, M.J. The sulfur shift: An activation mechanism for periplasmic nitrate reductase and formate dehydrogenase. *Inorg. Chem.* **2013**, *52*, 10766–10772. [CrossRef]
133. Coelho, C.; Gonzalez, P.J.; Moura, J.J.G.; Moura, I.; Trincao, J.; Romao, M.J. The crystal structure of cupriavidus necator nitrate reductase in oxidized and partially reduced states. *J. Mol. Biol.* **2011**, *408*, 932–948. [CrossRef]
134. Schrapers, P.; Hartmann, T.; Kositzki, R.; Dau, H.; Reschke, S.; Schulzke, C.; Leimkuhler, S.; Haumann, M. 'Sulfido and cysteine ligation changes at the molybdenum cofactor during substrate conversion by formate dehydrogenase (FDH) from rhodobacter capsulatus. *Inorg. Chem.* **2015**, *54*, 3260–3271. [CrossRef]
135. Robinson, W.E.; Bassegoda, A.; Reisner, E.; Hirst, J. Oxidation-state-dependent binding properties of the active site in a mo-containing formate dehydrogenase. *J. Am. Chem. Soc.* **2017**, *139*, 9927–9936. [CrossRef] [PubMed]
136. Raaijmakers, H.C.A.; Romao, M.J. Formate-reduced e-coli formate dehydrogenase h: The reinterpretation of the crystal structure suggests a new reaction mechanism. *J. Biol. Inorg. Chem.* **2006**, *11*, 849–854. [CrossRef] [PubMed]

137. Dong, G.; Ryde, U. Reaction mechanism of formate dehydrogenase studied by computational methods. *J. Biol. Inorg. Chem.* **2018**, *23*, 1243–1254. [CrossRef]
138. Axley, M.J.; Grahame, D.A. Kinetics for formate dehydrogenase of escherichia coli formate-hydrogenlyase. *J. Biol. Chem.* **1991**, *266*, 13731–13736. [PubMed]
139. Bergquist, C.; Storrie, H.; Koutcher, L.; Bridgewater, B.M.; Friesner, R.A.; Parkin, G. Factors influencing the thermodynamics of zinc alkoxide formation by alcoholysis of the terminal hydroxide complex, [tp(but,me)]znoh: An experimental and theoretical study relevant to the mechanism of action of liver alcohol dehydrogenase. *J. Am. Chem. Soc.* **2000**, *122*, 12651–12658. [CrossRef]
140. Kvassman, J.; Pettersson, G. Unified mechanism for proton-transfer reactions affecting the catalytic activity of liver alcohol-dehydrogenase. *Eur. J. Biochem.* **1980**, *103*, 565–575. [CrossRef] [PubMed]
141. Zhu, Y.H.; Liu, C.Y.; Cai, S.; Guo, L.B.; Kim, I.W.; Kalia, V.C.; Lee, J.K.; Zhang, Y.W. Cloning, expression and characterization of a highly active alcohol dehydrogenase for production of ethyl (S)-4-chloro-3-hydroxybutyrate. *Indian J. Med. Microbiol.* **2019**, *59*, 225–233. [CrossRef] [PubMed]
142. Kim, K.; Plapp, B.V. Substitution of cysteine-153 ligated to the catalytic zinc in yeast alcohol dehydrogenase with aspartic acid and analysis of mechanisms of related medium chain dehydrogenases. *Chem. Biol. Interact.* **2019**, *302*, 172–182. [CrossRef] [PubMed]
143. Moa, S.; Himo, F. Quantum chemical study of mechanism and stereoselectivity of secondary alcohol dehydrogenase. *J. Inorg. Biochem.* **2017**, *175*, 259–266. [CrossRef]
144. Agarwal, P.K.; Webb, S.P.; Hammes-Schiffer, S. Computational studies of the mechanism for proton and hydride transfer in liver alcohol dehydrogenase. *J. Am. Chem. Soc.* **2000**, *122*, 4803–4812. [CrossRef]
145. Hammes-Schiffer, S.; Benkovic, S.J. Relating protein motion to catalysis. *Annu. Rev. Biochem* **2006**, *75*, 519–541. [CrossRef] [PubMed]
146. Bogin, O.; Peretz, M.; Burstein, Y. Thermoanaerobacter brockii alcohol dehydrogenase: Characterization of the active site metal and its ligand amino acids. *Protein Sci.* **1997**, *6*, 450–458. [CrossRef] [PubMed]
147. Sjoberg, B.M. Biochemistry. A never-ending story. *Science* **2010**, *329*, 1475–1476. [CrossRef] [PubMed]
148. Cerqueira, N.M.; Fernandes, P.A.; Ramos, M.J. Ribonucleotide reductase: A critical enzyme for cancer chemotherapy and antiviral agents. *Recent Pat. Anticancer. Drug Discov.* **2007**, *2*, 11–29. [CrossRef] [PubMed]
149. Cerqueira, N.M.F.S.A.; Pereira, S.; Fernandes, P.A.; Ramos, M.J. Overview of ribonucleotide reductase inhibitors: An appealing target in anti-tumour therapy. *Curr. Med. Chem.* **2005**, *12*, 1283–1294. [CrossRef] [PubMed]
150. Cerqueira, N.M.F.S.A.; Femandes, P.A.; Eriksson, L.A.; Ramos, M.J. New insights into a critical biological control step of the mechanism of ribonucleotide reductase. *J. Mol. Struct.* **2004**, *709*, 53–65. [CrossRef]
151. Perez, M.A.S.; Cerqueira, N.M.F.S.A.; Fernandes, P.A.; Ramos, M.J. Ribonucleotide reductase: A mechanistic portrait of substrate analogues inhibitors. *Curr. Med. Chem.* **2010**, *17*, 2854–2872. [CrossRef]
152. Zhang, C.; Liu, G.; Huang, M. Ribonucleotide reductase metallocofactor: Assembly, maintenance and inhibition. *Front. Biol. (Beijing)* **2014**, *9*, 104–113. [CrossRef] [PubMed]
153. Bollinger, J.M., Jr.; Jiang, W.; Green, M.T.; Krebs, C. The manganese(iv)/iron(iii) cofactor of chlamydia trachomatis ribonucleotide reductase: Structure, assembly, radical initiation, and evolution. *Curr. Opin. Struct. Biol.* **2008**, *18*, 650–657. [CrossRef] [PubMed]
154. Cotruvo, J.A.; Stubbe, J. Escherichia coli class ib ribonucleotide reductase contains a dimanganese(iii)-tyrosyl radical cofactor in vivo. *Biochemistry* **2011**, *50*, 1672–1681. [CrossRef] [PubMed]
155. Rose, H.R.; Maggiolo, A.O.; McBride, M.J.; Palowitch, G.M.; Pandelia, M.E.; Davis, K.M.; Yennawar, N.H.; Boal, A.K. Structures of class id ribonucleotide reductase catalytic subunits reveal a minimal architecture for deoxynucleotide biosynthesis. *Biochemistry* **2019**, *58*, 1845–1860. [CrossRef] [PubMed]
156. Kolberg, M.; Strand, K.R.; Graff, P.; Andersson, K.K. Structure, function, and mechanism of ribonucleotide reductases. *Biochim. Biophys. Acta* **2004**, *1699*, 1–34. [CrossRef]
157. Eklund, H.; Uhlin, U.; Farnegardh, M.; Logan, D.T.; Nordlund, P. Structure and function of the radical enzyme ribonucleotide reductase. *Prog. Biophys. Mol. Biol.* **2001**, *77*, 177–268. [CrossRef]
158. Minnihan, E.C.; Nocera, D.G.; Stubbe, J. Reversible, long-range radical transfer in *E. Coli* class ia ribonucleotide reductase. *Acc. Chem. Res.* **2013**, *46*, 2524–2535. [CrossRef] [PubMed]
159. Hogbom, M.; Huque, Y.; Sjoberg, B.M.; Nordlund, P. Crystal structure of the di-iron/radical protein of ribonucleotide reductase from corynebacterium ammoniagenes. *Biochemistry* **2002**, *41*, 1381–1389. [CrossRef] [PubMed]

160. Merkx, M.; Kopp, D.A.; Sazinsky, M.H.; Blazyk, J.L.; Muller, J.; Lippard, S.J. Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: A tale of two irons and three proteins a list of abbreviations can be found in section 7. *Angew. Chem. (International ed. in English)* **2001**, *40*, 2782–2807. [CrossRef]
161. Buckel, W. Enzymatic reactions involving ketyls: From a chemical curiosity to a general biochemical mechanism. *Biochemistry* **2019**. [CrossRef] [PubMed]
162. Kashlan, O.B.; Scott, C.P.; Lear, J.D.; Cooperman, B.S. A comprehensive model for the allosteric regulation of mammalian ribonucleotide reductase. Functional consequences of ATP- and dATP-induced oligomerization of the large subunit. *Biochemistry* **2002**, *41*, 462–474. [CrossRef]
163. Ge, J.; Yu, G.; Ator, M.A.; Stubbe, J. Pre-steady-state and steady-state kinetic analysis of E. Coliclass i ribonucleotide reductase. *Biochemistry* **2003**, *42*, 10071–10083. [CrossRef]
164. Cotruvo, J.A.; Stich, T.A.; Britt, R.D.; Stubbe, J. Mechanism of assembly of the dimanganese-tyrosyl radical cofactor of class ib ribonucleotide reductase: Enzymatic generation of superoxide is required for tyrosine oxidation via a Mn(iii)Mn(iv) intermediate. *J. Am. Chem. Soc.* **2013**, *135*, 4027–4039. [CrossRef]
165. Boal, A.K.; Cotruvo, J.A., Jr.; Stubbe, J.; Rosenzweig, A.C. The dimanganese(ii) site of bacillus subtilis class ib ribonucleotide reductase. *Biochemistry* **2012**, *51*, 3861–3871. [CrossRef] [PubMed]
166. Banerjee, R.; Ragsdale, S.W. The many faces of vitamin b12: Catalysis by cobalamin-dependent enzymes. *Annu. Rev. Biochem.* **2003**, *72*, 209–247. [CrossRef] [PubMed]
167. Fernandes, H.S.; Ramos, M.J.; Cerqueira, N.M.F.S.A. Catalytic mechanism of the serine hydroxymethyltransferase: A computational oniom qm/mm study. *ACS Catal.* **2018**, *8*, 10096–10110. [CrossRef]
168. Bandarian, V.; Pattridge, K.A.; Lennon, B.W.; Huddler, D.P.; Matthews, R.G.; Ludwig, M.L. Domain alternation switches b(12)-dependent methionine synthase to the activation conformation. *Nat. Struct. Mol. Biol.* **2002**, *9*, 53–56. [CrossRef] [PubMed]
169. Drennan, C.L.; Huang, S.; Drummond, J.T.; Matthews, R.G.; Ludwig, M.L. How a protein binds b 12: A 3.0 å x-ray structure of b 12 -binding domains of methionine synthase. *Science* **1994**, *266*, 1669–1674. [CrossRef]
170. Drennan, C.L.; Matthews, R.G.; Ludwig, M.L. Cobalamin-dependent methionine synthase: The structure of a methylcobalamin-binding fragment and implications for other b12-dependent enzymes. *Curr. Opin. Struct. Biol.* **1994**, *4*, 919–929. [CrossRef]
171. Kumar, N.; Kozlowski, P.M. Mechanistic insights for formation of an organometallic Co-C bond in the methyl transfer reaction catalyzed by methionine synthase. *J. Phys. Chem. B* **2013**, *117*, 16044–16057. [CrossRef]
172. Smith, A.E.; Matthews, R.G. Protonation state of methyltetrahydrofolate in a binary complex with cobalamin-dependent methionine synthase. *Biochemistry* **2000**, *39*, 13880–13890. [CrossRef]
173. Jensen, K.P.; Ryde, U. Conversion of homocysteine to methionine by methionine synthase: A density functional study. *J. Am. Chem. Soc.* **2003**, *125*, 13970–13971. [CrossRef]
174. Kozlowski, P.M.; Kuta, J.; Galezowski, W. Reductive cleavage mechanism of methylcobalamin: Elementary steps of Co-C bond breaking. *J. Phys. Chem. B* **2007**, *111*, 7638–7645. [CrossRef]
175. Alfonso-Prieto, M.; Biarnes, X.; Kumar, M.; Rovira, C.; Kozlowski, P.M. Reductive cleavage mechanism of Co-C bond in cobalamin-dependent methionine synthase. *J. Phys. Chem. B* **2010**, *114*, 12965–12971. [CrossRef] [PubMed]
176. Ghosh, A.P.; Mamun, A.A.; Lodowski, P.; Jaworska, M.; Kozlowski, P.M. Mechanism of the photo-induced activation of CoC bond in methylcobalamin-dependent methionine synthase. *J. Photochem. Photobiol. B Biol.* **2018**, *189*, 306–317. [CrossRef] [PubMed]
177. Jarrett, J.T.; Amaratunga, M.; Drennan, C.L.; Scholten, J.D.; Sands, R.H.; Ludwig, M.L.; Matthews, R.G. Mutations in the b12-binding region of methionine synthase: How the protein controls methylcobalamin reactivity. *Biochemistry* **1996**, *35*, 2464–2475. [CrossRef] [PubMed]
178. Kapoor, A.; Shandilya, M.; Kundu, S. Structural insight of dopamine beta-hydroxylase, a drug target for complex traits, and functional significance of exonic single nucleotide polymorphisms. *PLoS ONE* **2011**, *6*, e26509. [CrossRef] [PubMed]
179. Vendelboe, T.V.; Harris, P.; Zhao, Y.; Walter, T.S.; Harlos, K.; El Omari, K.; Christensen, H.E. The crystal structure of human dopamine beta-hydroxylase at 2.9 a resolution. *Sci. Adv.* **2016**, *2*, e1500980. [CrossRef] [PubMed]

180. Cowley, R.E.; Tian, L.; Solomon, E.I. Mechanism of O_2 activation and substrate hydroxylation in noncoupled binuclear copper monooxygenases. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12035–12040. [CrossRef]
181. Wimalasena, K.; Dharmasena, S.; Wimalasena, D.S.; Hughbanks-Wheaton, D.K. Reduction of dopamine β-monooxygenase: A unified model for apparent negative cooperativity and fumarate activation. *J. Biol. Chem.* **1996**, *271*, 26032–26043. [CrossRef] [PubMed]
182. Klinman, J.P.; Krueger, M. Dopamine beta-hydroxylase: Activity and inhibition in the presence of beta-substituted phenethylamines. *Biochemistry* **1982**, *21*, 67–75. [CrossRef]
183. Sirimanne, S.R.; Herman, H.H.; May, S.W. Interaction of dopamine β-mono-oxygenase with substituted imidazoles and pyrazoles. Catalysis and inhibition. *Biochem. J.* **1987**, *242*, 227–233. [CrossRef]
184. Bollinger, J.M., Jr.; Diao, Y.; Matthews, M.L.; Xing, G.; Krebs, C. Myo-inositol oxygenase: A radical new pathway for O_2 and C-H activation at a nonheme diiron cluster. *Dalton Trans.* **2009**, 905–914. [CrossRef]
185. Hirao, H.; Morokuma, K. Insights into the (superoxo)Fe(iii)Fe(iii) intermediate and reaction mechanism of myo-inositol oxygenase: Dft and ONIOM(DFT:MM) study. *J. Am. Chem. Soc.* **2009**, *131*, 17206–17214. [CrossRef] [PubMed]
186. Brown, P.M.; Caradoc-Davies, T.T.; Dickson, J.M.J.; Cooper, G.J.S.; Loomes, K.M.; Baker, E.N. Crystal structure of a substrate complex of myo-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15032–15037. [CrossRef] [PubMed]
187. Arner, R.J.; Prabhu, K.S.T.; Thompson, J.R.; Hildenbrandt, G.D.; Liken, A.; Reddy, C.C. Myo-inositol oxygenase: Molecular cloning and expression of a unique enzyme that oxidizes myo-inositol and D-chiro-inositol. *Biochem. J.* **2001**, *360*. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

EXPERT OPINION ON THERAPEUTIC PATENTS, 2016
http://dx.doi.org/10.1080/13543776.2017.1254194

Taylor & Francis
Taylor & Francis Group

REVIEW

Amino acid deprivation using enzymes as a targeted therapy for cancer and viral infections

H. S. Fernandes, C. S. Silva Teixeira, P. A. Fernandes, M. J. Ramos and N. M. F. S. A. Cerqueira

UCIBIO-REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

ABSTRACT

Introduction: Amino acid depletion in the blood serum is currently being exploited and explored for therapies in tumors or viral infections that are auxotrophic for a certain amino acid or have a metabolic defect and cannot produce it. The success of these treatments is because normal cells remain unaltered since they are less demanding and/or can synthesize these compounds in sufficient amounts for their needs by other mechanisms.

Areas covered: This review is focused on amino acid depriving enzymes and their formulations that have been successfully used in the treatment of several types of cancer and viral infections. Particular attention will be given to the enzymes *L*-asparaginase, *L*-arginase, *L*-arginine deiminase, and *L*-methionine-γ-lyase.

Expert opinion: The immunogenicity and other toxic effects are perhaps the major limitations of these therapies, but they have been successfully decreased either through the expression of these enzymes from other organisms, recombination processes, pegylation of the selected enzymes or by specific mutations in the proteins. In 2006, FDA has already approved the use of *L*-asparaginase in the treatment of acute lymphoblastic leukemia. Other enzymes and in particular *L*-arginase, *L*-arginine deiminase, and *L*-methioninase have been showing promising results *in vitro* and *in vivo* studies.

ARTICLE HISTORY
Received 29 July 2016
Accepted 25 October 2016

KEYWORDS
Amino acid deprivation; *L*-asparaginase; *L*-methioninase; *L*-arginase; *L*-arginine deiminase; cancer; viral infections

1. Introduction

According to the World Health Organization, cancer figures among the leading causes of morbidity and mortality in the world, with approximately 14 million new cases and 8.2 million cancer-related deaths in 2014. In 2016, this trend has still not been reversed, and it is expected an increase of 70% of new cases during the next two decades [1].

During the last five decades, the field of oncology has witnessed a steady flow of exciting developments. A whole new range of molecules have been developed, which can be efficiently used to treat tumors, or significantly extend the survival of patients who previously had no effective treatment options. At the same time, the understanding of tumor biology is continuously evolving, allowing a better evaluation of tumor growth during and after treatments, which also has a direct impact on the efficacy of the treatment, and at the same time, improves the life quality of the patients.

Alterations to cellular metabolism seem to constitute a nearly universal feature in many types of cancers. For instance, many tumors exhibit deficiencies in their enzymatic armamentarium and cannot biosynthesize one or more amino acids that are essential for its development, survival, and proliferation (Figure 1(a)). In order to overcome these problems, tumors rely on the extracellular pool of these amino acids to satisfy protein biosynthesis demands and continue to grow without being affected (Figure 1(b)). This means that if the concentration of those amino acids is decreased, the development of tumor cells can be impaired or even annihilated. At the same time, normal cells would remain unchanged since they are less demanding or they can synthesize these compounds in sufficient amounts by alternative pathways (Figure 1(c)). This key difference between the nutritional needs of tumors and normal tissues creates a metabolic vulnerability of tumors that can be exploited to preclude their survival or proliferation. This vulnerability is the base of the amino acid depletion therapies.

Amino acid depletion therapy involves the degradation of a specific amino acid in the bloodstream and is fundamental for the survival and/or proliferation of a tumor (Figure 1(d–g)). This type of therapy has a long history in both clinical practice and experimental therapeutic settings but only now is gaining popularity because of the advances in protein engineering technology and on more sophisticated approaches that enable the study of genetic and metabolic differences between tumor and normal cells [2].

The studies made so far indicate that a simple diet restriction is not enough to achieve the required level of amino acid depletion for therapeutical purposes. The best results are obtained with the introduction of specific enzymes into the bloodstream of the patients that can decrease the concentration of a certain amino acid in the plasma (e.g. converting it into another molecule) that has been shown to be essential for the development of the tumor (Figure 1(d–g)). Unfortunately, the human genome does not encode enzymes with the pharmacological or catalytic requisites that are

CONTACT N. M. F. S. A. Cerqueira nscerque@fc.up.pt
© 2016 Informa UK Limited, trading as Taylor & Francis Group

Article highlights

- Amino acid deprivation therapy constitutes a valuable approach to treat cancer, infection and inflammation.
- FDA has already approved the use of L-asparaginase in anti-leukemic treatments in 2006 and 2011.
- *L*-arginine deiminase is used in the therapy of hepatocellular carcinomas and melanomas
- *L*-methioninase is a novel niche that is currently being explored in the treatment of breast, colon, lung and renal cancers.
- The main limitation of these therapies is the immunogenic effect of enzymes from bacterial sources.
- Several modifications to these enzymes have been conducted aiming to decrease these undesirable effects, such as the expression of these enzymes from other organisms, recombination or pegylation processes, or even by specific mutations on the active site of the enzymes.

This box summarizes key points contained in the article.

essential for the pretended therapeutic purpose. For this reason, heterologous enzymes (from other organisms), recombinant, and/or engineered human enzymes are used instead.

The potential of the amino acid depletion therapies is often illustrated by the treatment of childhood leukemia with the bacterial enzyme, L-asparaginase (Figure 1(g)). In 1950, this enzyme was used essentially as a monotherapy and could be successfully used to treat 5% of the diagnosed cases. However, nowadays, with the advent of multimodality chemotherapy regimens containing L-asparaginase, the cure rate for acute lymphoblastic leukemia (ALL) increased from 5% in the 1950s to 90% in the present days. Other successful examples include the treatment of metastatic melanomas with arginine (ARG)-depleting enzymes with low toxicity and high-efficacy results (Figure 1(e,f)). Methionine auxotrophy is also a novel niche, currently under exploration for targeting central nervous system cancers [3] and ALL (Figure 1(d)).

The success obtained so far with the amino acid depletion therapies is currently being exploited and explored in the treatment of several types of cancer that are auxotrophic for a certain type of amino acid. In addition, the same type of treatment is also being used in the treatment of viral infections (Figure 1(h)). Some viruses are also auxotrophic for some specific amino acids, and therefore, their depletion can be used to prevent the pathogens growth (Figure 1(i)).

In this review, the therapeutic potential, function, and the intellectual property related to the use of amino acid depriving enzymes and their formulations in the treatment of several types of cancer and other diseases will be revised. Particular attention will be given to the heterologous enzymes L-arginase, L-arginine deiminase (ADI), L-methioninase, and L-asparaginase taking into account their current therapeutic potential.

2. Amino acid depletion therapies

Humans and most mammals have long ago ceased the capability to produce 9 of the 20 amino acids that are needed for the biosynthesis of proteins. This means that they need to be obtained from the diet and/or microbiota. These amino acids are normally denominated as essential and include L-phenylalanine, L-valine, L-

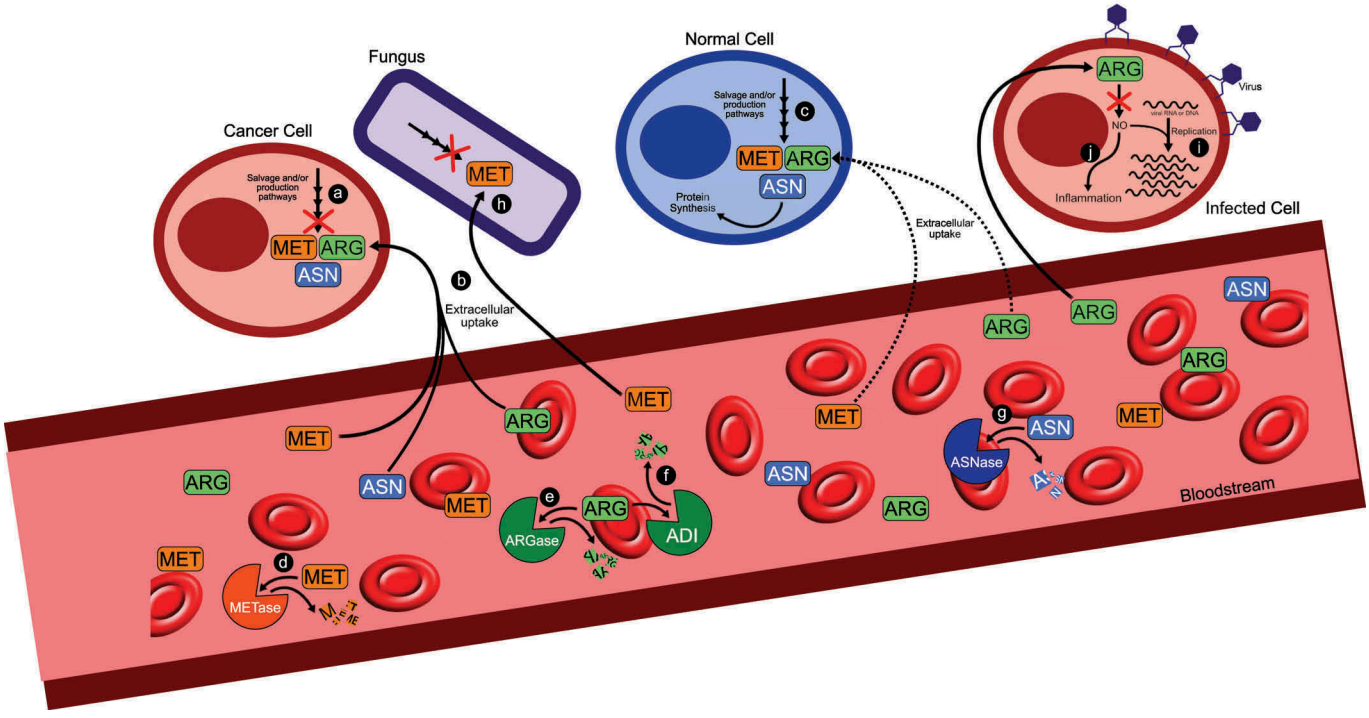


Figure 1. Schematic overview of the metabolic mechanisms involving the amino acid depletion therapies. Cancer cells (a) have some genetic modifications that compromise the *de novo* and/or salvage pathway synthesis of a key amino acid such as methionine (MET), arginine (ARG) or asparagine (ASN) and therefore require their uptake from the bloodstream (b). Normal cells are capable to synthesize endogenously MET, ARG, and ASN under normal conditions(c). The enzymes that have been tested for amino acid depletion therapy are methionine-γ-lyase (METase) (d), whose substrate is MET arginase (ARGase) (e) and L-arginine deiminase (ADI) (f) that metabolizes ARG, and asparaginase (ASNase) (g) that metabolizes ASN. Additionally, the depletion of MET can also be used as antifungal since fungus are auxotrophic for MET (h), and the depletion of ARG inhibits viral replication, also acting as anti-inflammatory (i).

threonine, L-tryptophan, L-methionine (MET), L-leucine, L-isoleucine, L-lysine, and L-histidine. Six other amino acids are termed conditionally essential; this means that their synthesis can be carried out by humans although limited by a variety of factors, such as prematurity in the infant or individuals in severe catabolic distress. These amino acids are L-arginine, L-cysteine, L-glycine, L-glutamine, L-proline, and L-tyrosine. The remaining five amino acids are dispensable from the human diet (also denoted as non-essential), meaning that human cells have the required enzymatic machinery to synthesize them, e.g. L-aspartic acid, L-asparagine (ASN), L-glutamic acid, L-alanine, and L-serine.

The optimal conditions for amino acid depletion therapies occur when the tumors are sensitive to one of the nonessential or conditionally essential amino acids. This will ensure that only tumor cells will be affected by the therapy whereas the normal cells remain unaltered since they can synthesize these compounds in sufficient amounts by their own means. The most notable and successful example of an amino acid depletion therapy is illustrated by the use of L-asparaginase (Figure 1(g)) in the treatment of childhood ALL and, to a lesser extent, non-Hodgkins lymphoma. Other examples involve the use of ADI (Figure 1(f)) in the hepatocellular carcinoma (HCC), melanoma, and other urea cycle-deficient cancer cells that have a high demand for L-arginine. The essential amino acids can also be used in amino acid depletion therapies, but it is not very common. This only happens when the normal metabolism of one essential amino acid in tumors is disrupted or when there is some sort of defect in the ability to use a certain amino acid. This is, for example, the case of MET that has been shown to be detrimental to the survival of a variety of tumor tissues, including colon, breast, prostate, ovary, lung, brain, kidney, stomach, and bladder cancers as well as larynx melanoma, sarcoma, leukemia, and lymphomas. In this case, the action of methionine-γ-lyase (Figure 1(d)) has been shown to be successfully used in the treatment of these types of cancers.

Taking into account the therapeutic potential of these enzymes in cancer therapies, they have been deeply studied in the last two decades. These therapies have been showing promising results, in spite of some side effects that were observed during clinical trials. In the following sections, a detailed analysis of each of these enzymes is done based on the amino acid that they metabolize.

2.1. Methionine

MET is a sulfur-containing amino acid that plays several important roles in the cell [4,5]. MET is crucial for protein synthesis, and it is also a precursor of several compounds, such as glutathione, polyamines spermine, and spermidine [6–9]. MET is also required for the synthesis of *S*-adenosylmethionine that it is essential in DNA methylation, and therefore, in genomic imprinting, X-chromosome inactivation, repression of repetitive elements, and carcinogenesis [10,11].

MET is a peculiar essential amino acid because its long-term deprivation does not inevitably compromise the life of the organism. Indeed, in normal cells, MET can be recycled by re-methylation of homocysteine (*de novo* synthesis), either by MET synthase or in the liver by betaine-homocysteine methyltransferase [6,12,13] (Figure 1(c)). In addition, MET can also be obtained by a salvage pathway in which it is generated from 5′-methylthioadenosine toward methylthioadenosine phosphorylase (Figure 1(c)). These two alternative sources of MET are available for normal cells from eukaryotes. However, numerous malignant cell lines, such as lung, breast, colon, kidney, and bladder cancers, melanoma, and glioblastoma [14,15] do not present an intact and functional mechanism capable of synthesizing MET (Figure 1(a)). In these conditions, cells only survive under an exogenous supply of MET (Figure 1(b)). As these cells divide more quickly than normal cells, their requirements for MET is very high [16,17], making them optimal for MET-depletion therapies.

Bacteria and fungi require a huge supply of MET (Figure 1(h)) which is required for the synthesis of *N*-formylmethionine (fMET) by the methionyl-tRNA formyltransferase [18,19]. fMET is required for the initiation of the protein synthesis in bacteria and fungus, and therefore, the depletion of MET can be used to impair their development or even to promote their destruction [20]. In addition, in these organisms, MET is required for the production of pathogenic agents, such as methanethiol [21], which turns MET depletion therapies very interesting for the development of new antibiotics.

During the last two decades, MET-restricted diets revealed a reasonable efficacy to prevent several diseases, such as aging [22–24], cancer [25,26], obesity, insulin resistance [27], and bacterial infection [21]. However, an MET-restricted diet implicates the consumption of a narrow range of nutrients that culminate in the emergence of some side effects, such as low bone density [27] and weight loss [28]. The usage of enzymes, like L-methionine-γ-lyase, has thus become a better alternative to efficiently decrease MET levels in the bloodstream of the patients having these diseases (Figure 1(d)).

L-methionine-γ-lyase (EC 4.4.1.11, METase), is an important enzyme involved in the catalysis of α- and γ-elimination of MET to α-ketobutyrate, ammonia, and methanethiol (Figure 2) [21]. METase is present intracellularly in bacteria and extracellularly in fungi, being absent in mammals [29]. Currently, the administration of this enzyme has been used as an attractive pharmacological approach to lower the MET levels, targeting the treatment of either cancer cells or other infections [30].

Indeed, the intravenous administration of recombinant and PEGylated form of METase was tested in phase I clinical trials. Patients in these trials experienced no significant toxicity, and MET levels significantly fell in plasma [31,32]. Nevertheless, the antitumor activity of this treatment was not tested, and to date, no other clinical trials have been published. Based on the results of these toxicological tests, some therapeutic applications have been assigned to METase, and they are given below.

2.1.1. Anticancer

It has been known for nearly a half century that human tumors, including those derived from the nervous system such as glioblastomas, medulloblastoma, and neuroblastomas, are much more sensitive to MET starvation than normal tissues.

Most studies on the antitumor action of MET depletion have used the METase from the potentially pathogen

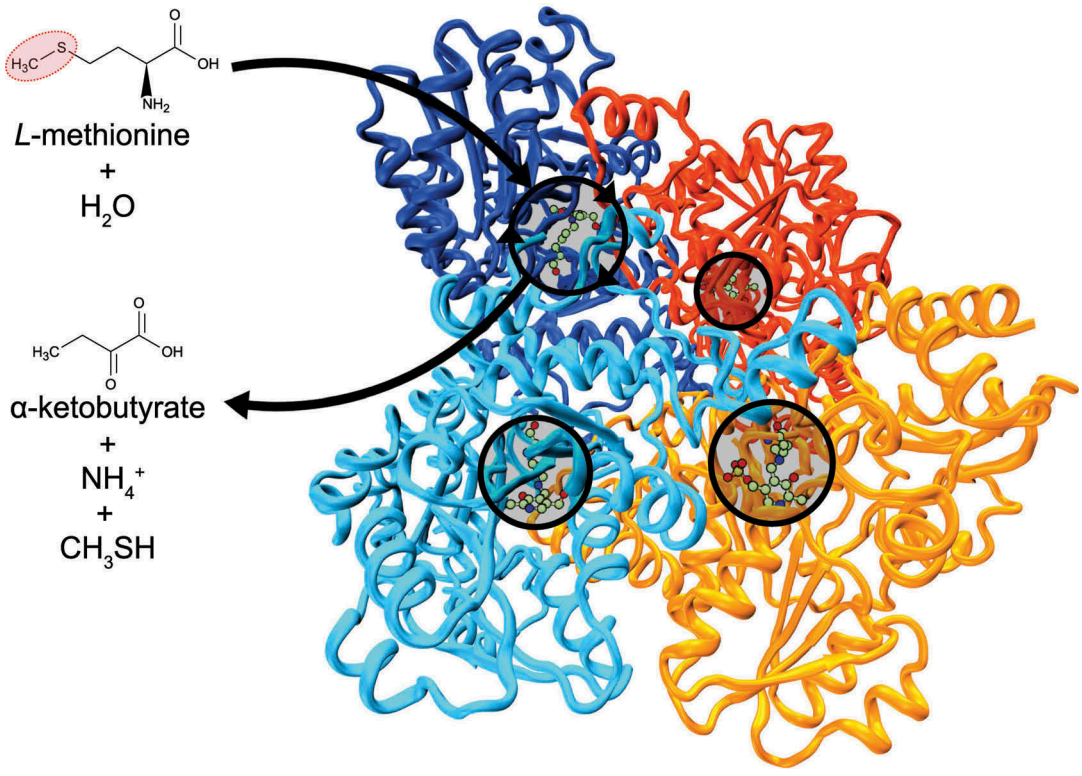


Figure 2. New Cartoon representation of the co-crystalized homotetramer structure of methionine-γ-lyase homotetramer with the PLP cofactor (PDB code: 2O7 C). The four subunits were represented with different colors (cyan, orange, blue, yellow). In the close-up of the active site, the PLP cofactor is depicted in ball and sticks. The reaction catalyzed by each active site is also represented with a highlight for the chemical group under catalytic attack. Full color available online.

pseudomonade, *Pseudomonas putida* (*Pp*-METase). Accordingly to patent US 6231854 B1, METase can be easily isolated and purified with less than a 1 ng of endotoxin per each mg of isolated, using a recombinant system, such as *Escherichia coli*, and the METase encoding gene from *Pseudomonas putida*. The purified recombinant METase could, then, be used *in vivo* in anticancer MET depletion therapies by direct intravascular administration. Besides the METase role in the cancer treatment, this enzyme could also be used to perform a global MET depletion in the organisms, followed by the radiolabeled MET administration in order to assess its distribution on the organism of patients and create a cancer location body map [28].

Various METases have also been produced from several microorganisms or by the recombination of genes encoding the enzyme originating from other bacterial species and protozoa. For example, the patent US 20140140978 A1 describes the isolation of a protein with METase and homocysteinase activities from *Brevibacterium aurantiacum*. This turns the MET deprivation in cancer therapies more efficient because homocysteine is a precursor of MET synthesis, enhancing, by this way, the desired therapeutic effect [33].

In spite of the success obtained using METase in several types of cancer in *in vitro* studies, its use *in vivo* reveals several problems. One of the main drawbacks of bacterial-derived METase is its poor stability in serum and highly immunogenicity in humans [34]. To overcome these issues, different strategies have been followed. One solution is presented in the patent US 6524571 B1 that describes a gene therapy in which the encoding gene of METase is inserted into cancer cells, and therefore, it is only produced inside the tumors. This presents several advantages since it leads to a local and more controlled MET deprivation and overcomes the instability of METase in the blood serum. In addition, the fusion of a fluorescent protein into the gene, such as GFP (Green Fluorescent Protein), can enable the control of the treatment [35]. The non-desirable immunogenicity of METase can be solved through the use of pegylated enzyme (5-kDa PEG). This posttranslational chemical modification to METase also increases the serum half-life of the enzyme, according to the registered patents US 7799549 B2, US 8465734 B2, US 20130252306 A1, US 20140205583 A1, US 6231854 B1, US 6461851 B1, and US 5891704 A [28,36–40].

Currently, a different pathway is being followed to overcome the disadvantages of METase from bacterial sources [41]. This includes the use of modified eukaryote enzymes with METase capabilities. The recent patent US 20150064159 A1 disclosed a primate modified cystathionine-γ-lyase, with amino acid substitution at positions 59, 63, 91, 119, 268, 331, 339, and/or 353 that induces a similar methionine-γ-lyase activity of the ones found on bacterial sources. This modified pegylated enzyme is currently being tested in several clinical trials in different types of cancers that have high requirements of MET. So far, the results are very promising since the immunogenicity effects and instability of the enzyme in the blood serum are not affecting the treatments [42].

2.1.2. Antiviral

The MET depletion therapies can also be expanded to the treatment of pneumonia, in particular to the one caused by *Pneumocystis carinii* (US 7264819 B2). *P. carinii* is a fungus that

causes pneumonia in patients in which the immune system is compromised, such as AIDS patients, patients undergoing chemotherapy, or patients being treated with immunosuppressants.

The use of METase is an interesting approach to treat this infection because *P. carinii* has a high MET exogenous dependency [43,44].

The patent US 7264819 B2 describes the successful use of METase in the treatment of this infection and proposes its administration with the conventional therapeutics in order to reduce their doses and minimize by this way the severe side effects from which results high mortality rates.

2.2. Arginine

ARG is a fundamental building block for protein synthesis in any organism, but in mammals, it is also an important precursor/mediator of a series of important biological pathways, such as nitrogen metabolism [45], creatine, agmatine, and polyamine synthesis [46], nitric oxide (NO) production [47], and anabolic hormone release. Additionally, it has immunostimulatory and thymotrophic effects [48,49]. ARG is classified as a semi or a conditionally essential amino acid. The term 'conditionally' is used because in healthy adults, the level of endogenous synthesis of ARG is sufficient to make it a non-essential amino acid [50]. However, under catabolic stress (e.g. inflammation, infection, etc.), the levels of endogenous synthesis may not be sufficient to ensure the metabolic demands of ARG, and in these cases, this amino acid becomes essential, requiring an exogenous supply.

The *de novo* biosynthetic pathway for ARG in mammals involves the conversion of citrulline to arginine catalyzed by argininosuccinate synthase (ASS) and argininosuccinate lyase (ASL) (Figure 3(a,b)). ASS catalyzes the condensation of citrulline and aspartate to form argininosuccinate (Figure 3(a)). ASL then splits argininosuccinate to release fumarate and ARG (Figure 3(a)). First identified in liver, as a rate-limiting enzyme in urea synthesis, ASS is now recognized as a ubiquitous enzyme in mammalian tissues [51].

In the last two decades, it was found that several types of tumors have a deficient expression of ASS (Figure 3(c)), a condition that causes ARG auxotrophy in those types of cancer. This deficiency was identified in a multitude of human cancers including metastatic melanoma, prostate carcinomas, HCC, cervical carcinoma, squamous cell carcinoma, breast carcinoma, ovarian carcinoma, prostate carcinoma, colon carcinoma, lung carcinoma, osteosarcoma, glioma/astrocytoma, glioblastoma, premyelocytic leukemia and lymphoblastic leukemia [52], non-Hodgkin's lymphoma, Hodgkin's lymphoma, pancreatic carcinoma, osteosarcoma, and malignant pleural mesothelioma [53].

Moreover, there are also tumors that have, additionally or alternatively, a deficient expression of ornithine transcarbamyl transferase (OCT) (Figure 3(d)), blocking the salvage pathway to ARG production from L-ornithine. It has been suggested for more than a decade that the lack of OTC expression was a typical feature of human epithelial cells of different organs of origin [54]. Later, Bobak and colleagues corroborated this hypothesis and reported negative OTC expression in several human cell lines, such as keratinocytic carcinoma A431, lung adenocarcinoma A549, HCC HepG2, breast adenocarcinoma MCF7, keratinocytes HaCaT, cervical carcinoma HeLa, ovarian carcinoma SKOV3, pancreatic carcinoma MIA PaCa-2, and melanomas MeWo, WM115, WM45 [55].

Deprivation of this amino acid is therefore investigated as a novel strategy for the treatment of these tumors [56–58].

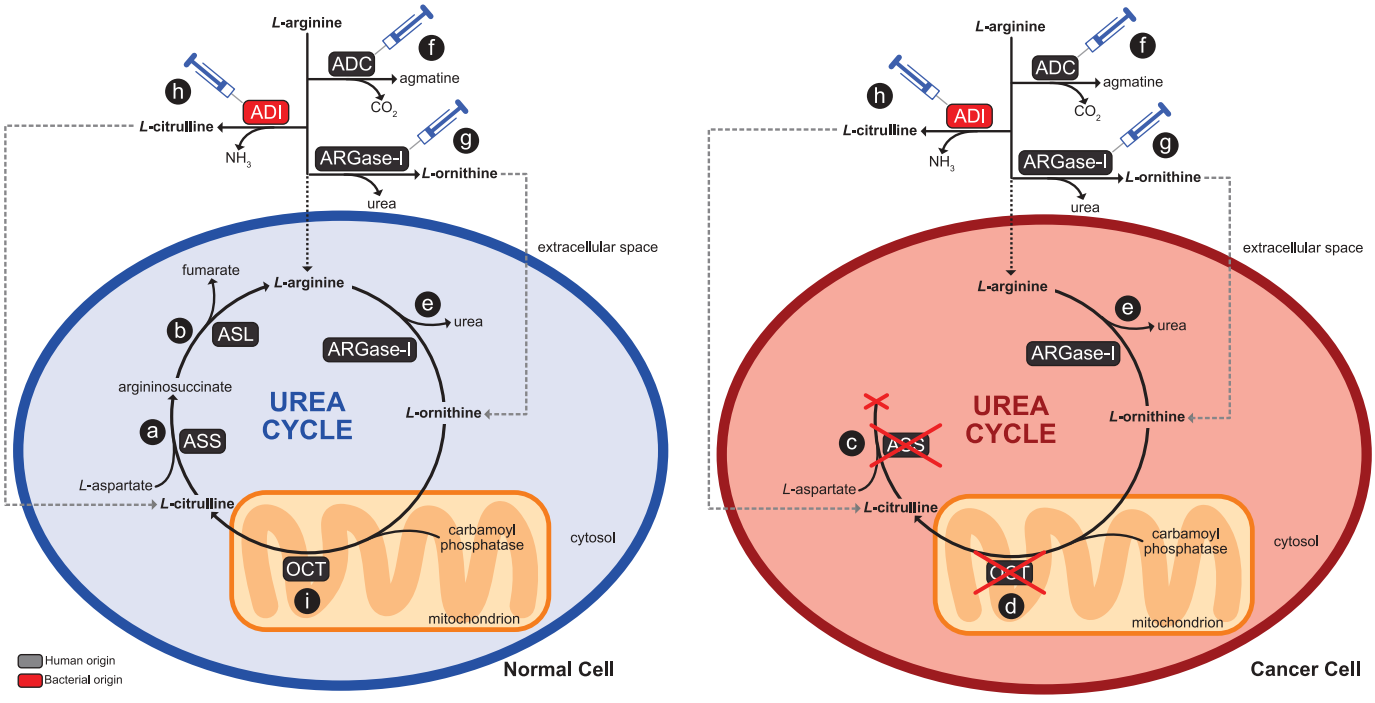


Figure 3. Schematic representation of the enzymes and metabolic pathways involved in ARG metabolism and degradation in normal cells in comparison with the genetically modified cancer cells. The enzymes with human origin are represented with a grey background while enzymes from bacterial sources are represented with red background. The enzymes engineered for amino acid depletion therapies are marked with a syringe. Full color available online.

The first step to implement an amino acid deprivation therapy is the choice of an enzyme that could metabolize ARG into another molecule, and at the same time, have optimal pharmacological characteristics and minimum side effects. ARG can be degraded by several mechanisms, either using enzymes available on the mammal cells, such as L-arginase (ARGase) (Figure 3(e)) or L-arginine decarboxylase (ADC) (Figure 3(f)), using human recombinant ARGase-I (Figure 3(g)), or using parasite–enzymes from a bacterial source such as ADI (Figure 3(h)) [59]. From these enzymes, ARGase-I and ADI are the most promising ones since they have therapeutic potential. ADC is relatively toxic to normal cells, and therefore, it is not suitable for therapeutic application [60].

ARG is also important in many viruses, where it has been shown to be essential for their replication [61–65]. The mechanism by which the ARG levels influence virus replication is still poorly understood and more studies are required [66–68]. However, several virus infections have been treated with ARG depletion therapies, using ARGase-I and ADI, with great success.

Despite the promising results in cancer and virus therapies, it must be noted that prolonged *in vivo* administration of arginine-depleting enzymes may impair NO synthesis by both constitutive and inducible nitric oxide synthase (NOS). Since NO participates in virtually every cellular and organ function in the body, it is the major endothelium-derived relaxing factor, a mediator of the immune response, a neurotransmitter, a cytotoxic free radical, and a widespread signaling molecule [69]. Its limitation may bring serious deleterious effects on the patient health.

2.2.1. Arginase

L-Arginase (EC 3.5.3.1, ARGase) is an enzyme that converts ARG into L-ornithine (Figure 4(a)). The isolation, purification, formulation, and production of a recombinant ARGase is described in the US 20100041101 A1 patent [70].

The use of this enzyme for amino acid depletion therapies in tumors (Figure 1: 1) or viral infections (Figure 1(i)) that have high demands of ARG would present many advantages. First, it is not immunogenic to humans and second, it has good serum stability. However, the overexpression of ARGase, increases considerably the concentration of L-ornithine in cells (product of the reaction of ARGase) (Figure 3(e)), which triggers a feedback mechanism in which of L-ornithine is converted back into L-citrulline by OCT (Figure 3(i)) and then recycle back to ARG by ASS/ASL (Figure 1(a,b)). This amino acid homeostatic mechanism keeps the various amino acid levels at constant ranges, and it is the main reason why normal cells and some tumors are resistant to therapies involving ARGase [71].

2.2.1.1. Anticancer.

The use of ARGase in the treatment of breast, rectal, and colon cancers has been shown with great success in *in vitro* studies; however, *in vivo* attempts were not always successful [56,72–75].

In order to improve ARGase activity *in vivo*, a pegylated variant of the enzyme was developed (US 8679479 B2 and US 20150010522 A1 patents) [76,77]. Other attempts to improve ARGase activity involved the mutation of a single amino acid in the enzyme structure as described in the US 20140242060 A1 patent [78]. However, the *in vivo* results indicate that no significant differences in activity of ARGase were obtained.

The amino acid homeostatic feedback mechanism is perhaps the main reason that limits the applicability of ARGase in the treatment of tumors in *in vivo* conditions. Several ways were developed in order to overcome this limitation. The US 6261557 B1 patent [79] describes a therapeutic composition in which ARGase is used combined with protein breakdown inhibitors, such as insulin, in order to prevent the muscles of the body from replenishing the depleted ARG. In the US 20080292609 A1 patent [80], it is described a combination therapy where the ARG depletion is achieved by a pegylated-ARGase administration together with an antineoplastic compound such as 5-fluorouracil or an alkylating agent [80,81]. In the WO 2014001956 A2 patent, a synergic therapy is described in which a recombinant ARGase with a mammalian target of rapamycin inhibitor, an alkylating agent, a mitotic inhibitor, or a small molecular inhibitor of tyrosine protein kinases and Raf kinases are used. The results of this combinatory therapy showed an interesting efficiency to treat liver and prostate cancers [82]. The patent CN 103402537 A describes the combination of a recombinant and pegylated-ARGase-I with doxorubicin in the treatment of patients suffering from lymphocytic leukemia and/or myeloid leukemia with great success [83].

A composition comprising an isolated human ARGase-I and a nonnative metal cofactor, wherein the nonnative metal cofactor is Co^{2+} was also patented [78]. The authors claimed that the substitution of the Mn^{2+} cofactor with Co^{2+} resulted in marked increase in catalytic activity and a drastic reduction in K_m at physiological pH.

The same author publicized a study showing that the substitution of the Mn(II) metal center in h-ARGase-I by Co(II) (Co-r-h-ARGase-I) results in an enzyme that displays 10-fold higher catalytic efficiency for ARG hydrolysis and a 12–15-fold reduction in the IC50 toward a variety of malignant cell lines, including melanomas and HCCs *in vitro* assays [84]. They also demonstrated that the modification of Co-r-h-ARGase-I with PEG-5K esters has shown an increase retention of the enzyme in circulation by about two orders of magnitude. The use of PEG(5K)-Co-h-ARGase-I was later on found to be effective in the treatment of hepatocellular and pancreatic carcinomas in xenograft models [85].

2.2.1.2. Antiviral.

ARG deprivation is also pointed as a therapeutic approach to treat infections, namely ocular lesions, hepatitis, and other inflammatory diseases.

In 2014, the eye lesions and/or infections gained a new type of treatment by the ocular administration of a pegylated-ARGase-I formulation in order to attenuate inflammation and neovascularization (US 8877183 B2 and WO 2014003850 A2). This treatment shows a good alternative to treat an eye that has been infected by a drug-resistant infectious agent [86,87]. Moreover, it can also be applied to an ocular trauma resulting from a medical procedure or from an accidentally caused injury (US 20130344049 A1) [86]. The attenuation of the inflammation is a consequence of the decrease of the NO production, which is a product of the NOS that uses ARG as

substrate. Consequently, if the ARG availability decreases, a lower amount of NO is produced, and the inflammatory process is attenuated (Figure 1(j)) [88].

The ARG-degrading enzyme, ARGase-I, can also be used to treat hepatitis B when it is used in its pegylated form, according to the patent WO 2006026915 A1. Actually, ARGase can be used in this type of treatment in the pegylated form in order to increase their serum half-life time up to 3 days. The pharmacological formulation could include a solution, a solid, an emulsion, a micelle, a dispersion, or a liposome, suitable for oral use or injection [89]. The usage of ARGase-I to treat hepatitis happens through a double-role mediated by the inhibition of the NO production similarly to the previous case. On one hand, the decrease of the NO levels attenuates the inflammation (Figure 1(j)), and on the other hand, the viral replication (Figure 1(i)) is also inhibited because viral replication requires NO to be successfully accomplished.

A worldwide patent WO 2015165374 A1 also describes an invention wherein a purified human pegylated-ARGase-I or a functional fragment is used in inflammatory diseases treatment, such as rheumatoid arthritis or multiple sclerosis. The procedure starts with an intravenous administration of the pharmacological formulation containing recombinant ARGase-I leading to a decrease of the ARG levels. The treatment can initiate an inhibition of T-cell polarization as well as modulation of cytokine release, namely IL-6 and IFN-γ. Furthermore, this treatment can also be taken by patients under immunosuppression in cases of cell, tissue or organ transplant. Osteoporosis associated with an osteoclast dysfunction was also reported as a healthy condition wherein the ARGase administration can be used as a treatment [90].

2.2.2. Arginine deiminase

ADI (EC 3.5.3.6) (Figure 4(b)) is an enzyme widely distributed among prokaryotic organisms and some anaerobic eukaryotes, which use it to convert ARG to L-citrulline and ammonia [91]. L-Citrulline is further degraded, forming ATP (adenosine triphosphate), carbon dioxide, and L-ornithine [92]. The so-called ADI pathway is used to generate energy, carbon, and nitrogen and also to protect some bacteria from acidic conditions, through the production of ammonia [93].

2.2.2.1. Anticancer.

The interest in ADI for therapeutic purposes began in the 90s when Miyazaki et al. observed a decrease in tumor development in Rous sarcoma virus-transformed buffalo–rat liver cells. They identified the mycoplasmal ADI as the growth inhibitory activity agent and suggested the possibility of its chemotherapeutic use for human cancers that have high demands of ARG such as hepatoma, cervix squamous cell carcinoma, and melanoma [94].

Soon after the identification of the ADI potential, a mycoplasmal proteinase K-resistant ADI, presenting optimum pH at physiological conditions and high stability, was developed and patented (US5372942 A). When applied to leukemia cell lines *in vitro*, this enzyme is 100% toxic and does not present any signs of toxicity to mice when administered intraperitoneally at much higher doses [95]. From there on, several studies, including the patent US5474928 A, confirmed the potential of ADI purified from *Mycoplasma arginine* (Figure 1(f)) as an antitumor agent in different types of human cancers, such as hepatoma, malignant fibrosarcoma, squamous cell carcinoma, malignant melanoma, nasopharyngeal carcinoma, and lung carcinoma [96–98].

Some years later, it was also found that beyond the antitumor activity, ADI has anti-angiogenic activity via suppression of NO generation [99–101]. In line with these studies, ADI was patented (US5474928 A) to be used as an active component of a pharmaceutical composition to inhibit angiogenesis. In that patent, the inventors use a pharmaceutical composition comprising ADI obtained from *M. arginini* or prepared by a genetic recombination technique, with an activated polymer to lower its immunogenicity and increase its life time [102].

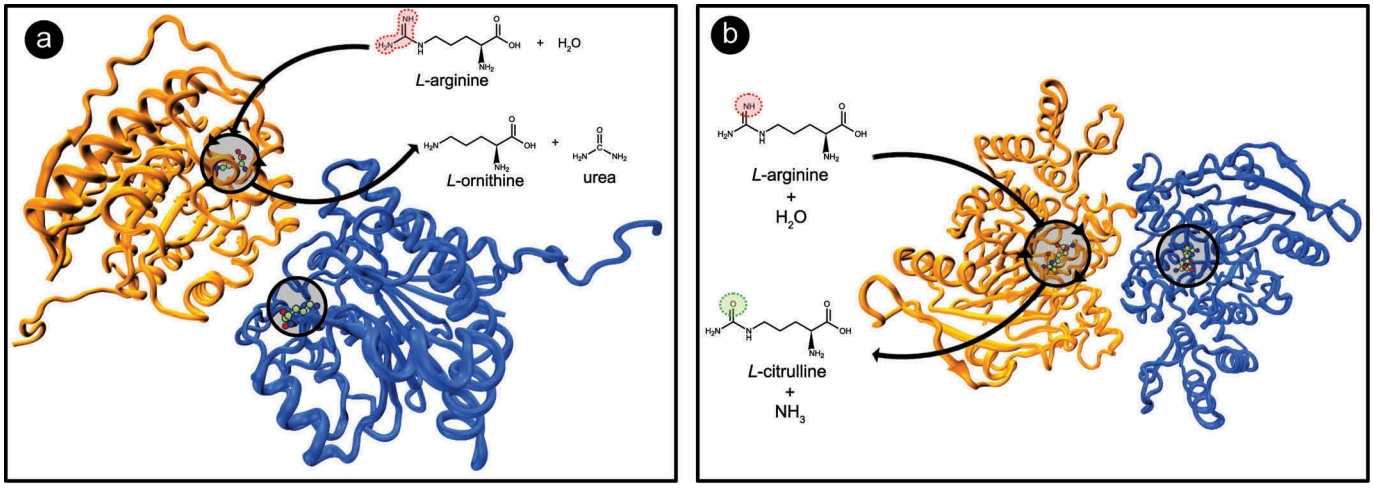


Figure 4. (a) New Cartoon representation of the co-crystalized structure of the human arginase I with L-ornithine (PDB code: 3GMZ). The two subunits were represented with different colors (orange and blue). In the close-up of the active site, it is shown, using the ball and sticks representation, the product L-ornithine. The reaction catalyzed by each active site is also represented with a highlight for the chemical group under catalytic attack. (b) New Cartoon representation of the co-crystalized structure of arginine deiminase from Mycoplasma arginini with a tetrahedral reaction intermediate (PDB Code: 1S9R).The four subunits were represented with different colors (cyan, orange, blue, yellow). In the close-up of the active site, it is shown, using the ball and sticks representation, the enzyme substrate, ASN. The reaction catalyzed by each active site is also represented with a highlight for the chemical group under catalytic attack. Full color available online.

The cooperative antiproliferative and antiangiogenic activities of ADI from *M. arginini* are believed today, to be the key assets that turn ADI very effective in the treatment of several tumors and other diseases. To date, cancers that present a sub-expression of ASS (Figure 3(c)) have already been shown to be sensitive to ARG deprivation via ADI, such as melanoma, lung cancer, renal cell carcinomas, and HCC [103].

In spite of the promising results, the application of ADI in the treatment of several cancers presents some limitations. The therapeutic efficiency of ADI is limited to tumors that do not express ASS (Figure 3(c)) and/or have an inactivated L-citrulline to ARG recycling pathway [104,105]. In addition, ADI has a short serum half-life and it is also highly immunogenic when used in prolonged treatments. In order to reduce this effect, several solutions were patented (US5372942 A; US5804183 A; US5916793 A; US6132713 A) proposing the attachment of ADI trough a covalent bond to a polyethylene glycol (pegylated-ADI) or similar water-soluble polymers [95,106–108]. In patent US6183738 B1, it was shown that pegylated-ADI is far less immunogenic, and has a greatly extended circulating half-life, and it is much more efficacious in the treatment of tumors [109]. Particular success has been obtained with the PEG(20)-ADI (molecular weight: 20 kDa) that has a remarkable efficiency *in vivo* assays with melanomas and HCC, comparing with native ADI [104,105,110–116]. The only limitation of ADI pegylated-modification is a partial decline in the activity of the modified proteins in relation to the native one, apparently because polyethylene glycol (PEG) attaches to sites on the enzyme, which interfere with the reaction that is catalyzed by the enzyme. Another problem is the formation of non-homogenous products during the pegylation process due to the random attachment of PEG on the protein surface. This makes the enzyme difficult to characterize and perform quality control during the manufacturing process. In order to overcome these issues, several modifications to pegylated-ADI have been proposed. First, in US6635462 B1 patent, it was proposed that one or more amino acid mutations in ADI structure could facilitate both the renaturation and formulation (i.e. pegylation) of the enzyme, thereby improving its manufacturing processes while retaining the ability to convert ARG to L-citrulline. These modified pegylated-ADI provided excellent results in the treatment of certain types of cancer, inhibiting the metastasis of cancer, and treating other disease states [117]. Good results have also been published in patent US8663967 B2 with the mutation of lysine residues present on the protein surface. The authors of this invention claim that when compared with pegylated-ADI, the modified pegylated-ADI has better bioactivity. Since the quantity of lysine in ADI is reduced, the PEG-modified products are more uniform and can be applied to the clinical treatment of various cancers such as leukemia, pancreatic, head and neck, colorectal, lung, breast, liver, nasopharyngeal, esophageal, prostate, stomach, and brain cancers [118].

Some studies have also shown that after the intravenous injection of pegylated-ADI *in vivo*, leakage or detachment of free PEG is observed and the ADI (without PEG) can elicit the immunogenicity problem. In an attempt to overcome this situation, in patent US9255262 B2, it is proposed the fusion of albumin to the ADI protein. This does not appear to influence the catalytic activity of the enzyme and seems to increase the circulating half-life, and at the same time, decrease the immunogenicity [119].

To date, the therapeutic efficacy of PEG(20)-ADI has already been validated through phase I/II clinical trials in advanced HCC [120–122] and melanoma patients [112,123,124]. Additionally, there are several ongoing phase I and/or II clinical trials using PEG20-ADI in patients with other tumors such as advanced non-Hodgkin's lymphoma, acute myeloid leukaemia, malignant pleural mesothelioma, prostate and non-small cell lung, advanced gastrointestinal, HER2 (Human Epidermal growth factor Receptor 2) negative metastatic breast, advanced pancreatic, and small-cell lung cancers.

2.2.2.2. Antiviral. In spite of some controversy, it is known that the elimination of extracellular ARG will inhibit the production of NO (Figure 1(i,j)), whose absence has been shown to protect mammals from the lethal effects of several viri such as influenza [125,126], polio virus [127], rabies virus [128], and flavivirus [129].

In this context, in patent US7204980 B2, it is shown that the administration of a therapeutically or prophylactically effective amount of a composition comprising modified pegylated-ADI inhibits Hepatitis C virus replication *in vitro*. Contrary to other existing inhibitors that preclude NO production and are toxic to animal and humans, ADI-PEG 20 is safer and more effective to inhibit NO production and can be used to gather a better role of this biomediator in protection against viral infections [130].

2.3. Asparaginase

ASN is an important amino acid that is involved in the synthesis of proteins. It also plays an important role in the biosynthesis of glycoproteins where it provides key sites for N-linked glycosylation and in the metabolic control of many cell functions in nerve and brain tissue.

ASN is a nonessential amino acid in humans, since it can be synthesized endogenously. The precursor in the biosynthesis of ASN is oxaloacetate, which reacts with L-glutamate to form L-aspartate and α-ketoglutarate. In the presence of a transaminase enzyme, L-aspartate then reacts with L-glutamine to form ASN in an ATP-dependent reaction that is catalyzed by ASN synthetase (EC 6.3.5.4, ASNS).

Certain types of tumors, such as leukemic cells, cannot synthesize ASN because they lack or have a deficient expression of the enzyme ASNS (Figure 1(a)). This means that the proliferation and survival of these tumor cells are dependent on the external supplies of this amino acid (Figure 1(b)). As normal cells can synthesize ASN (Figure 1(c)), this means that ASN deprivation therapies would mainly interfere with tumor cells that are ASN deficient.

L-Asparaginase (EC 3.5.1.1, ASNase) (Figure 5) is a relatively wide-spread enzyme found in many microorganisms (mostly in bacteria) that has been used with great success in leukemia treatment regimens since 1950. This enzyme performs the opposite reaction of ASNS, hydrolyzing ASN to L-aspartic acid

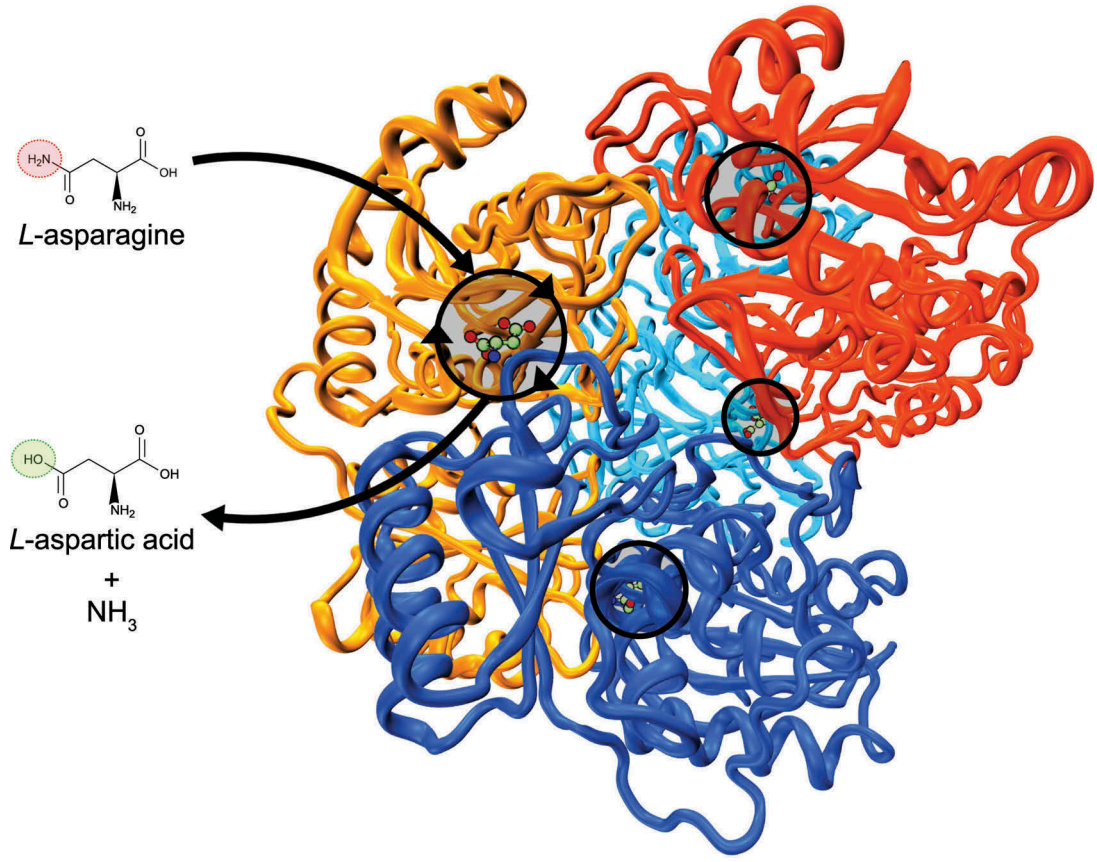


Figure 5. New Cartoon representation of the co-crystalized structure of asparaginase II from Escherichia coli with L-asparagine (PDB code: 3ECA). The two subunits were represented with different colors (orange and blue). In the close-up of the active site it is shown, using the ball and sticks representation, the product L-aspartic acid. The reaction catalyzed by each active site is also represented with a highlight for the chemical group under catalytic attack. Full color available online.

and ammonia [131]. By reducing the levels of plasmatic ASN (Figure 1(g)), ASNase starves leukemic cells and promotes tumor cell apoptosis. ASNase has also been proposed for the treatment of NK tumors, subtypes of myeloid leukemia, and T-cell lymphomas respond to ASNase, ovarian carcinomas, and other solid tumors [132]. ASNase has also been proposed in the treatment of AIDS.

2.3.1. Anticancer

ASNases derived from *E. coli* (Ecoli-ASNase) and *Erwinia chrysanthemi* (Echry-ASNase) are the ones that are commonly used in cancer treatments. For a large number of years, only the native Ecoli-ASNase (Elspar®, Lundbeck Inc.) was commercially available. This happened because this enzyme could be easily produced in large amounts as described in patents US3597323 A [133] and US3511755A [134], and efficiently used in the treatment of ALL. However, some patients showed hypersensitivity reactions and other toxicities (e.g. hepatic, renal, splenic, pancreatic dysfunction, and blood coagulation), which limited the application of Elspar® [135]. In 2006, Food and Drug Administration (FDA) approved the use of a pegylated form of ASNase from *E. coli* (Ecoli-PEG-ASNase), manufactured by Enzon Pharmaceuticals, Inc. (Oncaspar®) for the first-line treatment of patients with ALL. This new form of the ASNase (Oncaspar®) is much more stable and more efficient than Elspar®, requiring only a single dose to obtain a similar antileukemic activity of Elspar® that requires 6–9 doses of native Ecoli-ASNase.

Echry-ASNase is often used in the event of hypersensitivity to Ecoli-ASNase in the pegylated or not pegylated forms [136]. It is therefore used as a second- or third-line treatment protocol. In 2011, the FDA approved the usage of Echry-ASNase (Erwinaze, EUSA Pharma) as a component of a multi-agent chemotherapeutic regimen for the treatment of patients with ALL who have developed hypersensitivity to Ecoli-ASNase. A pegylated form of Echry-PEG-ASNase has also been claimed and it is described in patent WO 2011003886 A1 [137].

A number of comparative clinical studies have been conducted with Echry-ASNase and Ecoli-ASNase [138,139]. The dose and schedules of ASNase in these studies have been inconsistent and outcomes have been variable. However, the efficacy of Erwinia asparaginase following hypersensitivity to *E. coli*-asparaginase preparations has been well demonstrated [140]. The Echry-PEG-ASNase is currently under preclinical studies, and so far, it is showing promising results, similar to the ones obtained with Ecoli-PEG-ASNase but with less immunogenic effects [141,142].

In spite of the advantages promoted by the use of EColi-PEG-ASNase and Echry-ASNase, some patients continue to present adverse side effects, such as acute pancreatitis, thrombotic complications, and immunosuppression. This happens

because the main therapeutic effect of some forms of recombinant ASNases is also accompanied by glutamine deamination [143]. In an attempt to overcome this issue, new ASNases are currently being developed. The idea is to mutate some amino acids on the active site of ASNase in order to turn the enzyme more thermostable and specific to the substrate and, at the same time, preclude the reverse reaction that involves the conversion of L-glutamine into ASN [144,145]. One of such examples is described in the US 20130330316 A1 patent [146] where several mutations on the three-dimensional structure ASN isolated from *Pyrococcus furiosus* are conducted. Some of the mutated enzymes present high thermostability, pH stability, and no glutaminase activity that is important either for therapeutics or also for the process of preparing the same. These mutated enzymes also show high cytotoxicity on the leukemic cell lines making them promising candidates to help the treatment of leukemia [147].

2.3.2. Antiviral

The use of an EColi-PEG-ASN has also been proposed to inhibit or treat human immunodeficiency virus (HIV) infection. The patent WO 1999039732 A1 [148] describes a new pharmaceutically composition comprising a PEG-ASNase compound and optionally at least one compound selected from the group consisting of a protease inhibitor, a ribonucleotide reductase inhibitor compounds and an HIV reverse transcriptase inhibitor.

3. Conclusions

A prerequisite for making an effective medication for the treatment of cancer is that some fundamental difference between normal cells and cancer cells must be defined. An optimal chemotherapeutic agent must thus exploit this cellular difference in such a way that normal cells are spared and only cancer cells are injured. In this context, the amino acid depriving enzymes are promising anticancer drugs that have proven to be active and very specific against various types of cancers. Their mode of action is simple: they decrease the concentration of certain amino acids in the bloodstream and thus impair the development or even destroy tumor cells that are auxotrophic for those amino acids. Normal cells remain unaltered since they are less demanding and/or can synthesize these compounds in sufficient amounts by other mechanisms.

ASNase is probably the best known example of amino acid depletion chemotherapies using enzymes. The application of this enzyme in the treatment of childhood ALL has already been approved by FDA and shows very good results. ARG deprivation by ADI is a novel approach to target tumors that lack ASS expression. ADI showed promising results during initial clinical trials for the therapy of HCC and melanomas [149]. Some studies also suggest that ADI has potential to become a better therapeutic agent for the treatment of leukemia than ASNase [150,151].

Currently, the major drawback of ADI and ASNase when used for therapeutic purposes is their heterologous origin. As these enzymes are derived from a microorganism, they are immunogenic in humans and have a very short circulating half-life. The chemotherapeutic treatments using these enzymes thus require multiple injections to produce the desired effect and from which results in a variety of adverse effects including hypersensitivity reactions, anaphylactic shock, and the inactivation and clearance of the enzyme itself [152]. In order to overcome these problems and allow a safer application in cancer treatments, these enzymes have been linked to PEG to improve the pharmacokinetic and pharmacodynamic properties. The therapeutic results of these modified enzymes are much safer and have prompted the use of ADI in many clinical trials.

METase is another heterologous enzyme discussed in this review. Similarly to ASNase and ADI, this enzyme presents several immunogenic effects, but contrarily to those enzymes, the pegylated form does not seem to overcome this problem. Several attempts were made to decrease the toxicity of this enzyme by the coadministration of pyridoxal-5-phosphate and oleic acid or dithiothreitol [153,154]. However, several side effects persist and its utilization in cancer therapies continues to be inappropriate [155,156].

From the discussed enzymes, ARGase is the only enzyme that is from human origin. This means that it presents low immunogenicity, a great advantage over ADI in the treatment of cancer that is auxotrophic for ARG. Indeed, several authors consider ARGase a better chemotherapeutic agent for cancer treatment than ADI because, apart from its low immunogenicity, this enzyme is effective in tumors that are OCT-deficient and it is independent of ASS gene expression. Since most tumors are OCT-deficient, and ADI is ineffective in ASS-positive tumors, ARGase has a clear advantage over ADI to be used in therapeutic. Its major disadvantage when compared with ADI is its high K_m and high-optimum pH, which requires large quantities of enzyme to be included in the treatment in order to become effective. Actually this problem may be overcome by the utilization of cobalt-substituted ARGase, since the replacement of the two manganese ions at the active site by cobalt (II) leads to a 10-fold increase in enzymatic activity under biological conditions [84].

Despite the problems arising from the use of amino acid depleting, the recent progresses made in this field of research will undoubtedly facilitate the pursuit of new treatments to combat cancers and virus infections. ASNase was the first enzyme approved by FDA for the chemotherapeutic treatment of leukemia with remarkable success. Taking into account the on-going clinical trials made with other enzymes such as, ARGase or ADI, and the promising results that are being obtained, one might expect that the clinical approval of these enzymes should not be too distant in the future and will have undoubtedly a profound influence on human health.

4. Expert opinion

Cancer therapy has long relied on the rapid proliferation of tumor cells for an effective treatment. However, the lack of specificity in this approach often leads to undesirable side effects. Many reports have described various 'metabolic transformation' events that enable cancer cells to survive, suggesting that metabolic pathways might be good targets.

The amino acid depletion therapies using enzymes uses the uncommon metabolism of certain types of cancer that are auxotrophy for a certain amino acid or have a metabolic

defect and cannot produce it. This metabolic vulnerability has been used with great success to prevent the development or even the survival of some types of cancer. At the same time, the normal cells are spared because they are less demanding and/or can synthesize these compounds in sufficient amounts for their needs by other mechanisms.

The benefits of enzymatic therapies in amino acid depletion are well established in the literature and FDA has already approved the use of different ASNases in antileukemic treatments. Based on the ASNase paradigm, several ARG catabolizing enzymes have also been developed, including mycoplasma-derived ADI and human recombinant ARGase, for cancer therapy. Although the data are preliminary, early signs suggest that ARG deprivation has the potential to make an impact across a broad range of human malignancies. METase is another enzyme that is being efficiently employed in tumors that are autotrophic for methionine.

In spite of the success of amino acid depletion therapies using enzymes in auxotrophy tumors, they present many side effects, such as high immunogenic effects and other toxicities that preclude a generalized administration to patients. This is particularly evident in the treatments of tumors employing METase, ARGase, and ADI, but also in certain cases when ASNase is used.

Several modifications to these enzymes have been conducted aiming to decrease these undesirable effects, such as the expression of these enzymes from other organisms, recombination or pegylation processes, or even by specific mutations in the active site of the enzymes. In spite of the recent developments in the field, further drug optimizations are still needed to gain traction in the clinic. This is particularly true in the case of the enzymes are used to decrease the plasma levels of ARG and MET.

It is our opinion that the current limitations on the amino acid depletion therapies lie in the insufficient understanding about the pharmacokinetic and pharmacodynamic effects of these enzymes when they are applied to patients. Nowadays, this problem has been partially solved by the conjugation of PEG to the enzyme. However, while this approach can delay the formation of antibodies to the enzyme, it does not completely eliminate the eventual production of neutralizing antibodies, and it compromises enzyme activity. Furthermore, a precisely defined and reproducible conjugation of PEG to proteins is a laborious and expensive process. An alternative solution has been recently proposed through the use of fusion proteins containing albumin-binding proteins but more studies are required to test its effectiveness.

Despite all of these limitations of the amino acid therapies using enzymes, these treatments continue to be very attractive because it is less aggressive and more innocuous (apart of some immunogenicity associated with the enzymes that are not from human origin) when compared with other anticancer therapies. In addition, the promising results that are being obtained with the use of the same type of therapies in the treatment of several viral infections may turn them into the next antiviral generation of medications.

Funding

This work was supported by the European Union (FEDER funds POCI/01/0145/FEDER/007728) and National Funds (FCT/MEC, Fundação para a Ciência e a Tecnologia and Ministério da Educação e Ciência) under the Partnership Agreement PT2020 UID/MULTI/04378/2013.UID/MULTI/04378/2013; NORTE-01-0145-FEDER-000024, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF); Fundação para a Ciência e a Tecnologia (FCT) through project IF/01310/2013/CP1208/CT0002.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. McGuire S. World cancer report 2014. Geneva, Switzerland: World Health Organization, international agency for research on cancer, WHO Press, 2015. Adv Nutr. 2016;7(2):418–419.
2. Cantor JR, Panayiotou V, Agnello G, et al. Engineering reduced-immunogenicity enzymes for amino acid depletion therapy in cancer. In: Wittrup KD, Gregory LV, eds. Methods Enzymol. USA: Academic Press. 2012. p. 291–319.
3. Tan Y, Xu M, Hoffman RM. Broad selective efficacy of recombinant methioninase and polyethylene glycol-modified recombinant methioninase on cancer cells In Vitro. Anticancer Res. 2010;30(4):1041–1046.
4. Guedes RL, Prosdocimi F, Fernandes GR, et al. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. BMC Genomics. 2011;12(Suppl 4):S2.
5. Nozaki T, Ali V, Tokoro M. Sulfur-containing amino acid metabolism in parasitic protozoa. Adv Parasitol. 2005;60:1–99.
6. Cellarier E, Durando X, Vasson MP, et al. Methionine dependency and cancer treatment. Cancer Treat Rev. 2003;29(6):489–499.
7. Davis CD, Uthus EO. DNA methylation, cancer susceptibility, and nutrient interactions. Exp Biol Med (Maywood). 2004;229(10):988–995.
8. Ho E, Beaver LM, Williams DE, et al. Dietary factors and epigenetic regulation for prostate cancer prevention. Adv Nutr. 2011;2(6):497–510.
9. Thivat E, Durando X, Demidem A, et al. A methionine-free diet associated with nitrosourea treatment down-regulates methylguanine-DNA methyl transferase activity in patients with metastatic cancer. Anticancer Res. 2007;27(4C):2779–83.
10. Geiman TM, Muegge K. DNA methylation in early development. Mol Reprod Dev. 2010;77(2):105–113.
11. Ramani K, Yang H, Kuhlenkamp J, et al. Changes in the expression of methionine adenosyltransferase genes and S-adenosylmethionine homeostasis during hepatic stellate cell activation. Hepatology. 2010;51(3):986–995.
12. Guo HY, Herrera H, Groce A, et al. Expression of the biochemical defect of methionine dependence in fresh patient tumors in primary histoculture. Cancer Res. 1993;53(11):2479–2483.
13. Hoffman RM. Altered methionine metabolism, DNA methylation and oncogene expression in carcinogenesis. Biochimica et biophysica acta. 1984;738(1–2):49–87.
14. Breillout F, Antoine E, Poupon MF. Methionine dependency of malignant tumors: a possible approach for therapy. J Natl Cancer Inst. 1990;82(20):1628–1632.
15. Poirson-Bichat F, Gonçalves RA, Miccoli L, et al. Methionine depletion enhances the antitumoral efficacy of cytotoxic agents in drug-resistant human tumor xenografts. Clin Cancer Res. 2000;6(2):643–653.
16. Baylin SB, Herman JG, Graff JR, et al. Alterations in DNA methylation: a fundamental aspect of neoplasia. Adv Cancer Res. 1998;72:141–196.

17. Zingg JM, Jones PA. Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis. Carcinogenesis. 1997;18(5):869–882.
18. Gualerzi CO, Pon CL. Initiation of mRNA translation in prokaryotes. Biochemistry. 1990;29(25):5881–5889.
19. Stipanuk MH. Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. Annu Rev Nutr. 2004;24(1):539–577.
20. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999;234(2):187–208.
21. Sato D, Nozaki T. Methionine gamma-lyase: the unique reaction mechanism, physiological roles, and therapeutic applications against infectious diseases and cancers. IUBMB Life. 2009;61 (11):1019–1028.
22. Miller RA. Evaluating evidence for aging. Science. 2005;310 (5747):441–443. author reply 41-3.
23. Orentreich N, Matias JR, DeFelice A, et al. Low methionine ingestion by rats extends life span. J Nutr. 1993;123(2):269–274.
24. Richie JP, Leutzinger Y, Parthasarathy S, et al. Methionine restriction increases blood glutathione and longevity in F344 rats. Faseb J. 1994;8(15):1302–1307.
25. Cavuoto P, Fenech MF. A review of methionine dependency and the role of methionine restriction in cancer growth control and lifespan extension. Cancer Treat Rev. 2012;38(6):726–736.
26. Komninou D, Leutzinger Y, Reddy BS, et al. Methionine restriction inhibits colon carcinogenesis. Nutr Cancer. 2006;54(2):202–208.
27. Ables GP, Perrone CE, Orentreich D, et al. Methionine-restricted C57BL/6J mice are resistant to diet-induced obesity and insulin resistance but have low bone density. Plos One. 2012;7(12):e51357.
28. Yuying T, inventor; Methioninase formulations and use in diagnostic methods patent.US 6231854. 2001.
29. Sharma B, Singh S, Kanwar SS. L-methionase: a therapeutic enzyme to treat malignancies. Biomed Res Int. 2014;2014:506287.
30. Smiraglia DJ, Rush LJ, Fruhwald MC, et al. Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. Hum Mol Genet. 2001;10(13):1413–1419.
31. Tan Y, Zavala J Sr., Xu M, et al. Serum methionine depletion without side effects by methioninase in metastatic breast cancer patients. Anticancer Res. 1996 Nov–Dec;16(6c):3937–3942.
32. Tan Y, Zavala J Sr., Han Q, et al. Recombinant methioninase infusion reduces the biochemical endpoint of serum methionine with minimal toxicity in high-stage cancer patients. Anticancer Res. 1997; Sep-Oct;17(5b):3857–60.
33. Machover D, Bonnarme P, inventors; Polypeptides isolated from brevibacterium aurantiacum and their use for the treatment of cancer patent.US 20140140978. 2014.
• **US 6524,571: The patent describes a gene therapy approach that overcomes the stability problem associated with METase, and also allows a more specific targeting for the therapy.**
34. Kreis W, Hession C. Biological effects of enzymatic deprivation of l-methionine in cell-culture and an experimental tumor. Cancer Res. 1973;33(8):1866–1869.
35. Xu M, Tan Y, inventors; Methioninase gene therapy for tumor treatment patent.US 6524571. 2003.
36. Li S, Yang Z, Sun X, et al., inventors; Methods for increasing protein polyethylene glycol (PEG) conjugation patent.US7799549. 2010.
37. Li S, Yang Z, Sun X, et al., inventors; Methods for increasing protein polyethylene glycol (PEG) conjugation patent.US 8465734. 2013.
38. Tan Y, inventor; High expression modules containing twoor more tandem copies of a methioninase encoding sequence patent.US 6461851. 2002
39. Yagi S, Yang Z, Li S, et al., inventors; USE OF PLP WITH PEG-rMETase IN VIVO FOR ENHANCED EFFICACY patent.US 20140205583. 2014.
40. Yuying T, inventor; Method to produce high levels of methioninase patent.US 5891704 A. 1999.
•• **US 20,150,064,159: The patent proposes the use of an engineered form of methionine-γ-lyase in order to gain methioninase activity and solve the immunogenic effects of METase.**
41. Stone E, Paley O, Hu J, et al. De novo engineering of a human cystathionine-γ-lyase for systemic l-methionine depletion cancer therapy. ACS Chem Biol. 2012;7(11):1822–1829.
42. Georgiou G, Stone E, Lu WC, inventors; Engineered primate l-methioninase for therapeutic purposes patent.US 20150064159. 2015.
43. Merali S, Vargas D, Franklin M, et al. S-adenosylmethionine and pneumocystis carinii. J Biol Chem. 2000;275(20):14958–14963.
44. Tan Y, Yang Z, Sun X, et al., inventors; Lyase treatment for P. carinii patent.US 7264819. 2007.
45. Rogers QR, Visek WJ. Metabolic role of urea cycle intermediates: nutritional and clinical aspects. Introduction J Nutr. 1985;115 (4):505–508.
46. Klein D, Morris DR. Increased arginase activity during lymphocyte mitogenesis. Biochem Biophys Res Commun. 1978;81(1):199–204.
47. Billiar TR. Nitric oxide. Novel Biology with Clinical Relevance Ann Surg. 1995;221(4):339–349.
48. Barbul A, Rettura G, Levenson SM, et al. Arginine: a thymotropic and wound-healing promoting agent. Surg Forum. 1977;28:101–103.
49. Barbul A, Rettura G, Levenson SM, et al. Wound healing and thymotropic effects of arginine: a pituitary mechanism of action. Am J Clin Nutr. 1983;37(5):786–794.
50. Castillo L, Chapman TE, Sanchez M, et al. Plasma arginine and citrulline kinetics in adults given adequate and arginine-free diets. Proc Natl Acad Sci USA. 1993;90(16):7749–7753.
51. Haines RJ, Pendleton LC, Eichler DC. Argininosuccinate synthase: at the center of arginine metabolism. Int J Biochem Mol Biol. 2011;2 (1):8–23.
52. Dillon BJ, Prieto VG, Curley SA, et al. Incidence and distribution of argininosuccinate synthetase deficiency in human cancers. Cancer. 2004;100(4):826–833.
53. Qiu F, Huang J, Sui M. Targeting arginine metabolism pathway to treat arginine-dependent cancers. Cancer Lett. 2015 Aug 1;364 (1):1–7.
54. Wheatley DN. Arginine deprivation and metabolomics: important aspects of intermediary metabolism in relation to the differential sensitivity of normal and tumour cells. Semin Cancer Biol. 2005 Aug;15(4):247–253.
55. Bobak YP, Vynnytska BO, Kurlishchuk YV, et al. Cancer cell sensitivity to arginine deprivation in vitro is not determined by endogenous levels of arginine metabolic enzymes. Cell Biol Int. 2010;34 (11):1085–1089.
56. Feun LG, Kuo MT, Savaraj N. Arginine deprivation in cancer therapy. Curr Opin Clin Nutr Metab Care. 2015 Jan;18(1):78–82.
57. Shen L-J, Lin W-C, Beloussow K, et al. Resistance to the anti-proliferative activity of recombinant arginine deiminase in cell culture correlates with the endogenous enzyme, argininosuccinate synthetase. Cancer Lett. 2003;191(2):165–170.
58. Wheatley DN, Philip R, Campbell E. Arginine deprivation and tumour cell death: arginase and its inhibition. Mol Cell Biochem. 2003;244(1–2):177–185.
59. Schimke RT, Barile MF. Arginine metabolism in pleuropneumonia-like organisms isolated from mammalian cell culture. J Bacteriol. 1963;86:195–206.
60. Savaraj N, You M, Wu C, et al. Arginine deprivation, autophagy, apoptosis (AAA) for the treatment of melanoma. Curr Mol Med. 2010;10(4):405–412.
61. Archard LC, Williamson JD. The effect of arginine deprivation on the replication of vaccinia virus. J Gen Virol. 1971;12(3):249–258.
62. Goldblum N, Ravid Z, Becker Y. Effect of withdrawal of arginine and other amino acids on the synthesis of tumour and viral antigens of SV 40 virus. J Gen Virol. 1968;3(1):143–146.
63. Levine S, Buthala DA, Hamilton RD. Late stage synchronization of respiratory syncytial virus replication. Virology. 1971;45(2):390–400.
64. Minamishima Y, Benyesh-Melnick M. Arginine-dependent events in cytomegalovirus infection. Bacteriol Proc Abstr 69th Annu Meet Am Soc Microbiol. 1969;135:170.

65. Winters AL, Consigli RA, Rogers QR. A non-functional arginine biosynthetic pathway in polyoma infected-mouse embryo cells. Biochem Biophys Res Commun. 1972;47(5):1044–1050.
66. Akaike T, Maeda H. Nitric oxide and virus infection. Immunology. 2000;101(3):300–308.
67. Blond D, Raoul H, Le Grand R, et al. Nitric oxide synthesis enhances human immunodeficiency virus replication in primary human macrophages. J Virol. 2000;74(19):8904–8912.
68. Campbell IL, Samimi A, Chiang C-S. Expression of the inducible nitric oxide synthase. Correlation with neuropathology and clinical features in mice with lymphocytic choriomeningitis. J Immunol. 1994;153(8):3622–3629.
69. Ignarro LJ, Cirino G, Casini A, et al. Nitric oxide as a signaling molecule in the vascular system: an overview. J Cardiovasc Pharmacol. 1999;34(6):879–886.
70. Huang YL, Xian ZS, inventors; Expression system for recombinant human arginase i patent.US 20100041101. 2010.
71. Cheng PN-M, Lam T-L, Lam W-M, et al. Pegylated recombinant human arginase (rhArg-peg5,000mw) inhibits the in vitro and in vivo proliferation of human hepatocellular carcinoma through arginine depletion. Cancer Res. 2007;67(1):309–317.
72. Scott L, Lamb J, Smith S, et al. Single amino acid (arginine) deprivation: rapid and selective death of cultured transformed and malignant cells. Br J Cancer. 2000;83(6):800–810.
73. Lam T-L, Wong GKY, Chow H-Y, et al. Recombinant human arginase inhibits the in vitro and in vivo proliferation of human melanoma by inducing cell cycle arrest and apoptosis. Pigment Cell Melanoma Res. 2011;24(2):366–376.
• **US 8679,479: Describes the pegylation of ARGase in order to increase the serum half-life time.**
74. Hsueh EC, Knebel SM, Lo W-H, et al. Deprivation of arginine by recombinant human arginase in prostate cancer cells. J Hematol Oncol. 2012;5(1):1–6.
•• **WO 2014,001,956 describes the combined therapy using ARGase and others antitumor drugs that are commonly used.**
75. Li L, Wang Y, Chen J, et al. An engineered arginase FC protein inhibits tumor growth in vitro and in vivo. Evid Based Complement Alternat Med. 2013;2013:423129.
76. Leung YC, Lo W-H, inventors; Site-directed pegylation of arginases and the use thereof as anti-cancer and anti-viral agents patent.US 20140023628. 2014.
77. Georgiou G, Stone E, inventors; Engineered enzymes with methionine-gamma-lyase enzymes and pharmacological preparations thereof patent.US 8709407. 2014.
78. Georgiou G, Stone E, inventors; Compositions of engineered human arginases and methods for treating cancer patent.US 8440184. 2013.
79. Tepic S, Pyk P, inventors; Arginine decomposing enzyme therapeutic composition patent.US 6261557. 2001.
80. Cheng NM, Leung YC, Lo WH, inventors; Use of arginase in combination with 5FU and other compounds for treatment of human malignancies patent.US 20080292609. 2008.
81. Georgiou G, Stone E, inventor Methods for purifying pegylated arginase patent.US 8679479. 2014.
82. Cheng NM, Chen L, inventors; Utilisation combinatoire d'arginase humaine recombinante pégylée avec un médicament chimiothérapeutique/thérapeutique cible dans le traitement du cancer patent.WO 2014001956A3. 2014.
83. Cheng NM, inventor; 聚乙二醇化重组人精氨酸酶治疗白血病的应用patent.CN 103402537 A. 2013.
84. Stone EM, Glazer ES, Chantranupong L, et al. Replacing Mn(2+) with Co(2+) in human arginase i enhances cytotoxicity toward l-arginine auxotrophic cancer cell lines. ACS Chem Biol. 2010;5(3):333–342.
85. Glazer ES, Stone EM, Zhu C, et al. Bioengineered human arginase I with enhanced activity and stability controls hepatocellular and pancreatic carcinoma xenografts. Transl Oncol. 2011;4(3):138–146.
86. Foster TP, Rodriguez PC, Hill JM, et al., inventors; Methods for treatment of ocular diseases patent.US 20130344049. 2013.
87. Foster TP, Rodriguez PC, Hill JM, et al., inventors; Methods for treatment of ocular diseases patent.WO 2014003850 A2. 2014.
88. Knowles RG, Moncada S. Nitric oxide synthases in mammals. Biochem J. 1994;298(Pt 2):249–258.
89. Cheng NM, inventor; Pharmaceutical composition and method of treating hepatitis with arginases patent.WO 2006026915 A1. 2006.
90. Schabbauer G, Bluml S, Sahin E, et al., inventors; Methods and Compositions for Modulating the Immune System with Arginase I patent.US 20150315561. 2015.
91. Horn F. The breakdown of arginine to citrulline by Bacillus pyocyaneus. Hoppe Seylers Z Physiol Chem. 1933;216:244–247.
92. Arena ME, Saguir FM. Manca de Nadra MC. Arginine dihydrolase pathway in Lactobacillus plantarum from orange. Int J Food Microbiol. 1999;47(3):203–209.
93. Casiano-Colon A, Marquis RE. Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. Appl Environ Microbiol. 1988;54(6):1318–1324.
94. Miyazaki K, Takaku H, Umeda M, et al. Potent growth inhibition of human tumor cells in culture by arginine deiminase purified from a culture medium of a Mycoplasma-infected cell line. Cancer Res. 1990;50(15):4522–4527.
95. McGarrity GJ, Butler GH, inventors; Protease K resistant arginine deiminase, its method of preparation and its use as an anti-neoplastic agent patent.US 5372942. 1994
96. Ashikaga T, Ashikaga T, Wang Z, et al. Development and characterization of macrophage hybridomas derived from murine peritoneal exudate cells. Biosci Biotech Bioch. 2014;58(5):839–842.
97. Takaku H, Takase M, Abe S, et al. In vivo anti-tumor activity of arginine deiminase purified from Mycoplasma arginini. Int J Cancer Suppl. 1992;51(2):244–249.
98. Sugimura K, Ohno T, Kusuyama T, et al. High sensitivity of human melanoma cell lines to the growth inhibitory activity of mycoplasmal arginine deiminase in vitro. Melanoma Res. 1992;2(3):191–196.
99. Beloussow K, Wang L, Wu J, et al. Recombinant arginine deiminase as a potential anti-angiogenic agent. Cancer Lett. 2002;183(2):155–162.
100. Park IS, Kang SW, Shin YJ, et al. Arginine deiminase: a potential inhibitor of angiogenesis and tumour growth. Br J Cancer. 2003;89(5):907–914.
101. Yoon CY, Yoon C-Y, Shim YJ, et al. Renal cell carcinoma does not express argininosuccinate synthetase and is highly sensitive to arginine deprivation viaarginine deiminase. Int J Cancer Suppl. 2007;120(4):897–905.
102. Min B-H, Park M-O, Kim M-Y, et al., inventors; Pharmaceutical composition comprising arginine deiminase for inhibiting angiogenesis patent.US 7413735. 2008.
103. Liu J, Ma J, Wu Z, et al. Arginine deiminase augments the chemosensitivity of argininosuccinate synthetase-deficient pancreatic cancer cells to gemcitabine via inhibition of NF-kappaB signaling. BMC Cancer. 2014;14:686.
104. Ensor CM, Holtsberg FW, Bomalaski JS, et al. Pegylated arginine deiminase (ADI-SS PEG20,000 mw) inhibits human melanomas and hepatocellular carcinomas in vitro and in vivo. Cancer Res. 2002;62(19):5443–5450.
105. Qiu F, Chen YR, Liu X, et al. Arginine starvation impairs mitochondrial respiratory function in ASS1-deficient breast cancer cells. Sci Signal. 2014;7(319):ra31–ra31.
106. Fiipula DR, Wang M, inventors; Arginine deiminase derived from mycoplasma arthritidis and polymer conjugates containing the same patent.US 6132713. 2000.
107. Filpula DR, Wang M, inventors; Arginine deiminase derived from Mycoplasma arthritidis and polymer conjugates containing the same patent.US 5916793. 1999.
108. Filpula DR, Wang M, inventors; Arginine deminase derived from mycoplasma arthritidis and polymer conjugates containing the same patent.US 5804183. 1998.
109. Clark MA, inventor; Modified arginine deiminase patent.US 6183738. 2001.
•• **US 9255,262 B2: This patent describes an albumin-binding arginine deiminase (AAD) fusion protein that has been genetically modified to create an enzyme that is not immunogenic and has higher activity and longer *in vivo* half-life when compared to ADI native form. Besides, its production is cheaper**

and the product is easier to characterize when compared to the pegylated ADI.

110. Bowles TL, Bowles TL, Kim R, et al. Pancreatic cancer cell lines deficient in argininosuccinate synthetase are sensitive to arginine deprivation by arginine deiminase. Int J Cancer Suppl. 2008;123 (8):1950–1955.
111. Delage B, Luong P, Maharaj L, et al. Promoter methylation of argininosuccinate synthetase-1 sensitises lymphomas to arginine deiminase treatment, autophagy and caspase-dependent apoptosis. Cell Death Dis. 2012;3(7):e342.
112. Feun LG, Marini A, Walker G, et al. Negative argininosuccinate synthetase expression in melanoma tumours may predict clinical benefit from arginine-depleting therapy with pegylated arginine deiminase. Br J Cancer. 2012;106(9):1481–1485.
113. Huang CC, Huang -C-C, Tsai ST, et al. Arginine deprivation as a new treatment strategy for head and neck cancer. Oral Oncol. 2012;48 (12):1227–1235.
114. Huang HY, Wu WR, Wang YH, et al. ASS1 as a novel tumor suppressor gene in myxofibrosarcomas: aberrant loss via epigenetic DNA methylation confers aggressive phenotypes, negative prognostic impact, and therapeutic relevance. Clin Cancer Res. 2013;19 (11):2861–2872.
115. Kelly MP, Jungbluth AA, Wu BW, et al. Arginine deiminase PEG20 inhibits growth of small cell lung cancers lacking expression of argininosuccinate synthetase. Br J Cancer. 2011;106(2):324–332.
116. Syed N, Langer J, Janczar K, et al. Epigenetic status of argininosuccinate synthetase and argininosuccinate lyase modulates autophagy and cell death in glioblastoma. Cell Death Dis. 2013;4(1):e458.
117. Ensor CM, Holtsberg FW, Clark MA, inventors; Mutated form of arginine deiminase patent.US 6635462. 2003.
118. Huang Y, Qiu J, Fu X, et al., inventors; Arginine deiminase mutant and preparation and application thereof patent.US 8663967. 2014.
119. Wong BL, inventor; Albumin-binding arginine deminase and the use thereof patent.US 9255262. 2016.
120. Glazer ES, Piccirillo M, Albino V, et al. Phase II study of pegylated arginine deiminase for nonresectable and metastatic hepatocellular carcinoma. J Clinical Oncology. 2010 May 1;28 (13):2220–2226.
121. Izzo F, Marra P, Beneduce G, et al. Pegylated arginine deiminase treatment of patients with unresectable hepatocellular carcinoma: results from phase I/II studies. J Clinical Oncology. 2004 May 15;22 (10):1815–1822.
122. Yang TS, Lu SN, Chao Y, et al. A randomised phase II study of pegylated arginine deiminase (ADI-PEG 20) in Asian advanced hepatocellular carcinoma patients. Br J Cancer. 2010 Sep 28;103 (7):954–960.

- **EP 2,451,486: This patent describes a new ASNase derived from Erwinina that does not present the hypersensitivity reactions and other toxicities that are found in the ASNase isolated from *E. coli*.**

123. Ascierto PA, Scala S, Castello G, et al. Pegylated arginine deiminase treatment of patients with metastatic melanoma: results from phase I and II studies. J Clinical Oncology. 2005 Oct 20;23 (30):7660–7668.
124. Ott PA, Carvajal RD, Pandit-Taskar N, et al. Phase I/II study of pegylated arginine deiminase (ADI-PEG 20) in patients with advanced melanoma. Invest New Drugs. 2013 Apr;31(2):425–434.
125. Akaike T, Noguchi Y, Ijiri S, et al. Pathogenesis of influenza virus-induced pneumonia: involvement of both nitric oxide and oxygen radicals. Proc Natl Acad Sci USA. 1996;93(6):2448–2453.
126. Karupiah G, Chen JH, Mahalingam S, et al. Rapid interferon gamma-dependent clearance of influenza A virus and protection from consolidating pneumonitis in nitric oxide synthase 2-deficient mice. J Exp Med. 1998;188(8):1541–1546.
127. Lopez-Guerrero JA, Carrasco L. Effect of nitric oxide on poliovirus infection of two human cell lines. J Virol. 1998;72(3):2538–2540.

- **US 20,130,330,316: The present invention relates to a novel mutant of L-asparaginase enzyme characterized in having high thermostability, pH stability, and no glutaminase activity useful for therapeutics and the process of preparing the same.**

128. Sukwattanapan C, Ubol S, Ubol S, et al. Inducible nitric oxide synthase inhibition delays death of rabies virus-infected mice. J Med Microbiol. 2001;50(3):238–242.
129. Kreil TR, Eibl MM. Nitric oxide and viral infection: no antiviral activity against a flavivirus in vitro, and evidence for contribution to pathogenesis in experimental infection in vivo. Virology. 1996;219(1):304–306.
130. Clark MA, inventor Methods for inhibiting viral replication in vivo patent.US 7204980. 2007.
131. Gesto DS, Cerqueira NM, Fernandes PA, et al. Unraveling the enigmatic mechanism of L-asparaginase II with QM/QM calculations. J Am Chem Soc. 2013;135(19):7146–7158.
132. Covini D, Tardito S, Bussolati O, et al. Expanding targets for a metabolic therapy of cancer: L-asparaginase. Recent Pat Anticancer Drug Discov. 2012;7(1):4–13.
133. Roberts J, Bank WRIB, inventors; Method of purifying l-asparaginase patent US 3597323 A. 1971.
134. Ratych OT, Robison RS, Berk B, Inc SS, inventors; Synthesis of l-asparaginase patent US 3511755 A. 1970.
135. Spiers AS, Wade HE. Achromobacter L-glutaminase-L-asparaginase: human pharmacology, toxicology, and activity in acute leukemias. Cancer Treat Rep. 1979;63(6):1019–1024.
136. Sullivan MG. Asparaginase Erwinia chrysanthemi approved for acute lymphoblastic leukemia. Community Oncol. 2011;8(12):569–570.
137. Abribat T, inventor Pegylated l-asparaginase patent.EP 2451486 A1. 2011.
138. Moghrabi A, Levy DE, Asselin B, et al. Results of the Dana-Farber Cancer Institute ALL Consortium Protocol 95-01 for children with acute lymphoblastic leukemia. Blood. 2007 Feb 1;109(3):896–904.
139. Pieters R, Hunger SP, Boos J, et al. L-asparaginase treatment in acute lymphoblastic leukemia: a focus on Erwinia asparaginase. Cancer. 2011 Jan 15;117(2):238–249.
140. Kwok CS, Kham SK, Ariffin H, et al. Minimal residual disease (MRD) measurement as a tool to compare the efficacy of chemotherapeutic drug regimens using Escherichia coli-asparaginase or Erwinia-asparaginase in childhood acute lymphoblastic leukemia (ALL). Pediatr Blood Cancer. 2006 Sep;47(3):299–304.
141. Allas S, Sahakian P, Fichtner I, et al. Pharmacokinetics and Pharmacodynamics in Mice of a Pegylated Recombinant Erwinia Chrysanthemi-Derived L-Asparaginase. Blood. 2009;114(22):2033–33.
142. Chien -W-W, Allas S, Rachinel N, et al. Pharmacology, immunogenicity, and efficacy of a novel pegylated recombinant Erwinia chrysanthemi-derived L-asparaginase. Invest New Drugs. 2014;32(5):795–805.
143. Samudio I, Konopleva M. Asparaginase unveils glutamine-addicted AML. Blood. 2013 Nov 14;122(20):3398–3400.
144. Derst C, Henseling J, Röhm KH. Engineering the substrate specificity of Escherichia coliasparaginase II. Selective reduction of glutaminase activity by amino acid replacements at position 248. Protein Sci. 2000;9(10):2009–2017.
145. Li L-Z, Xie T-H, Li H-J, et al. Enhancing the thermostability of Escherichia coli l-asparaginase II by substitution with pro in predicted hydrogen-bonded turn structures. Enzyme Microb Technol. 2007;41(4):523–527.
146. Bansal S, Kundu B, Mishra P. Department of biotechnology MOST, Indian Institute Of Technology D, inventors. Mutants of l-Asparagine patent WO 2012028945 A3. 2012.
147. Gummadi C. Screening and isolation of novel glutaminase Free L-asparaginase from fungal endophytes. Research Journal of Microbiology. 2014;9(4):163–176.
148. Avramis VI, Cohen L, Aventis Pharmaceuticals Products I, inventors; Pharmaceutical compositions comprising peg-asparaginase for the treatment of hiv infections patent.WO 1999039732 A1. 1999.
149. Shen L-J, Shen W-C. Drug evaluation: ADI-PEG-20–a PEGylated arginine deiminase for arginine-auxotrophic cancers. Curr Opin Mol Ther. 2006;8(3):240–248.
150. Gong H, Zölzer F, von Recklinghausen G, et al. Arginine deiminase inhibits proliferation of human leukemia cells more potently than asparaginase by inducing cell cycle arrest and apoptosis. Leukemia. 2000;14(5):826–829.
151. Müller H. Use of ?-asparaginase in childhood ALL. Crit Rev Oncol Hematol. 1998;28(2):97–113.

152. Schellekens H. Immunogenicity of therapeutic proteins: clinical implications and future prospects. Clin Ther. 2002;24(11):1720–1740. discussion 19.
153. Cantor JR, Yoo TH, Dixit A, et al. Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. Proc Natl Acad Sci USA. 2011;108(4):1272–1277.
154. Sun X, Yang Z, Li S, et al. In vivo efficacy of recombinant methioninase is enhanced by the combination of polyethylene glycol conjugation and pyridoxal 5′-phosphate supplementation. Cancer Res. 2003;63(23):8377–8383.
155. Miki K, Xu M, Gupta A, et al. Methioninase cancer gene therapy with selenomethionine as suicide prodrug substrate. Cancer Res. 2001;61(18):6805–6810.
156. Yang Z, Wang J, Yoshioka T, et al. Pharmacokinetics, methionine depletion, and antigenicity of recombinant methioninase in primates. Clin Cancer Res. 2004;10(6):2131–2138.

View publication stats

Nanostructures in Therapeutic Medicine Series

Nanostructures for Cancer Therapy

Edited by

Anton Ficai
University Politehnica of Bucharest, Bucharest, Romania

Alexandru Mihai Grumezescu
University Politehnica of Bucharest, Bucharest, Romania

CHAPTER

CANCER THERAPIES BASED ON ENZYMATIC AMINO ACID DEPLETION

24

Carla S. Silva Teixeira, Henrique S. Fernandes, Pedro A. Fernandes, Maria J. Ramos, Nuno M.F.S.A. Cerqueira

University of Porto, Porto, Portugal

CHAPTER OUTLINE

1 Introduction 623
2 Amino Acid Deprivation Enzymes 624
2.1 L-Asparagine 625
2.2 L-Arginine 627
2.3 L-Methionine 637
3 Conclusions 641
Acknowledgments 642
References 642

1 INTRODUCTION

Cancer has become the leading cause of death in the developed world and has remained one of the most difficult diseases to treat. During the past 5 decades, cancer therapy has relied mainly on low molecular weight molecules. However, this class of compounds, very potent in their action, unfortunately lacks tumor specificity and often has unacceptable toxicities toward normal cells. All of these factors are often followed by side effects, which often undermine the effectiveness of the therapeutic treatment in patients (Cantor et al., 2012b). For this reason, the demand for new and improved therapeutics against cancer remains high, and at the same time a deeper knowledge regarding the development of tumors is required.

The development of tumor cells is still not fully understood, but many advances have been made in this area. Alterations to cellular metabolism seem to constitute a nearly universal feature in many types of cancers. For instance, many tumors exhibit deficiencies in their enzymatic armamentarium and cannot biosynthesize one or more amino acids that are essential for its development, survival, and spread. In order to overcome these problems, tumors rely on the extracellular pool of these amino acids to satisfy protein biosynthesis demands and continue to grow without being affected. This means that if the concentration of those amino acids is decreased, the development of tumor cells can be impaired or even annihilated. At the same time, normal cells remain unchanged because they are less demanding or can synthesize these compounds in sufficient amounts by alternative pathways.

These observations are the basis of the amino acid deprivation cancer therapy that involves the systemic depletion of tumor-essential amino acid. This knowledge is not new and was initially suggested

Copyright © 2017 Elsevier Inc. All rights reserved.

over 50 years ago, but only now is it gaining popularity with the advances in protein engineering technology and in more sophisticated approaches that enable the study of genetic and metabolic differences between tumor and normal cells (Cantor et al., 2012a).

In these tumors, diet is not enough to achieve a therapeutically relevant level for amino acid depletion, and therefore enzymes are used instead. Regrettably, the human genome does not encode enzymes with the pharmacological or catalytic requisites that are essential for the pretended therapeutic purpose. For this reason heterologous enzymes (from other organisms), recombinant, and/or engineered human enzymes are used both in animal studies and in clinical trials. Currently the enzymes of bacterial origin are the most frequently used.

The main function of heterologous or engineered human enzymes used in amino acid depletion therapies is to decrease the concentration of a certain amino acid in the bloodstream (e.g., converting it into another molecule) that is essential for the development of a tumor. Some examples include a large fraction of hepatocellular carcinomas (HCC) and metastatic melanomas that become apoptotic under conditions where the nonessential amino acid L-arginine (L-ARG) in serum is depleted (Ni et al., 2008), central nervous system cancers that respond to L-methionine (L-MET) deprivation (Tan et al., 2010), and acute lymphoblastic leukemia (ALL) for which depletion of L-asparagine (L-ASN) has been shown to significantly reduce cancer growth (Rytting, 2010).

In this review, the structure, function, and catalytic mechanism of therapeutic heterologous enzymes involved in L-ARG, L-MET, and L-ASP degradation will be revised. Particular attention will be given to the heterologous enzymes L-arginase, L-arginine deiminase, L-methioninase, and L-asparaginase, taking into account their current therapeutic potential.

2 AMINO ACID DEPRIVATION ENZYMES

At least 20 amino acids have been identified in cells and among these are 9 that the human body cannot manufacture de novo (from scratch), and consequently these must be obtained from the diet. These amino acids are normally denominated as essential and include L-phenylalanine, L-valine, L-threonine, L-tryptophan, L-methionine, L-leucine, L-isoleucine, L-lysine, and L-histidine. Six other amino acids are termed conditionally essential. This means that their synthesis can be carried out by humans but their synthesis can be limited by a variety of factors, such as prematurity in the infant or individuals in severe catabolic distress. These amino acids are L-arginine, L-cysteine, L-glycine, L-glutamine, L-proline, and L-tyrosine. The remaining five amino acids are dispensable in humans (also denoted as nonessential), meaning they can be synthesized in the body, for example, L-aspartic acid, L-asparagine, L-glutamic acid, L-alanine, and L-serine.

The optimal conditions for amino acid depletion therapies occur when tumors are sensitive to one of the nonessential or conditionally essential amino acids. This will ensure that only tumor cells will be affected by the therapy, whereas the normal cells remain unaltered because they can synthesize these compounds in sufficient amounts by their own means. The most notable example of target enzymatic amino acid depletion in cancer therapy is illustrated by the remarkable success of L-asparaginase in the treatment of childhood ALL and, to a lesser extent, in non-Hodgkins lymphoma. Other examples involve the use of arginine deiminase in the HCC, melanoma, and other urea cycle-deficient cancer cells that have a high demand for L-arginine. The essential amino acids can also be used in amino acid depletion therapies, but it is not very common. This only happens in tumors where the normal metabolism of

one essential amino acid is disrupted or when there is some sort of defect in the ability to use a certain amino acid. For example, this is the case of L-methionine, which has been shown to be detrimental for the survival of a variety of tumor tissues, including colon, breast, prostate, ovary, lung, brain, kidney, stomach, and bladder cancers, as well as larynx melanoma, sarcoma, leukemia, and lymphomas. In this case the action of methionine-γ-lyase has been shown to be successfully used in the treatment of these types of cancers.

Taking into account the therapeutic potential of these enzymes in cancer therapies, they have been deeply studied in the past 2 decades. These therapies have been showing promising results, but also some problems in clinical trials. In the following sections a detailed analysis of each of these enzymes is done based on the amino acid that they metabolize.

2.1 L-ASPARAGINE

L-ASN is a nonessential amino acid that is involved in the metabolic control of cell functions in nerve and brain tissue. This amino acid is important in the synthesis of a large number of proteins and plays an important role in the biosynthesis of glycoproteins where it provides key sites for *N*-linked glycosylation. ASN also plays an important role in the metabolism of ammonia, which is toxic in the human body.

Since L-asparagine is a nonessential amino acid, it can be easily synthesized by the cells. The precursor in the biosynthesis of ASN is oxaloacetate, which reacts with L-glutamate to form L-aspartate and α-ketoglutarate, in the presence of a transaminase enzyme. The L-aspartate then reacts with L-glutamine to form L-asparagine in an ATP-dependent reaction that is catalyzed by L-asparagine synthetase (Richards and Kilberg, 2006).

Certain types of tumors, such as leukemic cells, cannot synthesize ASN because they lack or have deficient expression of the enzyme L-Asparaginase synthase. This means that the proliferation and survival of these tumor cells are dependent on the external supplies of this amino acid. As normal cells can synthesise ASN, this means that L-asparagine deprivation therapies would mainly interfere with tumor cells that are ASN deficient.

L-Asparaginase (ASNase) is a bacterial enzyme that has been used with great success in leukemia treatment regimens for 4 decades. This enzyme performs the opposite reaction of L-asparagine synthetase by hydrolyzing ASN to L-aspartic acid and ammonia. By reducing the levels of plasmatic L-asparagine, ASNase starves leukemic cells and promotes tumor cell apoptosis.

The therapeutic role of ASNase was first noted by Kidd (1953), who found that guinea pig serum had antitumor activity against two kinds of lymphoma in a murine model and lymphosarcoma in rats. This fact was later attributed to the enzyme ASNase that was present in the serum (Broome, 1961, 1963). Later on, in 1966, DeLowrey and coworkers using a purified guinea pig serum treated a boy with acute lymphoblastic leukemia, and obtained an objective response (Dolowy et al., 1966). This turned ASNase into a promising anticancer drug, and therefore large amounts of the enzyme were required to start new studies and large-scale clinical trials. However, this was not an easy task at that time since the isolation and purification of the enzyme was difficult and not all the ASNases possess antitumor activity.

L-Asparaginases (EC 3.5.1.1) is a relatively widespread enzyme, found in many microorganisms, such as *Aerobacter*, *Bacillus*, *Pseudomonas*, *Serratia*, *Xanthomonas*, *Photobacterium* (Peterson and Ciegler, 1969), *Streptomyces* (Dejong, 1972), *Proteus* (Tosa et al., 1971), *Vibrio* (Kafkewitz and

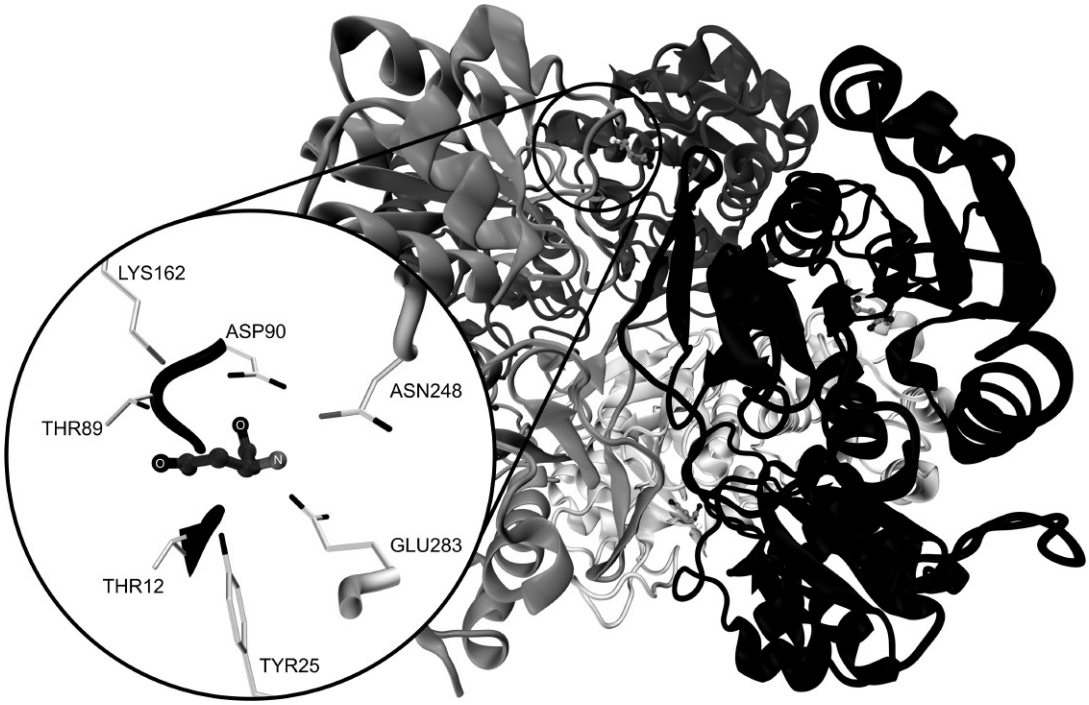


FIGURE 24.1 New Cartoon Representation of the Cocrystalized Ctructure of ASNase With L-Asparagine (PDB Code: 3ECA)

Each subunit is represented with a different black tonality and the substrate in ball and sticks and colored by CPK. In the close-up of the active site the residues of the active site that are important for the catalytic process, as well as the substrate are represented in licorice.

Goodman, 1974), and *Aspergillus* (Sarquis et al., 2004). After intense research ASNases derived from *Escherichia coli* (Ecoli-ASNase) and *Erwinia chrysanthemi* (Echry-ASNase) turned out to have the best anticancer capabilities. Ecoli-ASNase has thereby become the main source of ASNase because it is easy to produce in large amounts. Echry-ASNase is typically reserved for cases of Ecoli-ASNase hypersensitivity.

The *E. coli* expresses two ASNases denoted as Ecoli-ASNase-I and Ecoli-ASNase-II that are found in different intra- or extracellular localizations. Ecoli-ASNase-I (cytosolic) has a lower affinity for ASN, while Ecoli-ASNase-II (periplasmic) has a high ASN affinity. Due to the higher affinity of Ecoli-ASNase-II toward L-asparagine, this enzyme is the mostly studied enzyme and the one that is used for oncologic therapeutic purposes.

Ecoli-ASNase-II is a tetrameric protein that is composed of four identical subunits, each one containing 326 amino acid residues. The binding pocket of this enzyme involves residues from both subunits of each dimer, namely, THR12, TYR25, SER58, GLN59, THR89, ASP90, and LYS162 from one subunit (Fig. 24.1—black chain), and ASN248 and GLU283 from the other subunit (Fig. 24.1—light gray

chain). All of these residues are interconnected by a strong hydrogen bond network, including a water molecule that is structurally conserved. The position of the amino acid residues in the active site suggests several possible pathways for the catalytic reaction, but the almost symmetric location of two threonine residues, THR12 and THR89, observed in the X-ray structure 3ECA above and below carbon C2 of the substrate, suggests that one of them must be directly involved in the reaction (Fig. 24.1) (Swain et al., 1993).

The first step of the mechanism involves the formation of a tetrahedral intermediate (Fig. 24.2a) that results from the nucleophilic attack of the water molecule to α-carbon of the substrate. In the course of this reaction, LYS162 receives the proton from the water molecule and becomes positively charged. The second step (Step 2) of the reaction involves a concerted double proton transfer from LYS162 to THR89 and from THR89 to the substrate, from which results the formation of ammonia and aspartic acid. However, the ammonia molecule is still weakly bound to the substrate at this stage and its dissociation requires an additional step (Step 3). At the end of this process (Step 4), the enzyme is ready for the next turnover (Gesto et al., 2013).

Ecoli-ASNase-II and Echry-ASNase have been established as a general treatment for acute lymphoblastic leukemia for nearly 40 years. It is used for remission induction and intensification treatment in all pediatric regimens and in the majority of adult treatment protocols. Currently, it is estimated that these enzymes have contributed to significant improvements of therapy outcomes and to achieve complete remission of leukemia in about 90% of patients.

Notwithstanding its high therapeutic efficacy, the administration of Ecoli-ASNase-II for long periods produces acute allergic reactions and silent immunity, a condition that is characterized by circulating antibodies and rapid clearance of the enzyme from the blood (Avramis et al., 2002; Henriksen et al., 2015; Woo et al., 1998). In order to overcome this problem, a modification of Ecoli-ASNase-II with 5-kDa units of monomethoxypolyethylene glycol was introduced (PEG-Ecoli-ASNase-II) (Rytting, 2010). The modified enzyme preserves enzyme activity, but decreases immunogenicity of the protein and extends elimination half-life (5 times longer than native enzyme) (Keating et al., 1993).

Currently, three L-asparaginase agents are available. These are either derived from *E. coli* in its native form (Ecoli-ASNase-II) or as a pegylated enzyme (PEG-Ecoli-ASNase-II), or extracted from *Erwinia chrysanthemi* (Echry-ASNase). Due to the unique anticancer mechanism of action, PEG-Ecoli-ASNase-II and Echry-ASNase have already been approved by US Food and Drug Administration (FDA) for the treatment of several cancers, such as acute lymphoblastic leukemia (Hill et al., 1967), lymphosarcoma (Story et al., 1993), and a few subtypes of non-Hodgkin's lymphoma (Kobrinsky et al., 2001). Ovarian carcinomas and other solid tumors have also been proposed as additional targets for ASNase, with a potential role for its glutaminase activity (Covini et al., 2012).

2.2 L-ARGININE

L-ARG is a precursor/mediator of a series of biological pathways, some of them involved in important cellular functions, such as in nitrogen metabolism (Rogers and Visek, 1985), creatine, agmatine, and polyamine synthesis (Klein and Morris, 1978). In addition, ARG is the major substrate for the production of nitric oxide (Billiar, 1995), stimulates anabolic hormone release, and is immunostimulatory and thymotrophic (Barbul et al., 1977, 1983).

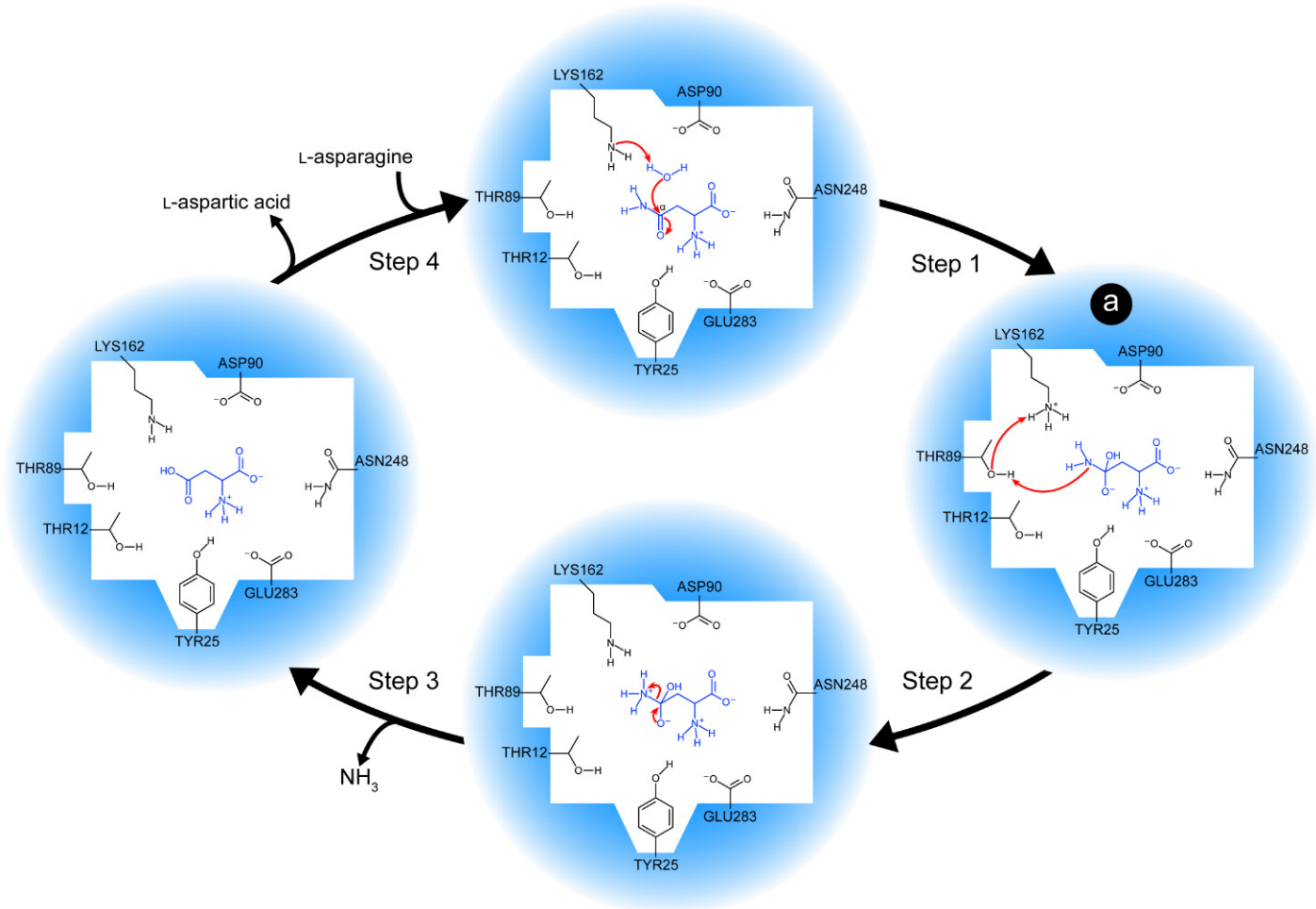


FIGURE 24.2 Schematic Representation of the ASNase Catalytic Mechanism

Each circle represents a different reaction intermediate. The important reaction intermediates are labeled with labels (a).

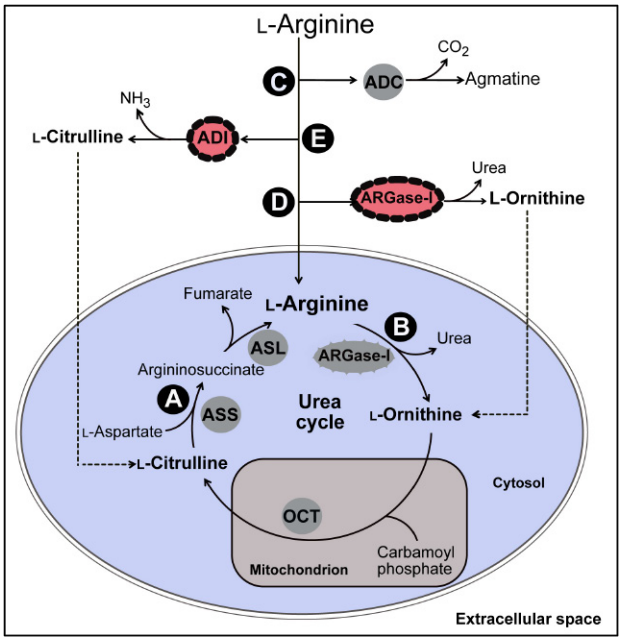


FIGURE 24.3 Schematic Representation of the Enzymes Involved in ARG Metabolism and Degradation

The enzymes available on the cell have grey background. Recombinant enzymes or from bacterial sources are represented with dashed outline.

The enzyme argininosuccinate synthetase (ASS) catalyzes the condensation of L-Citrulline and L-aspartic acid to argininosuccinate which is subsequently converted to ARG and fumaric acid by argininosuccinate lyase (ASL) (Fig. 24.3A) (Haines et al., 2011).

In healthy adults the level of endogenous synthesis is sufficient to make ARG a nonessential dietary amino acid (Castillo et al., 1993). However under catabolic stress (e.g., inflammation, infection, etc.), the levels of endogenous synthesis may not be sufficient to meet metabolic demands and ARG becomes an essential amino acid. For this reason, this amino acid is often denoted as a semiessential amino acid instead of nonessential (Reeds, 2000).

Certain tumor cells, such as metastatic melanoma, prostate carcinomas, HCC (Dillon et al., 2004), non-Hodgkin's lymphoma, Hodgkin's lymphoma, pancreatic carcinoma, osteosarcoma, and malignant pleural mesothelioma and some breast tumors (Qiu et al., 2015) have an elevated requirement for ARG due to the deficient expression of ASS. This causes ARG autotrophy and therefore these tumors rely only on exogenous sources of ARG for growth and proliferation. Deprivation of this amino acid is, for this reason, investigated as a novel strategy for cancer therapy and has been showing promising efficacy against the ARG-auxotrophic tumors mentioned earlier (Feun et al., 2015; Shen et al., 2003; Wheatley and Campbell, 2003).

ARG can be degraded by several mechanisms, either using enzymes available on the cell, such as L-arginase (ARGase) (Fig. 24.3B) or L-arginine decarboxylase (ADC) (Fig. 24.3C), using human recombinant ARGase-I (Fig. 24.3D), or using parasite-enzymes from a bacterial source, such as arginine

deiminase (ADI, Fig. 24.3E). From these enzymes, ARGase-I and ADI are the most promising ones since they have therapeutic potential. L-arginine decarboxylase is relatively toxic to normal cells and therefore it is not used for therapeutic purposes (Savaraj et al., 2010).

2.2.1 L-Arginase

L-ARG can be degraded directly to L-ornithine by ARGase (EC 3.5.3.1), an enzyme of the urea cycle (Fig. 24.3B). Thus the insertion of ARGase on the extracellular moiety could be used to decrease the levels of ARG in cells. However, under salvage conditions, higher concentration of L-ornithine (product of the reaction of ARGase) can be converted back to L-citrulline by ornithine transcarbamyl transferase and then recycle back to ARG by ASS/ASL (Fig. 24.3A). This feedback mechanism makes normal cells and many tumors resistant to ARGase inhibition. Fortunately, there are many tumors that present deficient expressions of OCT 0and/or ASS enzymes and in these cases ARGase can be efficiently used to decrease ARG intracellular levels and therefore tumor growth (Cheng et al., 2007).

ARGase is found in organisms from the five kingdoms (Reczkowski and Ash, 1992). In most mammals, two isoforms of this enzyme exist and are denoted as types I and II. They share approximately 60% amino acid sequence homology and differ in their expression regulation, tissue distribution, subcellular locations, immunologic reactivity, physiologic function, and certain enzymatic properties (Dizikes et al., 1986; Herzfeld and Raper, 1976; Jenkinson et al., 1996; Spector et al., 1983; Vockley et al., 1996). ARGase-I is a cytosolic enzyme located predominantly in the liver (Krebs, 1973). ARGase-II is a mitochondrial enzyme that does not appear to function in the urea cycle and is more widely distributed in numerous tissues (Morris et al., 1997).

In mammals, ARGase-I is present in the liver and it is one of the most important enzymes involved in nitrogen metabolism. It is involved in the principal route for disposal of excess nitrogen resulting from amino acid and nucleotide metabolism. The flux of nitrogen through this pathway is considerable, given that the average individual excretes about 10 kg of urea per year (Ash et al., 1998).

In humans, h-ARGase-I is an oligomer of 107 kDa and shows optimum pH at 9.3 and has a Km for ARG of about 10.5 mM (Beruter et al., 1978). The key structural features of mammalian ARGase-I include a binuclear manganese cluster (Mn), and an S-shaped tail composed by 19 amino acids at the C terminus of each monomer. The S-shaped tails are located at the subunit-subunit interface, and 54% of the intersubunit interactions are mediated by this region of the protein (Lavulo et al., 2001).

The catalytic mechanism of this enzyme is still not fully understood but several suggestions have already been proposed, some of them more consensually accepted (Christianson and Cox, 1999; Cox et al., 2001; Leopoldini et al., 2009) than others (Khangulov et al., 1998). Among the active site residues, GLU277 is apparently implicated in the substrate binding through the establishment of electrostatic interactions with the guanidium group of the substrate. This interaction appears to orientate the subtract in a favorable position with respect to the metal-bound hydroxide anion with whom it must react (Cavalli et al., 1994; Kanyo et al., 1996; Scolnick et al., 1997). Both manganese ions are coordinated to a hydroxide anion and with two aspartates (ASP232 and ASP128) and one glutamate (GLU277). These structural features can be observed in the crystallographic structure with the PDB code 3GMZ, where the human ARGase is in complex with L-ornithine. (Fig. 24.4) (Ilies et al., 2011).

According to (Leopoldini et al., 2009) the hydrolysis of ARG mediated by ARGase-I is kinetically favored when the ASP128 residue is deprotonated. The hydrolysis starts with the binding of ARG to the Mn(II) ions through the gaunidium groups. These interactions place the substrate in a good position

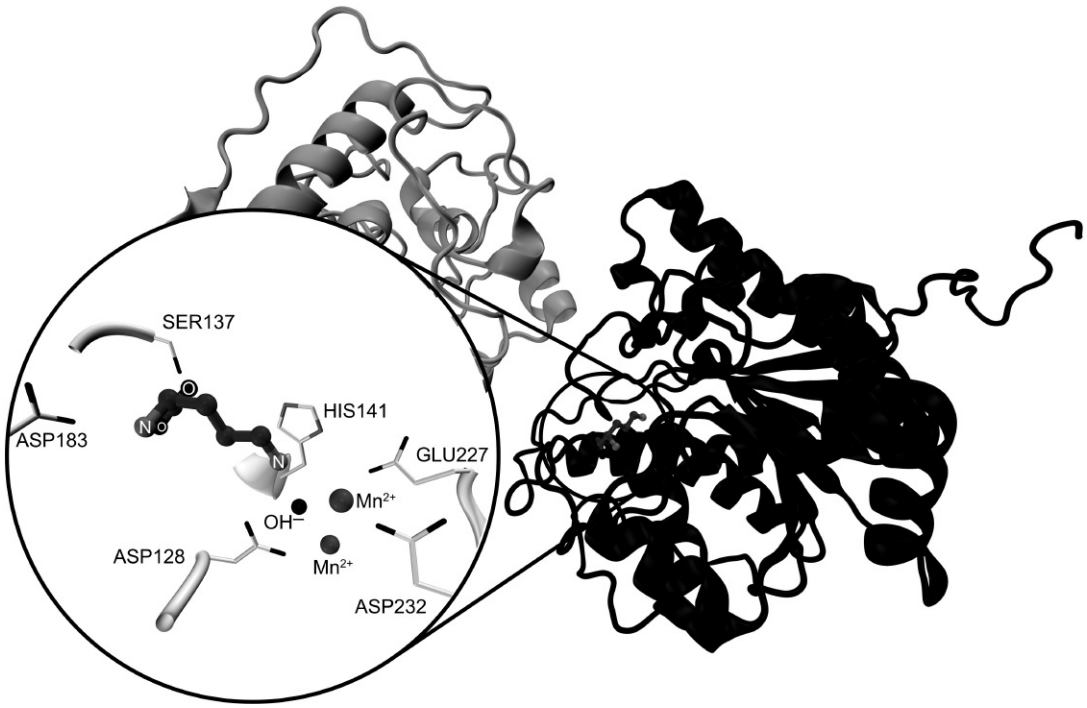


FIGURE 24.4 New Cartoon Representation of the Cocrystalized Structure of h-ARGase With L-Ornithine (PDB Code: 3GMZ)

Each subunit is represented with a different color *(gray and black)* and the substrate in ball and sticks and colored by CPK. In the close-up of the active site are represented: the residues of the active site that are important for the catalytic process in licorice, and the oxygen atom from the hydroxide ion and the two magnesium ions in VDW.

to undergo nucleophilic attack by the bridging hydroxy anion to carbon Cζ (Step 1) from which results a tetrahedral intermediate (Fig. 24.5a). In the second step of the mechanism, ASP128 acts as an acid/base catalyst, abstracting a proton from the hydroxyl group and shuttling it to the nitrogen Nε of the substrate (Step 2). This step is then followed by the bond break between nitrogen Nε and carbon Cζ of the substrate from which results L-ornithine and urea (Step 3). In the following step (Step 4), a water molecule enters to bridge the binuclear Mn(II) cluster, causing the urea product to move to a terminal coordination site on Mn(II). HIS141 appears to facilitate product release by serving as a proton shuttle from bulk solvent to nitrogen Nε of ornithine and regenerating the nucleophilic metal-bridging hydroxide anion (Step 5). Once this step is complete, ornithine dissociates, the enzymatic turnover takes place, and the enzyme is ready for a new catalytic cycle (Step 6).

h-ARGase-I has been tested with success in experimental animals since 1950, and showed promising activity in in vitro conditions (Simon-Reuss, 1953; Wiswell, 1951). However, the therapy failed to produce major responses when tested in in vivo conditions. The reason for this failure was related with the low affinity of this enzyme for ARG with the consequent requirement of a large quantity of enzyme in order to obtain the expected therapeutic results. In addition, this enzyme has a short circulatory

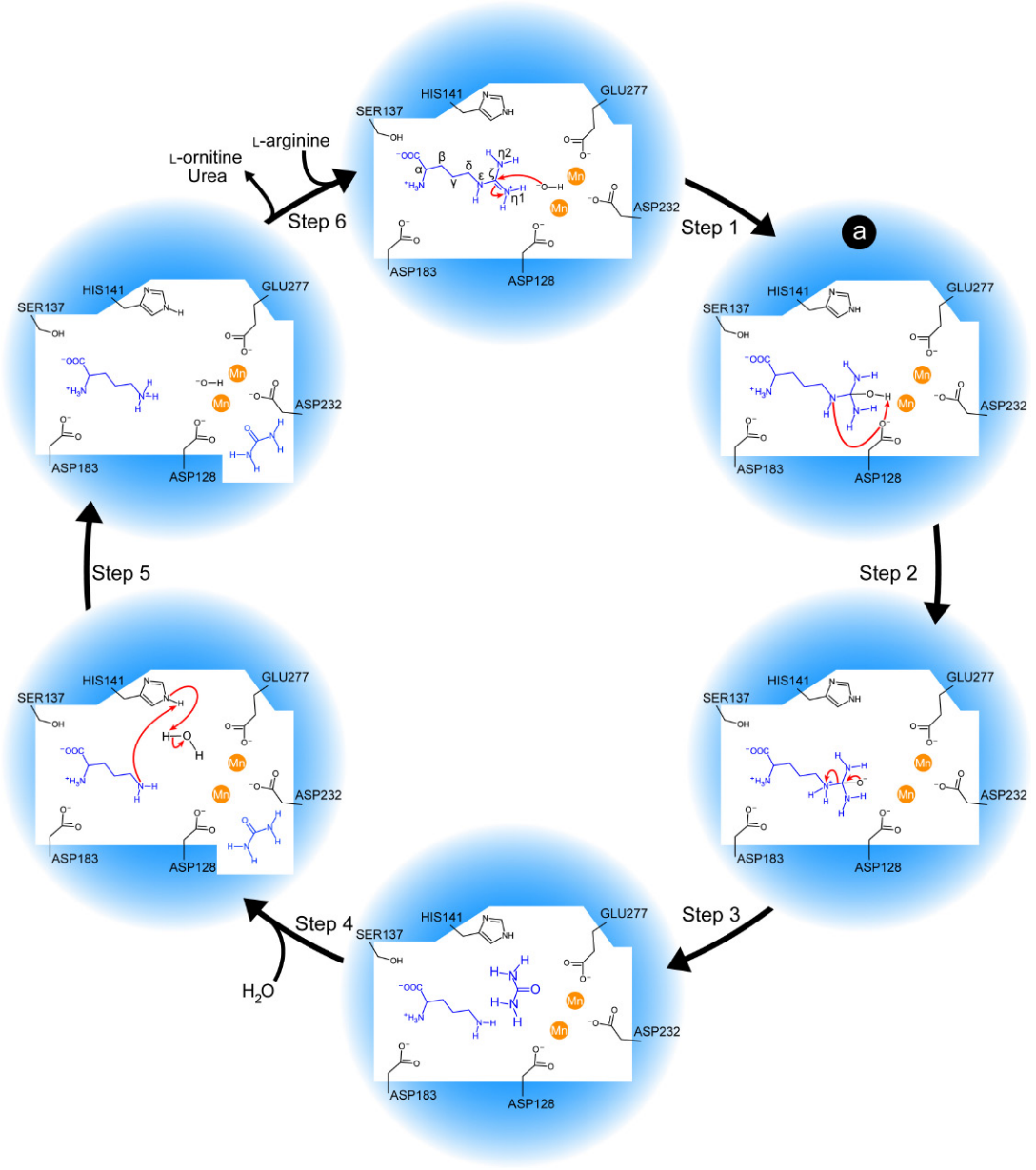


FIGURE 24.5 Schematic Representation of the ARGase Catalytic Mechanism

Each circle represents a different reaction intermediate. The important reaction intermediate is indicated with label (a).

half-life (only a few minutes) (Savoca et al., 1979) and its optimal pH is very high (9.5), which cannot be achieved under physiological conditions. All of these drawbacks limited the use of h-ARGase-I as an antitumor agent in the last decades (Currie et al., 1979). In spite of these drawbacks it was shown in 2010 that a pegylated recombinant human arginase I (PEG-r-h-ARGase-I) impairs the proliferation of malignant T cells in acute lymphoblastic T cell leukemia in in vivo and in vitro studies (Hernandez et al., 2010).

Later on, a pegylated cobalt-modified enzyme was developed and showed that the substitution of the Mn(II) metal center in h-ARGase-I by Co(II) (Co-r-h-ARGase-I) makes the enzyme more efficient for ARG hydrolysis and lowers the IC50 in a variety of malignant cell lines, including melanomas and hepatocellular carcinomas in in vitro assays (Stone et al., 2010). The only-same group also reported a modification of Co-r-h-ARGase-I with PEG-5K esters that showed an increase retention of the enzyme in circulation by about 2 orders of magnitude. The use of PEG(5K)-*co*-h-ARGase-I was later found to be effective in the treatment of hepatocellular and pancreatic carcinomas in xenograft models (Glazer et al., 2011).

2.2.2 L-arginine deiminase

L-Arginine deiminase (ADI, EC 3.5.3.6) is widely distributed among prokaryotic organisms and some anaerobic eukaryotes, but has never been identified in humans or other higher eukaryotes (Fig. 24.3E). The organisms that employ this pathway use ADI to convert ARG to L-citrulline and ammonia (Horn, 1933). L-citrulline is then degraded further, forming ATP (Adenosine Tri-Phosphate), carbon dioxide, and L-ornithine (Arena et al., 1999). The ADI pathway is thus used by these organisms to get energy, carbon, and nitrogen. In addition, it has also been suggested that this pathway can protect some bacteria from acidic conditions, by the production of ammonia (Casianocolon and Marquis, 1988) (Fig. 24.6).

The interest in ADI for therapeutic purposes began in the 1990s when Takaku and coworkers successfully used this enzyme to inhibit the growth of several murine and human tumor cell lines in vitro and in vivo, by exhausting the supplies of ARG (Takaku et al., 1995; Takaku et al., 1992). From that time, several studies confirmed the potential of ADI, purified from Mycoplasma, as an antitumor agent in different types of tumors (Ashikaga et al., 1994; Sugimura et al., 1992; Takaku et al., 1992). Later on, it was also found that beyond the antitumor activity, ADI has an antiangiogenic activity via suppression of nitric oxide generation (Beloussow et al., 2002; Park et al., 2003; Yoon et al., 2007). The cooperative antiproliferative and antiangiogenic activities of ADI are believed to be the key assets that turn ADI very effective in the treatment of several tumors and other diseases.

ADI has been structurally and pharmacologically characterized from various bacteria, but the most studied enzyme is ADI from *Mycoplasma arginini*, taking into account its therapeutic potential (Zhang et al., 2015). ADI from *M. arginini* has 410 amino acids and is composed of two identical subunits with a molecular weight of 45 kDa each and has an isoelectric point of 4.7. Its V_{max} value and K_m value for ARG is estimated to be 50 units/mg protein and 0.2 mM, respectively. It exerted maximal enzyme activity at pH 6.0–7.5 and at 50°C (Takaku et al., 1992).

A proposal for the catalytic mechanism of ADI from M. arginini was published by Das et al. (2004). The crystallographic structures available on the protein databank show that the active site is deeply buried on the protein surface and the substrate binds with the guanidium group pointing toward two aspartates (ASP161 and ASP271). These interactions endorse a specific orientation of the substrate that becomes aligned with HIS296 and CYS398, two residues that play an important role in the catalytic

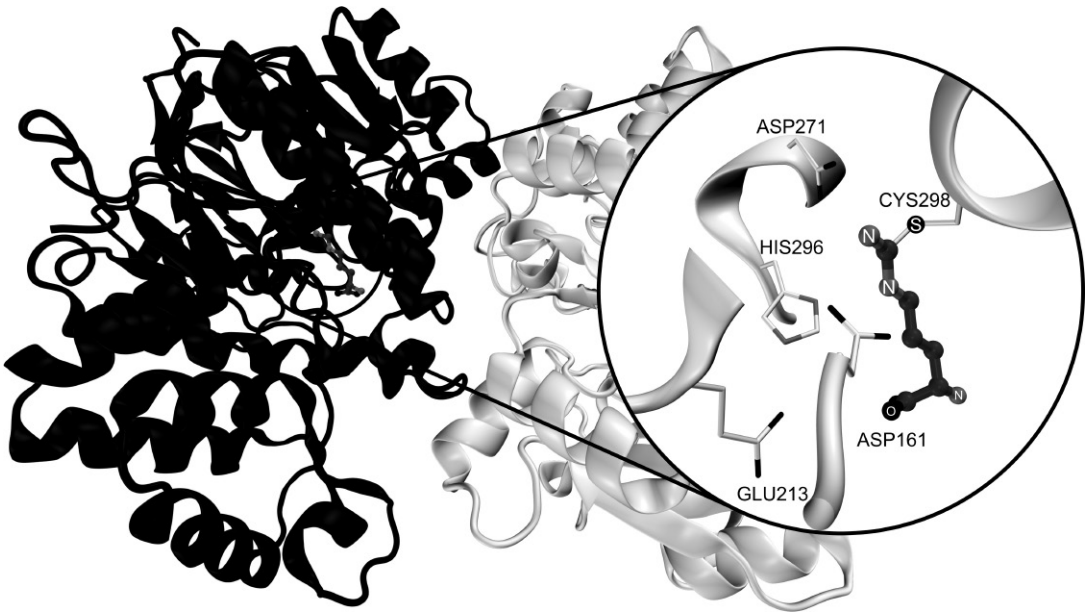


FIGURE 24.6 New Cartoon Representation of the Cocrystalized Structure of ADI Homodimer Structure With a Tetrahedal Reaction Intermediate (PDB Code: 1S9R)

Each subunit was represented with a different color *(gray and black)* and the reaction intermediate in ball and sticks and colored by CPK. In the close-up of the active site the residues of the active site that are important for the catalytic process are represented in licorice, as well as the reaction intermediate that should be formed after the first step of the mechanism.

process. These structural features are represented in Fig. 24.7 (PDB code 1S9R) (Das et al., 2004). The importance of the CYS398 in the catalytic reaction is well established, since it has already been demonstrated through structural studies, mutagenesis, and thiol modifying agents (Das et al., 2004; Galkin et al., 2005; Lu et al., 2004, 2006; Smith et al., 1978).

In the first step of the mechanism there is a nucleophilic attack of CYS398 to carbon Cζ of the substrate. Simultaneously, the proton from CYS398 migrates to nitrogen Nη2 of the substrate. At the same time, one proton from the same nitrogen migrates to HIS296 and the proton from HIS296 is transferred to GLU213. In the end of the first step, a tetrahedral intermediate (Fig. 24.7a) is generated with carbon Cζ of the substrate covalent bonded to CYS398.

In the second step of the catalytic mechanism occurs the formation of ammonia through the migration of one proton from HIS296 to Nη2 of the substrate. Simultaneously, one proton is transferred from GLU213 to HIS296. In the end of this reaction, the reaction intermediate (Fig. 24.7b) remains covalently bonded to CYS398 and one molecule of ammonia is released from the active site.

The next steps of the mechanism (Fig. 24.7, Steps 4–6) involve the release of the product of the reaction (L-Citrulline) and the enzymatic turnover. This requires three sequential steps and involves the reaction of one water molecule that is in close proximity to ASP271 and HIS296. In the fourth step of the catalytic mechanism occurs the nucleophilic attachment of the water molecule to carbon Cζ of

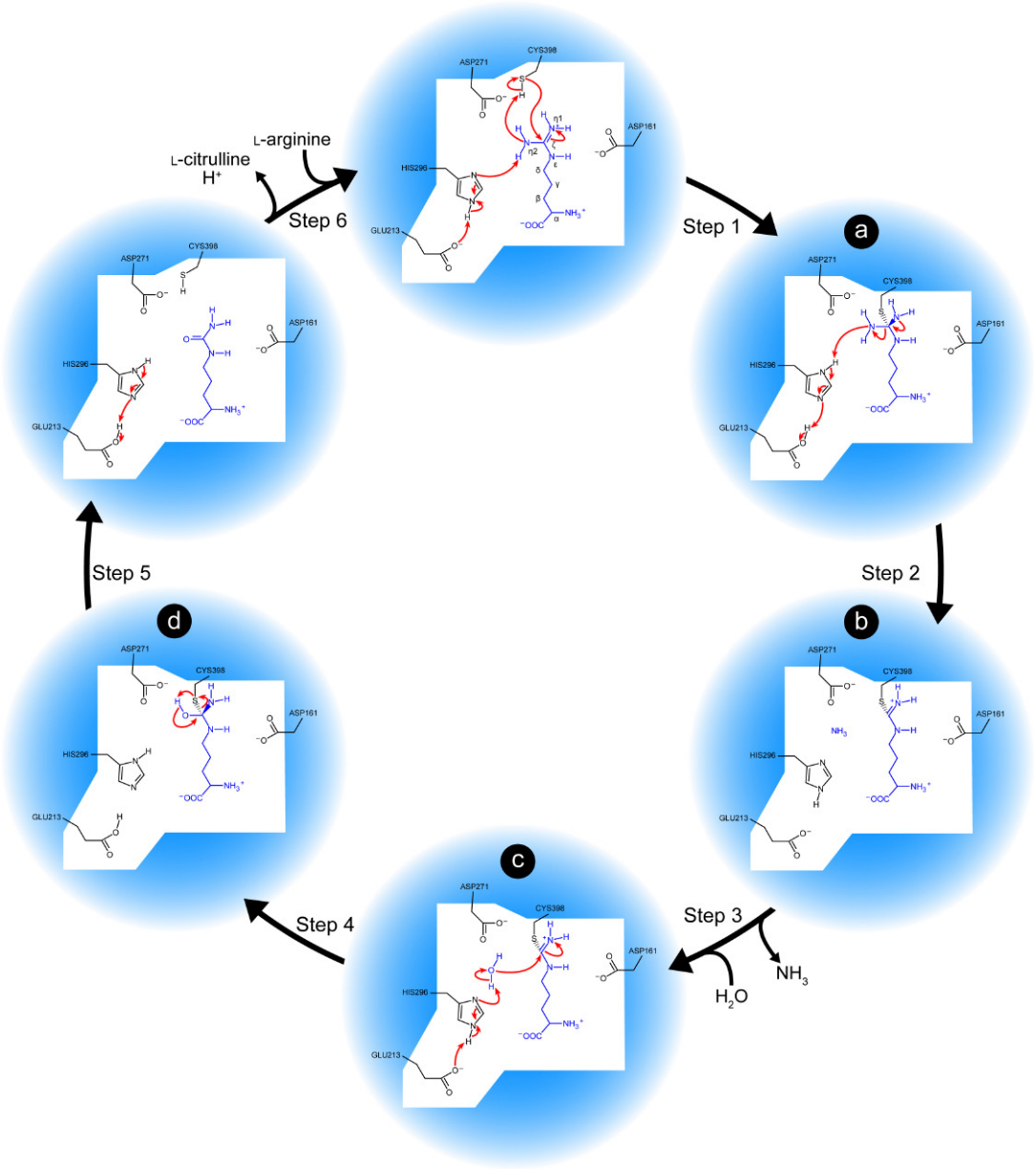


FIGURE 24.7 Schematic Representation of the ADI Catalysis Mechanism

Each circle represents a different reaction intermediate. The important reaction intermediates are labelled (a–d).

the reaction intermediate (Fig. 24.7c). At the same time one proton from the water molecule is transferred to HIS296 and one proton migrates from HIS296 to GLU213. The reaction is endorsed by the close proximity of ASP161 that establishes two hydrogen bonds with the NH groups of the intermediate that turns carbon Cζ more electrophilic. In the end of this reaction, another tetrahedral intermediate is generated (Fig. 24.7d), still bonded to CYS398 but now with a hydroxyl group attached to carbon Cζ instead of an amino group that was found after step 1. Step 5 involves the cleavage of the covalent bond between CYS398 and the substrate. This requires the proton transfer from hydroxyl group of the reaction intermediate to CYS398. Simultaneously, occurs the proton transfer from GLU213 to HIS296. In the end of this reaction, L-citrulline is generated and is ready to leave the active site. The final step of the reaction (Step 6) involves the enzymatic turnover that requires HIS398 to lose a proton to the solvent or any neighbor active site residue.

In recent years there has been an increased interest in ADI as a potential antineoplastic therapy since various studies have demonstrated disturbances in both ARG synthesis and metabolism in malignant processes. To date several types of cancer have already been shown to be sensitive to arginine deprivation via ADI, such as melanoma, lung cancer, renal cell carcinomas, and HCC (Liu et al., 2014). These tumors do not express ASS and are therefore highly sensitive to ARG deprivation and optimal targets for ADI chemotherapies. In spite of the promising results, the application of these treatments presents some limitations. The therapeutic efficiency of ADI is limited to tumors that do not express ASS and/or have an inactivated L-citrulline to ARG recycling pathway (Ensor et al., 2002; Qiu et al., 2014). In addition, ADI has a short serum half-life and when used in prolonged treatments is highly immunogenic. In order to overcome some of these limitations a pegylated form of ADI was formulated, PEG(20)-ADI (molecular weight: 20 kDa), that serves to reduce the immunogenicity of the enzyme while greatly improving its pharmacokinetic half-life in serum (Feun and Savaraj, 2006; Feun et al., 2008; Ni et al., 2008). This enzyme revealed a similar efficiency in in vitro assays with melanomas and HCC, comparing with native ADI, however, it shows a much more effectiveness in in vivo assays (Ensor et al., 2002). Currently, the antitumor activity of PEG(20)-ADI has been also observed in pancreatic (Bowles et al., 2008), prostate (Kim et al., 2009), small cell lung (Kelly et al., 2012), head, neck (Huang et al., 2012), and breast cancers (Qiu et al., 2014), lymphoma (Delage et al., 2012; Huang et al., 2012), myxofibrosarcomas (Huang et al., 2013), melanoma (Feun et al., 2012), and glioblastoma (Qiu et al., 2014; Syed et al., 2013).

The therapeutic efficacy of PEG(20)-ADI has been validated through phase I/II clinical trials in advanced HCC (Glazer et al., 2010; Izzo et al., 2004; Yang et al., 2010) and melanoma patients (Ascierto et al., 2005; Feun et al., 2012; Ott et al., 2013). Additionally, there are several ongoing phase I and/or II clinical trials using PEG20-ADI in patients with advanced non-Hodgkin's lymphoma, acute myeloid leukemia, malignant pleural mesothelioma, prostate and nonsmall cell lung, advanced gastrointestinal, HER2 (Human Epidermal growth factor Receptor 2) negative metastatic breast, advanced pancreatic, and small cell lung cancers.

PEG(20)-ADI has also been reported as a potentially better therapeutic agent for the treatment of leukemia than ASNase, which has been used for the treatment of acute leukemia for over 20 years. ADI can inhibit the growth of cultured leukemia cells at concentrations of 5–10 ng/mL, which are about 20–100 times lower than those of ASNase (Feun et al., 2008). In addition, in vivo studies showed that ADI treatment does not have serious side effects, such as anaphylactic shock, coagulopathies, and liver toxicity, which have been reported for ASNase (Vrooman et al., 2010).

2.3 L-METHIONINE

L-MET is a sulfur containing amino acid that plays many key roles in cells (Guedes et al., 2011; Nozaki et al., 2005). In every cell, MET is essential for protein synthesis and it is the precursor of many compounds, such as Glutathione (a tripeptide that reduces reactive oxygen species, thereby protecting cells from oxidative stress), polyamines spermine and spermidine (which are detrimental on nuclear and cell division) and S-adenosylmethionine (the major source of methyl groups in a myriad of biological and biochemical events) (Cellarier et al., 2003; Davis and Uthus, 2004; Ho et al., 2011; Thivat et al., 2007). One of the most relevant biological role of MET in the cell is DNA methylation, where it serves as precursor for S-adenosylmethionine synthesis. DNA methylation is a common epigenetic signalling tool that cells use to control gene expression and has been associated with a number of key processes including genomic imprinting, X-chromosome inactivation, repression of repetitive elements, and carcinogenesis (Geiman and Muegge, 2010; Ramani et al., 2010).

In spite of its importance in cells, MET is a special essential amino acid as its long-term dietary deprivation is still compatible with life. This happens because unlike other essential amino acids that must be ingested, MET can be recycled by remethylation of homocysteine (de novo synthesis), either by L-methionine synthase (MetS) or in the liver by betaine-homocysteine methyltransferase (BHMT) (Cellarier et al., 2003; Hoffman, 1984; Guo et al., 1993a). In addition, MET can also be obtained by a *salvage* pathway in which MET is generated from 5′-methylthioadenosine toward methylthioadenosine phosphorylase (MTAP). This is true for all normally functioning cells, however, numerous malignant cell lines, such as breast, lung, colon, kidney and bladder cancers, melanoma, and glioblastoma (Breillout et al., 1990; Cellarier et al., 2003; Hoffman, 1985; Kokkinakis et al., 2001; Poirson-Bichat et al., 2000) do not have a functional methionine cycle intact. As a result, they cannot synthesize MET and in the absence of an exogenous supply they experience growth inhibition or eventually die.

The mechanisms responsible for MET dependence in malignant cell lines are still not fully understood. Several studies have shown that MET-dependent tumor cell lines presented relatively low amounts of L-methionine synthase or/and have deficient methionine synthase activity (Christa et al., 1986; Fiskerstrand et al., 1994; Kenyon et al., 2002). Other studies emphasized the importance of the inhibition of the salvage MET pathway in which MTAP is a key enzyme and is also responsible for the production of polyamines, which are critical for cell proliferation. Indeed, the loss of MTAP activity through gene deletion is a common feature in many kinds of human cancers, including non-small cell lung cancer, glioma, T-cell acute leukemia, bladder cancer, osteosarcoma and endometrial cancer (Albers, 2009; Nobori et al., 1993; Subhi et al., 2003; Traweek et al., 1988). In contrast, other studies suggest that MET dependence would result not from an enzymatic deficiency but rather on an increased requirement for this amino acid. In fact, MET requirements of tumor cells are higher than that of normal cells because of increased protein synthesis and DNA trans-methylation reactions. This promotes multiple biochemical reactions necessary for fast growth of tumor cells, and can change DNA expression. Indeed, in several tumors, a variety of growth inhibitory and proaapoptotic genes are transcriptionally silenced as a result of DNA hypermethylation (Baylin et al., 1998; Zingg and Jones, 1997).

The observation that some human tumors are MET-dependent in in vivo presents a therapeutic target for cancer growth control and has been studied in great detail in the past 2 decades (Hoshiya et al., 1995). As MET is sourced mainly from diet, initial studies observed that simple methionine dietary restriction caused regression of a variety of animal tumors and inhibits metastasis in animal models (Breillout et al., 1987, 1990; Goseki et al., 1987; Guo et al., 1993b; Millis et al., 1998). Phase

I clinical trials of dietary methionine restriction for adults with advanced cancer indicate that dietary methionine restriction is safe and provided promising results. However, reduction of methionine levels in humans solely by dietary intervention presents several limitations as MET is practically ubiquitous for all food and the price of such diet is about US$1000 per month.

A more attractive pharmacological approach to lowering MET levels in vitro and in vivo involves the use of MET degrading enzymes. L-methionine-α-amino-γ-mercaptoethane lyase or methioninase (MGL, EC 4.4.1.11) is an enzyme that specifically degrades L-methionine to α-ketobutyrate, methanethiol and ammonia (Kreis and Hession, 1973) very efficiently. In addition, it can also degrade L-homocysteine, a key factor that prevents endogenous MET production (Sato and Nozaki, 2009).

MGL is found in bacterial species and in fungi, but it is absent in mammals (Sharma et al., 2014). This enzyme was originally purified from *Clostridium sporogene*, but later on it was found that MGL from *Pseudomonas putida* (Pp-MGL) yielded a more stable enzyme with a relatively low Km (Esaki and Soda, 1987). Due to problems in purification of Pp-MGL, the MGL gene, isolated from *P. putida*, is now cloned and expressed in great quantities from *Escherichia coli* (r-Pp-MGL) (Hori et al., 1996).

Pp-MGL has 389 to 441 amino acids, and it is found as a homotetramer with a molecular weight of about 149–173 kDa (Nakayama et al., 1984). The active site of this enzyme is located at the interface of the two neighboring subunits, each one containing one Pyridoxal 5'-phosphate (PLP) cofactor that is required for the catalytic process (Kudou et al., 2007a; Mamaeva et al., 2005; Motoshima et al., 2000; Toney, 2005). The cofactor is stabilized by six residues TYR59, ARG61, TYR114, CYS116, LYS240, and ASP241 that have been found to be important for catalysis (Fig. 24.8) (Kudou et al., 2007a).

MGL is classified into the γ-family of PLP-dependent enzymes. However, it possesses a key cysteine (CYS116) in the active site that is not conserved in other PLP γ-family enzymes. Mutagenic studies revealed that this active site CYS is essential for substrate specificity (McKie et al., 1998; Kudou et al., 2007a, 2008; Sato et al., 2008) and also for enzymatic activity (Inoue et al., 2000; Kudou et al., 2008).

The enzymatic mechanism of MGL is proposed based on the knowledge of other PLP γ-family enzymes and the enzymological analyses of *P. putida* wildtype and mutant forms (Cerqueira et al., 2011; Fukumoto et al., 2012; Kudou et al., 2007a; Oliveira et al., 2011). Similar to all PLP-dependent enzymes, in the steady state the PLP cofactor is covalently bonded to an active site LYS, forming what is generally known as an internal aldimine (Fig. 24.9a). When MET is available in the active site of MGL, it binds to PLP (Step 1) generating a geminal diamine (Fig. 24.9b). Afterward, the cleavage of the covalent bond between the PLP cofactor and the active site LYS211 takes place (Step 2) and the substrate becomes covalent bonded to the PLP cofactor through the primary amino group, forming an external aldimine (Fig. 24.9c). The next step (Step 3) involves the abstraction of one proton from carbon Cα of the reaction intermediate by an active site base resulting in the formation of a quinonoide intermediate (Fig. 24.9d). Later on, carbon 7 from PLP is protonated by an active site base (Step 4), from which results a ketamine intermediate (Fig. 24.9e). An active site base then abstracts a second proton from carbon Cβ (Step 5) and an enamine intermediate is generated (Fig. 24.9f) that triggers the γ-elimination (Step 6) and the generation of unprotonanted methanethiol (Fig. 24.9g). Methanethiol is subsequently protonated by the active site base and an aminocrotonate intermediate is obtained (Fig. 24.9h) (Step 7). In the final step of the catalytic process (Step 8), a water molecule mediates the α-ketobutyrate and ammonia generation and the concomitant release of methanethiol. At the same time, the enzymatic

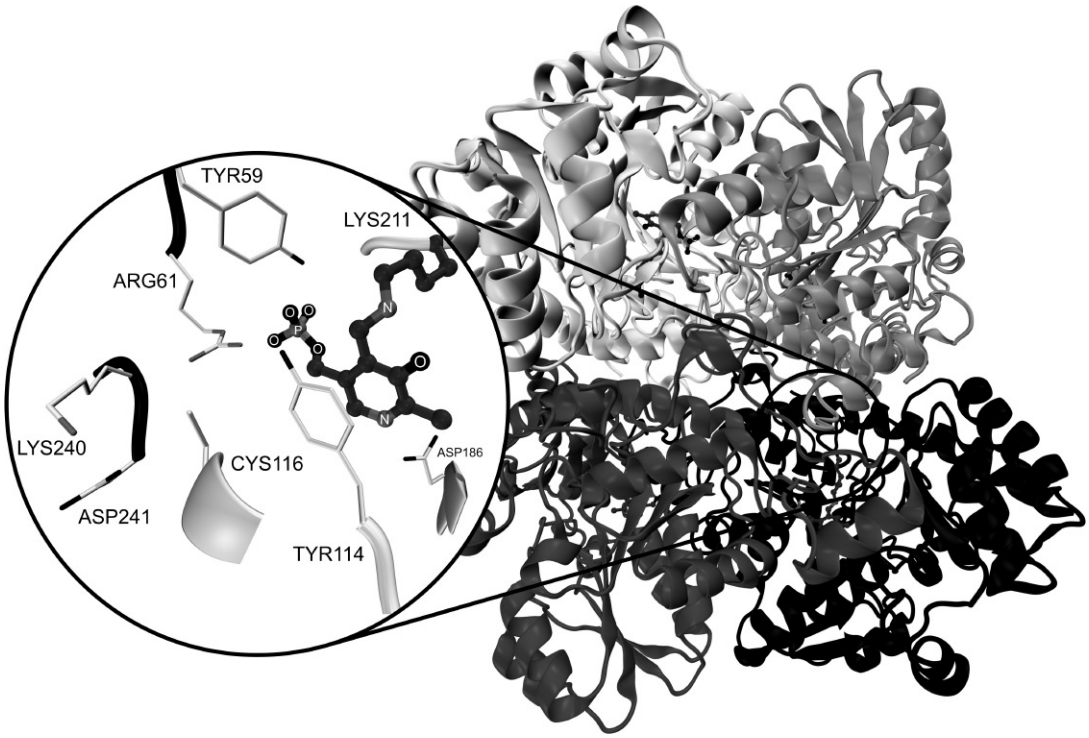


FIGURE 24.8 New Cartoon Representation of the Cocrystalized Structure of the MGL Homotetramer Structure With the PLP Cofactor (PDB Code: 2O7C)

Each subunit was represented with a different black tonality. In the close-up of the active site, the internal aldimine is shown in ball and sticks, where PLP cofactor covalent bonded to an active site LYS. The other active site residues that are important for the catalytic process are represented in licorice.

turnover takes place recovering the steady state of the enzyme, that is, internal aldimine in which the active site LYS is covalent bonded to the PLP cofactor. (Inoue et al., 2000; Kuznetsov et al., 2015; Sato et al., 2008; Toney, 2005).

The therapeutic manipulation of r-Pp-MGL to deplete plasma MET has been widely investigated in the past 2 decades (Jeanblanc et al., 2005) with promising results in a number of L-methionine-dependent cancer cell lines including lung, colon, kidney, brain and prostate cancers, melanoma, and fibrosarcoma (Miki et al., 2000a, 2000b, 2001; Smiraglia et al., 2001; Tan et al., 1996, 1997, 1999; Yamamoto et al., 2003). Preclinical studies in animal models have also demonstrated the potential of r-Pp-MGL to efficiently inhibited the growth of Yoshida sarcomas and a human lung tumor in nude mice xenograft models (Tan et al., 1996).

Several studies have also shown promising results combining the chemotherapeutic effect of r-Pp-MGL with other compounds, such as cisplatin, 5-fluorouracil (5-FU), 1-3-bis(2-chloroethyl)-1-nitrosourea (BCNU), cisplatin, and vincristine (Tan et al., 1999). These studies have shown synergic effects in mouse models of colon, lung, and brain cancers and glioblastoma (Hoshiya et al., 1997;

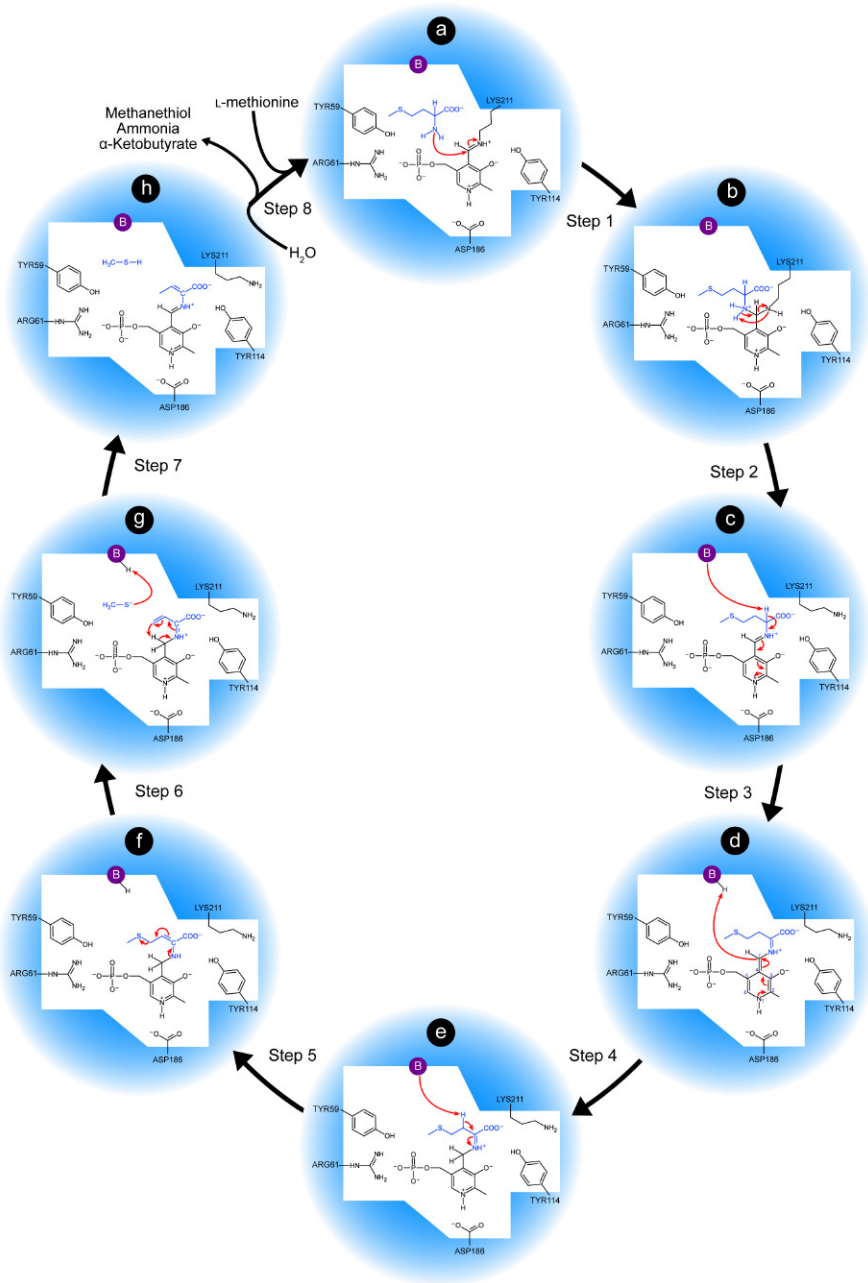


FIGURE 24.9 Schematic Representation of the MGL Catalysis Mechanism

Each circle represents a different reaction intermediate. The important reaction intermediates are labeled (a–h).

Hu and Cheung, 2009; Kokkinakis et al., 2001; Tan et al., 1999; Yoshioka et al., 1998). Other authors have also studied the administration of r-Pp-MGL with selenomethionine (a nontoxic compound). These studies have shown that r-Pp-MGL catalyses the conversion of selenomethionine into methylselenol, and the latter compound is able to oxidate thiol groups and generate toxic superoxide. This compound is very toxic to cells and can cause apoptosis via a mitochondrial pathway (Guo et al., 1993a; Kim et al., 2007). Because in cancer cells the metabolism is faster than in normal cells, this approach has the advantage of being target specific.

In spite of the initial success of r-Pp-MGL for cancer therapy it presents some problems that precludes an easy manipulation, namely, the instability of r-Pp-MGL in plasma caused by immunogenicity, proteolysis, oxidation of the active site residues, release of PLP and the instability of the active dimer (Kudou et al., 2007b). Modification of this enzyme with polyethylene glycol (PEG) has reduced immunogenicity, lengthened the half-life of the recombinant enzyme, and possibly inhibited proteolytic enzymes that could degrade MGL (Tan et al., 1998). The intravenous administration of PEG-r-Pp-MGL was tested in phase I clinical trials. Patients in these trials experienced no significant toxicity, and plasma MET levels fell dramatically. However, the authors did not assess the antitumor activity of this treatment and to date no other clinical trials have been published (Sun et al., 2003; Yang et al., 2004b, 2004c).

3 CONCLUSIONS

A prerequisite for making an effective medication in the treatment of cancer is that some fundamental difference between normal cells and cancer cells must be defined. An optimal chemotherapeutic agent must thus exploit this cellular difference in such a way that normal cells are spared and only cancer cells are injured. In this context, the amino acid depriving enzymes are promising anticancer drugs that have proved to be active and very specific against various types of cancers. Their mode of action is simple: they decrease the concentration of certain amino acids in the bloodstream and thus impair the development or even destroy tumor cells that are autotrophic for those amino acids. Normal cells remain unaltered since they are less demanding and/or can synthesize these compounds in sufficient amounts by other mechanisms.

ASNase is probably the best-known example of amino acid depletion chemotherapies using enzymes. The application of this enzyme in the treatment of childhood acute lymphoblastic leukemia has already been approved by FDA and shows very good results. ARG deprivation by ADI is a novel approach to target tumors that lack argininosuccinate synthetase (ASS) expression. ADI showed promising results during initial clinical trials for the therapy of HCC and melanomas and has currently been granted “orphan status” by the FDA and EMEA for the potential treatment of these diseases (Shen and Shen, 2006). Some studies also suggest that ADI has the potential to become a better therapeutic agent than ASNase in the treatment of leukemia (Gong et al., 2000; Muller and Boos, 1998).

Currently, the major drawback of ADI and ASNase for therapeutic purposes is the heterologous origin. As these enzymes are derived from a microorganism, they are immunogenic in humans and have very short circulating half-life. The chemotherapeutic treatments using these enzymes thus require multiple injections to produce the desired effect and this results in a variety of adverse effects including hypersensitivity reactions, anaphylactic shock, and the inactivation and clearance of the enzyme itself

(Schellekens, 2002). In order to overcome these problems and allow a safer application in cancer treatments, these enzymes have been linked to PEG to improve the pharmacokinetic and pharmacodynamic properties. The therapeutic results of these modified enzyme are much safer and have prompted the use of ADI in many clinical trials and the approval of ASNase by FDA and EMEA for the treatment of some types of tumors.

MGL is another heterologous enzyme discussed in this review. Similar to ASNase and ADI, this enzyme presents several immunogenic effects, but unlike those enzymes the pegylated form of MGL does not seem to overcome this problem. Several attempts were made to decrease the toxicity of this enzyme by the coadministration of pyridoxal 5-phosphate and oleic acid or dithiothreitol (Cantor et al., 2011; Sun et al., 2003). However, several side effects persist and its utilization in cancer therapies continues to be inappropriate (Miki et al., 2001; Yang et al., 2004a).

Among the discussed enzymes, ARGase is the only enzyme that is from human origin. This means that it presents low immunogenicity, a great advantage over ADI in the treatment of cancers that are autotrophic for ARG. Indeed, several authors consider ARGase a better chemotherapeutic agent for cancer treatment than ADI because apart from its low immunogenicity, this enzyme also is effective in tumors for which ADI is ineffective (although in these cases OCT cannot be expressed). Its major disadvantage when compared with ADI is its high Km and high optimum pH, which requires large quantities of enzyme in order to accomplish the expected therapeutic results. This problem may be overcome by the utilization of cobalt-substituted ARGase, but more studies are required to evaluate its potential in cancer therapy.

Despite the problems arising from the use of amino acid depleting enzymes in amino acid depriving chemotherapies, the recent progresses made in this field will undoubtedly facilitate the pursuit of new treatments to combat cancers. ASNase was the first enzyme approved by FDA and EMEA for the chemotherapeutic treatment of leukemia with remarkable success. Taking into account the ongoing clinical trials made with other enzymes, such as ARGase or ADI, and the promising results that are being obtained, one might expect that the clinical approval of these enzymes will not be too distant in the future and will have a profound influence on human health.

ACKNOWLEDGMENTS

This work has been funded by FEDER/COMPETE and the Fundação para a Ciência e a Tecnologia (FCT) through projects IF/01310/2013 and UID/Multi/04378/2013.

REFERENCES

Albers, E., 2009. Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5′-methylthioadenosine. IUBMB Life 61, 1132–1142.

Arena, M.E., Saguir, F.M., De Nadra, M.C.M., 1999. Arginine, citrulline, and ornithine metabolism by lactic acid bacteria from wine. Int. J. Food Microbiol. 52, 155–161.

Ascierto, P.A., Scala, S., Castello, G., Daponte, A., Simeone, E., Ottaiano, A., Beneduce, G., De Rosa, V., Izzo, F., Melucci, M.T., Ensor, C.M., Prestayko, A.W., Holtsberg, F.W., Bomalaski, J.S., Clark, M.A., Savaraj, N., Feun, L.G., Logan, T.F., 2005. Pegylated arginine deiminase treatment of patients with metastatic melanoma: results from phase I and II studies. J. Clin. Oncol. 23, 7660–7668.

Ash, D.E., Scolnick, L.R., Kanyo, Z.F., Vockley, J.G., Cederbaum, S.D., Christianson, D.W., 1998. Molecular basis of hyperargininemia: structure-function consequences of mutations in human liver arginase. Mol. Genet. Metab. 64, 243–249.

Ashikaga, T., Wang, Z., Yamamoto, M., Yamasaki, M., Magae, J., Nagai, K., 1994. Development and characterization of macrophage hybridomas derived from murine peritoneal exudate cells. Biosci. Biotechnol. Biochem. 58, 839–842.

Avramis, V.I., Sencer, S., Periclou, A.P., Sather, H., Bostrom, B.C., Cohen, L.J., Ettinger, A.G., Ettinger, L.J., Franklin, J., Gaynon, P.S., Hilden, J.M., Lange, B., Majlessipour, F., Mathew, P., Needle, M., Neglia, J., Reaman, G., Holcenberg, J.S., Stork, L., 2002. A randomized comparison of native *Escherichia coli* asparaginase and polyethylene glycol conjugated asparaginase for treatment of children with newly diagnosed standard-risk acute lymphoblastic leukemia: a Children's Cancer Group study. Blood 99, 1986–1994.

Barbul, A., Rettura, G., Levenson, S.M., Seifter, E., 1977. Arginine: a thymotropic and wound-healing promoting agent. Surg. Forum 28, 101–103.

Barbul, A., Rettura, G., Levenson, S.M., Seifter, E., 1983. Wound healing and thymotropic effects of arginine: a pituitary mechanism of action. Am. J. Clin. Nutr. 37, 786–794.

Baylin, S.B., Herman, J.G., Graff, J.R., Vertino, P.M., Issa, J.P., 1998. Alterations in DNA methylation: a fundamental aspect of neoplasia. Adv. Cancer Res. 72, 141–196.

Beloussow, K., Wang, L., Wu, J., Ann, D., Shen, W.C., 2002. Recombinant arginine deiminase as a potential antiangiogenic agent. Cancer Lett. 183, 155–162.

Beruter, J., Colombo, J.P., Bachmann, C., 1978. Purification and properties of arginase from human liver and erythrocytes. Biochem. J. 175, 449–454.

Billiar, T.R., 1995. Nitric oxide. novel biology with clinical relevance. Ann. Surg. 221, 339–349.

Bowles, T.L., Kim, R., Galante, J., Parsons, C.M., Virudachalam, S., Kung, H.J., Bold, R.J., 2008. Pancreatic cancer cell lines deficient in argininosuccinate synthetase are sensitive to arginine deprivation by arginine deiminase. Int. J. Cancer 123, 1950–1955.

Breillout, F., Hadida, F., Echinardgarin, P., Lascaux, V., Poupon, M.F., 1987. Decreased rat rhabdomyosarcoma pulmonary metastases in response to a low methionine diet. Anticancer Res. 7, 861–867.

Breillout, F., Antoine, E., Poupon, M.F., 1990. Methionine dependency of malignant-tumors: a possible approach for therapy. J. Nat. Cancer Inst. 82, 1628–1632.

Broome, J.D., 1961. Evidence that L-asparaginase activity of guinea pig serum is responsible for its antilymphoma effects. Nature 191, 1114.

Broome, J.D., 1963. Evidence that L-asparaginase of guinea pig serum is responsible for its antilymphoma effects. 1. Properties of L-asparaginase of guinea pig serum in relation to those of antilymphoma substance. J. Exp. Med. 118, 99.

Cantor, J.R., Yoo, T.H., Dixit, A., Iverson, B.L., Forsthuber, T.G., Georgiou, G., 2011. Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. Proc. Natl. Acad. Sci. USA 108, 1272–1277.

Cantor, J.R., Panayiotou, V., Agnello, G., Georgiou, G., Stone, E.M., 2012a. Engineering reduced-immunogenicity enzymes for amino acid depletion therapy in cancer. Methods Enzymol. 502, 291–319.

Cantor, J.R., Panayiotou, V., Agnello, G., Georgiou, G., Stones, E.M., 2012b. Engineering reduced-immunogenicity enzymes for amino acid depletion therapy in cancer. Methods Enzymol. 502, 291–319.

Casianocolon, A., Marquis, R.E., 1988. Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. Appl. Environ. Microbiol. 54, 1318–1324.

Castillo, L., Chapman, T.E., Sanchez, M., Yu, Y.M., Burke, J.F., Ajami, A.M., Vogt, J., Young, V.R., 1993. Plasma arginine and citrulline kinetics in adults given adequate and arginine-free diets. Proc. Natl. Acad. Sci. USA 90, 7749–7753.

Cavalli, R.C., Burke, C.J., Kawamoto, S., Soprano, D.R., Ash, D.E., 1994. Mutagenesis of rat liver arginase expressed in *Escherichia coli*: role of conserved histidines. Biochemistry 33, 10652–10657.

Cellarier, E., Durando, X., Vasson, M.P., Farges, M.C., Demiden, A., Maurizis, J.C., Madelmont, J.C., Chollet, P., 2003. Methionine dependency and cancer treatment. Cancer Treat. Rev. 29, 489–499.

Cerqueira, N.M.F.S.A., Fernandes, P.A., Ramos, M.J., 2011. Computational mechanistic studies addressed to the transimination reaction present in all pyridoxal 5′-phosphate-requiring enzymes. J. Chem. Theory Comput. 7, 1356–1368.

Cheng, P.N., Lam, T.L., Lam, W.M., Tsui, S.M., Cheng, A.W., Lo, W.H., Leung, Y.C., 2007. Pegylated recombinant human arginase (rhArg-peg5,000mw) inhibits the in vitro and in vivo proliferation of human hepatocellular carcinoma through arginine depletion. Cancer Res. 67, 309–317.

Christa, L., Kersual, J., Auge, J., Perignon, J.L., 1986. Salvage of 5′-deoxy-5′-methylthioadenosine and L-homocysteine into methionine in cells cultured in a methionine-free medium—a study of methionine-dependence. Biochem. Biophys. Res. Comm. 135, 131–138.

Christianson, D.W., Cox, J.D., 1999. Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. Ann. Rev. Biochem. 68, 33–57.

Covini, D., Tardito, S., Bussolati, O., Chiarelli, L.R., Pasquetto, M.V., Digilio, R., Valentini, G., Scotti, C., 2012. Expanding targets for a metabolic therapy of cancer: L-asparaginase. Recent Pat. Anticancer Drug Discov. 7, 4–13.

Cox, J.D., Cama, E., Colleluori, D.M., Pethe, S., Boucher, J.L., Mansuy, D., Ash, D.E., Christianson, D.W., 2001. Mechanistic and metabolic inferences from the binding of substrate analogues and products to arginase. Biochemistry 40, 2689–2701.

Currie, G.A., Gyure, L., Cifuentes, L., 1979. Microenvironmental arginine depletion by macrophages in vivo. Br. J. Cancer 39, 613–620.

Das, K., Butler, G.H., Kwiatkowski, V., Clark, Jr., A.D., Yadav, P., Arnold, E., 2004. Crystal structures of arginine deiminase with covalent reaction intermediates; implications for catalytic mechanism. Structure 12, 657–667.

Davis, C.D., Uthus, E.O., 2004. DNA methylation, cancer susceptibility, and nutrient interactions. Exp. Biol. Med. 229, 988–995.

Dejong, P.J., 1972. L-asparaginase production by streptomyces-griseus. Appl. Microbiol. 23, 1163.

Delage, B., Luong, P., Maharaj, L., O'riain, C., Syed, N., Crook, T., Hatzimichael, E., Papoudou-Bai, A., Mitchell, T.J., Whittaker, S.J., Cerio, R., Gribben, J., Lemoine, N., Bomalaski, J., Li, C.F., Joel, S., Fitzgibbon, J., Chen, L.T., Szlosarek, P.W., 2012. Promoter methylation of argininosuccinate synthetase-1 sensitises lymphomas to arginine deiminase treatment, autophagy and caspase-dependent apoptosis. Cell Death Dis. 3, e342.

Dillon, B.J., Prieto, V.G., Curley, S.A., Ensor, C.M., Holtsberg, F.W., Bomalaski, J.S., Clark, M.A., 2004. Incidence and distribution of argininosuccinate synthetase deficiency in human cancers: a method for identifying cancers sensitive to arginine deprivation. Cancer 100, 826–833.

Dizikes, G.J., Grody, W.W., Kern, R.M., Cederbaum, S.D., 1986. Isolation of human liver arginase cDNA and demonstration of nonhomology between the two human arginase genes. Biochem. Biophys. Res. Comm. 141, 53–59.

Dolowy, W.C., Henson, D., Cornet, J.A., Sellin, H., 1966. Toxic and antineoplastic effects of L-asparaginase: study of mice with lymphoma and normal monkeys and report on a child with leukemia. Cancer 19, 1813.

Ensor, C.M., Holtsberg, F.W., Bomalaski, J.S., Clark, M.A., 2002. Pegylated arginine deiminase (ADI-SS PEG20,000 mw) inhibits human melanomas and hepatocellular carcinomas in vitro and in vivo. Cancer Res. 62, 5443–5450.

Esaki, N., Soda, K., 1987. L-methionine gamma-lyase from *Pseudomonas putida* and *Aeromonas*. Methods Enzymol. 143, 459–465.

Feun, L., Savaraj, N., 2006. Pegylated arginine deiminase: a novel anticancer enzyme agent. Exp. Opin. Investig. Drugs 15, 815–822.

Feun, L., You, M., Wu, C.J., Kuo, M.T., Wangpaichitr, M., Spector, S., Savaraj, N., 2008. Arginine deprivation as a targeted therapy for cancer. Curr. Pharm. Des. 14, 1049–1057.

Feun, L.G., Marini, A., Walker, G., Elgart, G., Moffat, F., Rodgers, S.E., Wu, C.J., You, M., Wangpaichitr, M., Kuo, M.T., Sisson, W., Jungbluth, A.A., Bomalaski, J., Savaraj, N., 2012. Negative argininosuccinate synthetase expression in melanoma tumors may predict clinical benefit from arginine-depleting therapy with pegylated arginine deiminase. Br. J. Cancer 106, 1481–1485.

Feun, L.G., Kuo, M.T., Savaraj, N., 2015. Arginine deprivation in cancer therapy. Curr. Opin. Clin. Nutr. Metab. Care 18, 78–82.

Fiskerstrand, T., Christensen, B., Tysnes, O.B., Ueland, P.M., Refsum, H., 1994. Development and reversion of methionine dependence in a human glioma cell-line: relation to homocysteine remethylation and cobalamin status. Cancer Res. 54, 4899–4906.

Fukumoto, M., Kudou, D., Murano, S., Shiba, T., Sato, D., Tamura, T., Harada, S., Inagaki, K., 2012. The role of amino acid residues in the active site of L-methionine γ-lyase from *Pseudomonas putida*. Biosci. Biotechnol. Biochem. 76, 1275–1284.

Galkin, A., Lu, X., Dunaway-Mariano, D., Herzberg, O., 2005. Crystal structures representing the Michaelis complex and the thiouronium reaction intermediate of *Pseudomonas aeruginosa* arginine deiminase. J. Biol. Chem. 280, 34080–34087.

Geiman, T.M., Muegge, K., 2010. DNA methylation in early development. Mol. Reprod. Dev. 77, 105–113.

Gesto, D.S., Cerqueira, N.M.F.S.A., Fernandes, P.A., Ramos, M.J., 2013. Unraveling the enigmatic mechanism of L-asparaginase II with QM/QM calculations. J. Am. Chem. Soc. 135, 7146–7158.

Glazer, E.S., Piccirillo, M., Albino, V., Di Giacomo, R., Palaia, R., Mastro, A.A., Beneduce, G., Castello, G., De Rosa, V., Petrillo, A., Ascierto, P.A., Curley, S.A., Izzo, F., 2010. Phase II study of pegylated arginine deiminase for nonresectable and metastatic hepatocellular carcinoma. J. Clin. Oncol. 28, 2220–2226.

Glazer, E.S., Stone, E.M., Zhu, C., Massey, K.L., Hamir, A.N., Curley, S.A., 2011. Bioengineered human arginase I with enhanced activity and stability controls hepatocellular and pancreatic carcinoma xenografts. Trans. Oncol. 4, 138–146.

Gong, H., Zolzer, F., Von Recklinghausen, G., Havers, W., Schweigerer, L., 2000. Arginine deiminase inhibits proliferation of human leukemia cells more potently than asparaginase by inducing cell cycle arrest and apoptosis. Leukemia 14, 826–829.

Goseki, N., Onodera, T., Koike, M., Kosaki, G., 1987. Inhibitory effect of L-methionine-deprived amino-acid imbalance using total parenteral-nutrition on growth of ascites hepatoma in rats. Tohoku J. Exp. Med. 151, 191–200.

Guedes, R.L., Prosdocimi, F., Fernandes, G.R., Moura, L.K., Ribeiro, H.A., Ortega, J.M., 2011. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. BMC Genomics 4 (Suppl. 12), S2.

Guo, H., Lishko, V.K., Herrera, H., Groce, A., Kubota, T., Hoffman, R.M., 1993a. Therapeutic tumor-specific cell cycle block induced by methionine starvation in vivo. Cancer Res. 53, 5676–5679.

Guo, H.Y., Lishko, V.K., Herrera, H., Groce, A., Kubota, T., Hoffman, R.M., 1993b. Therapeutic tumor-specific cell-cycle block induced by methionine starvation in-vivo. Cancer Res. 53, 5676–5679.

Haines, R.J., Pendleton, L.C., Eichler, D.C., 2011. Argininosuccinate synthase: at the center of arginine metabolism. J. Biochem. Mol. Biol. 2, 8–23.

Henriksen, L.T., Harila-Saari, A., Ruud, E., Abrahamsson, J., Pruunsild, K., Vaitkeviciene, G., Jonsson, O.G., Schmiegelow, K., Heyman, M., Schroder, H., Albertsen, B.K., Nordic Society of Paediatric, H., Oncology, G., 2015. PEG-asparaginase allergy in children with acute lymphoblastic leukemia in the NOPHO ALL2008 protocol. Pediatr. Blood Cancer 62, 427–433.

Hernandez, C.P., Morrow, K., Lopez-Barcons, L.A., Zabaleta, J., Sierra, R., Velasco, C., Cole, J., Rodriguez, P.C., 2010. Pegylated arginase I: a potential therapeutic approach in T-ALL. Blood 115, 5214–5221.

Herzfeld, A., Raper, S.M., 1976. The heterogeneity of arginases in rat tissues. Biochem. J. 153, 469–478.

Hill, J.M., Roberts, J., Loeb, E., Khan, A., Maclellan, A., Hill, R.W., 1967. L-asparaginase therapy for leukemia and other malignant neoplasms: remission in human leukemia. JAMA 202, 882–888.

Ho, E., Beaver, L.M., Williams, D.E., Dashwood, R.H., 2011. Dietary factors and epigenetic regulation for prostate cancer prevention. Adv. Nutr. 2, 497–510.

Hoffman, R.M., 1984. Altered methionine metabolism: DNA methylation and oncogene expression in carcinogenesis: a review and synthesis. Biochim. Biophys. Acta 738, 49–87.

Hoffman, R.M., 1985. Altered methionine metabolism and transmethylation in cancer. Anticancer Res. 5, 1–30.

Hori, H., Takabayashi, K., Orvis, L., Carson, D.A., Nobori, T., 1996. Gene cloning and characterization of *Pseudomonas putida* L-methionine-alpha-deamino-gamma-mercaptomethane-lyase. Cancer Res. 56, 2116–2122.

Horn, F., 1933. The breakdown of arginine to citrulline by *Bacillus pyocyaneus*. Hoppe-Seylers Z. Physiol. Chem. 216, 244–247.

Hoshiya, Y., Guo, H., Kubota, T., Inada, T., Asanuma, F., Yamada, Y., Koh, J., Kitajima, M., Hoffman, R.M., 1995. Human tumors are methionine dependent in vivo. Anticancer Res. 15, 717–718.

Hoshiya, Y., Kubota, T., Inada, T., Kitajima, M., Hoffman, R.M., 1997. Methionine-depletion modulates the efficacy of 5-fluorouracil in human gastric cancer in nude mice. Anticancer Res. 17, 4371–4375.

Hu, J., Cheung, N.-K.V., 2009. Methionine depletion with recombinant methioninase: In vitro and in vivo efficacy against neuroblastoma and its synergism with chemotherapeutic drugs. Int. J. Cancer 124, 1700–1706.

Huang, C.C., Tsai, S.T., Kuo, C.C., Chang, J.S., Jin, Y.T., Chang, J.Y., Hsiao, J.R., 2012. Arginine deprivation as a new treatment strategy for head and neck cancer. Oral Oncol. 48, 1227–1235.

Huang, H.Y., Wu, W.R., Wang, Y.H., Wang, J.W., Fang, F.M., Tsai, J.W., Li, S.H., Hung, H.C., Yu, S.C., Lan, J., Shiue, Y.L., Hsing, C.H., Chen, L.T., Le, C.F., 2013. ASS1 as a novel tumor suppressor gene in myxofibrosarcomas: aberrant loss via epigenetic DNA methylation confers aggressive phenotypes, negative prognostic impact, and therapeutic relevance. Clin. Cancer Res. 19, 2861–2872.

Ilies, M., Di Costanzo, L., Dowling, D.P., Thorn, K.J., Christianson, D.W., 2011. Binding of alpha,alpha-disubstituted amino acids to arginase suggests new avenues for inhibitor design. J. Med. Chem. 54, 5432–5443.

Inoue, H., Inagaki, K., Adachi, N., Tamura, T., Esaki, N., Soda, K., Tanaka, H., 2000. Role of tyrosine 114 of L-methionine gamma-lyase from *Pseudomonas putida*. Biosci. Biotechnol. Biochem. 64, 2336–2343.

Izzo, F., Marra, P., Beneduce, G., Castello, G., Vallone, P., De Rosa, V., Cremona, F., Ensor, C.M., Holtsberg, F.W., Bomalaski, J.S., Clark, M.A., Ng, C., Curley, S.A., 2004. Pegylated arginine deiminase treatment of patients with unresectable hepatocellular carcinoma: results from phase I/II studies. J. Clin. Oncol. 22, 1815–1822.

Jeanblanc, M., Mousli, M., Hopfner, R., Bathami, K., Martinet, N., Abbady, A.Q., Siffert, J.C., Mathieu, E., Muller, C.D., Bronner, C., 2005. The retinoblastoma gene and its product are targeted by ICBP90: a key mechanism in the G1/S transition during the cell cycle. Oncogene 24, 7337–7345.

Jenkinson, C.P., Grody, W.W., Cederbaum, S.D., 1996. Comparative properties of arginases. Comp. Biochem. Physiol. 114, 107–132.

Kafkewitz, D., Goodman, D., 1974. L-asparaginase production by the rumen anaerobe *Vibrio succinogenes*. Appl. Microbiol. 27, 206–209.

Kanyo, Z.F., Scolnick, L.R., Ash, D.E., Christianson, D.W., 1996. Structure of a unique binuclear manganese cluster in arginase. Nature 383, 554–557.

Keating, M.J., Holmes, R., Lerner, S., Ho, D.H., 1993. L-asparaginase and PEG asparaginase: past, present, and future. Leuk. Lymphoma (Suppl. 10), 153–157.

Kelly, M.P., Jungbluth, A.A., Wu, B.W., Bomalaski, J., Old, L.J., Ritter, G., 2012. Arginine deiminase PEG20 inhibits growth of small cell lung cancers lacking expression of argininosuccinate synthetase. Br. J. Cancer 106, 324–332.

Kenyon, S.H., Waterfield, C.J., Timbrell, J.A., Nicolaou, A., 2002. Methionine synthase activity and sulphur amino acid levels in the rat liver tumor cells HTC and Phi-1. Biochem. Pharm. 63, 381–391.

Khangulov, S.V., Sossong, Jr., T.M., Ash, D.E., Dismukes, G.C., 1998. L-arginine binding to liver arginase requires proton transfer to gateway residue His141 and coordination of the guanidinium group to the dimanganese(II,II) center. Biochemistry 37, 8539–8550.

Kidd, J.G., 1953. Regression of transplanted lymphomas induced in vivo by means of normal guinea pig serum. 1. Course of transplanted cancers of various kinds in mice and rats given guinea pig serum, horse serum, or rabbit serum. J. Exp. Med. 98, 565.

Kim, A., Oh, J.H., Park, J.M., Chung, A.S., 2007. Methylselenol generated from selenomethionine by methioninase downregulates integrin expression and induces caspase-mediated apoptosis of B16F10 melanoma cells. J. Cell Physiol. 212, 386–400.

Kim, R.H., Coates, J.M., Bowles, T.L., McNerney, G.P., Sutcliffe, J., Jung, J.U., Gandour-Edwards, R., Chuang, F.Y., Bold, R.J., Kung, H.J., 2009. Arginine deiminase as a novel therapy for prostate cancer induces autophagy and caspase-independent apoptosis. Cancer Res. 69, 700–708.

Klein, D., Morris, D.R., 1978. Increased arginase activity during lymphocyte mitogenesis. Biochem. Biophys. Res. Comm. 81, 199–204.

Kobrinsky, N.L., Sposto, R., Shah, N.R., Anderson, J.R., Delaat, C., Morse, M., Warkentin, P., Gilchrist, G.S., Cohen, M.D., Shina, D., Meadows, A.T., 2001. Outcomes of treatment of children and adolescents with recurrent non-Hodgkin's lymphoma and Hodgkin's disease with dexamethasone, etoposide, cisplatin, cytarabine, and L-asparaginase, maintenance chemotherapy, and transplantation: Children's Cancer Group Study CCG-5912. J. Clin. Oncol. 19, 2390–2396.

Kokkinakis, D.M., Hoffman, R.M., Frenkel, E.P., Wick, J.B., Han, Q.H., Xu, M.X., Tan, Y.Y., Schold, S.C., 2001. Synergy between methionine stress and chemotherapy in the treatment of brain tumor xenografts in athymic mice. Cancer Res. 61, 4017–4023.

Krebs, H., 1973. The discovery of the ornithine cycle of urea synthesis. Biochem. Educ. 1, 19–23.

Kreis, W., Hession, C., 1973. Isolation and purification of L-methionine-alpha-deamino-gamma-mercaptomethane-lyase (L-methioninase) from *Clostridium sporogenes*. Cancer Res. 33, 1862–1865.

Kudou, D., Misaki, S., Yamashita, M., Tamura, T., Takakura, T., Yoshioka, T., Yagi, S., Hoffman, R.M., Takimoto, A., Esaki, N., Inagaki, K., 2007a. Structure of the antitumor enzyme L-methionine gamma-lyase from *Pseudomonas putida* at 1.8 A resolution. J. Biochem. 141, 535–544.

Kudou, D., Misaki, S., Yamashita, M., Tamura, T., Takakura, T., Yoshioka, T., Yagi, S., Hoffman, R.M., Takimoto, A., Esaki, N., Inagaki, K., 2007b. Structure of the antitumor enzyme L-methionine gamma-lyase from pseudomonas putida at 1.8 angstrom resolution. J. Biochem. 141, 535–544.

Kudou, D., Misaki, S., Yamashita, M., Tamura, T., Esaki, N., Inagaki, K., 2008. The role of cysteine 116 in the active site of the antitumor enzyme L-methionine gamma-lyase from Pseudomonas putida. Biosci. Biotechnol. Biochem. 72, 1722–1730.

Kuznetsov, N.A., Faleev, N.G., Kuznetsova, A.A., Morozova, E.A., Revtovich, S.V., Anufrieva, N.V., Nikulin, A.D., Fedorova, O.S., Demidkina, T.V., 2015. Pre-steady-state kinetic and structural analysis of interaction of methionine gamma-lyase from Citrobacter freundii with inhibitors. J. Biol. Chem. 290, 671–681.

Lavulo, L.T., Sossong, T.M., Brigham-Burke, M.R., Doyle, M.L., Cox, J.D., Christianson, D.W., Ash, D.E., 2001. Subunit-subunit interactions in trimeric arginase: generation of active monomers by mutation of a single amino acid. J. Biol. Chem. 276, 14242–14248.

Leopoldini, M., Russo, N., Toscano, M., 2009. Determination of the catalytic pathway of a manganese arginase enzyme through density functional investigation. Chemistry 15, 8026–8036.

Liu, J., Ma, J., Wu, Z., Li, W., Zhang, D., Han, L., Wang, F., Reindl, K.M., Wu, E., Ma, Q., 2014. Arginine deiminase augments the chemosensitivity of argininosuccinate synthetase-deficient pancreatic cancer cells to gemcitabine via inhibition of NF-kappaB signaling. BMC Cancer 14, 686.

Lu, X., Galkin, A., Herzberg, O., Dunaway-Mariano, D., 2004. Arginine deiminase uses an active-site cysteine in nucleophilic catalysis of L-arginine hydrolysis. J. Am. Chem. Soc. 126, 5374–5375.

Lu, X., Li, L., Wu, R., Feng, X., Li, Z., Yang, H., Wang, C., Guo, H., Galkin, A., Herzberg, O., Mariano, P.S., Martin, B.M., Dunaway-Mariano, D., 2006. Kinetic analysis of *Pseudomonas aeruginosa* arginine deiminase mutants and alternate substrates provides insight into structural determinants of function. Biochemistry 45, 1162–1172.

Mamaeva, D.V., Morozova, E.A., Nikulin, A.D., Revtovich, S.V., Nikonov, S.V., Garber, M.B., Demidkina, T.V., 2005. Structure of *Citrobacter freundii* L-methionine gamma-lyase. Acta Crystallogr. Sect. F 61, 546–549.

McKie, A.E., Edlind, T., Walker, J., Mottram, J.C., Coombs, G.H., 1998. The primitive protozoon *Trichomonas vaginalis* contains two methionine gamma-lyase genes that encode members of the gamma-family of pyridoxal 5′-phosphate-dependent enzymes. J. Biol. Chem. 273, 5549–5556.

Miki, K., Xu, M., An, Z., Wang, X., Yang, M., Al-Refaie, W., Sun, X., Baranov, E., Tan, Y., Chishima, T., Shimada, H., Moossa, A.R., Hoffman, R.M., 2000a. Survival efficacy of the combination of the methioninase gene and methioninase in a lung cancer orthotopic model. Cancer Gene Ther. 7, 332–338.

Miki, K., Xu, M.X., An, Z.L., Wang, X.E., Yang, M., Al-Refaie, W., Sun, X.H., Baranov, E., Tan, Y.Y., Chishima, T., Shimada, H., Moossa, A.R., Hoffman, R.M., 2000b. Survival efficacy of the combination of the methioninase gene and methioninase in a lung cancer orthotopic model. Cancer Gene Ther. 7, 332–338.

Miki, K., Xu, M., Gupta, A., Ba, Y., Tan, Y., Al-Refaie, W., Bouvet, M., Makuuchi, M., Moossa, A.R., Hoffman, R.M., 2001. Methioninase cancer gene therapy with selenomethionine as suicide prodrug substrate. Cancer Res. 61, 6805–6810.

Millis, R.M., Diya, C.A., Reynolds, M.E., Dehkordi, O., Bond, V., 1998. Growth inhibition of subcutaneously transplanted hepatomas without cachexia by alteration of the dietary arginine-methionine balance. Nutr. Cancer 31, 49–55.

Morris, Jr., S.M., Bhamidipati, D., Kepka-Lenhart, D., 1997. Human type II arginase: sequence analysis and tissue-specific expression. Gene 193, 157–161.

Motoshima, H., Inagaki, K., Kumasaka, T., Furuichi, M., Inoue, H., Tamura, T., Esaki, N., Soda, K., Tanaka, N., Yamamoto, M., Tanaka, H., 2000. Crystal structure of the pyridoxal 5′-phosphate dependent L-methionine gamma-lyase from *Pseudomonas putida*. J. Biochem. 128, 349–354.

Muller, H.J., Boos, J., 1998. Use of L-asparaginase in childhood ALL. Crit. Rev. Oncol. Hematol. 28, 97–113.

Nakayama, T., Esaki, N., Sugie, K., Beresov, T.T., Tanaka, H., Soda, K., 1984. Purification of bacterial L-methionine gamma-lyase. Anal. Biochem. 138, 421–424.

Ni, Y., Schwaneberg, U., Sun, Z.H., 2008. Arginine deiminase, a potential antitumor drug. Cancer Lett. 261, 1–11.

Nobori, T., Szinai, I., Amox, D., Parker, B., Olopade, O.I., Buchhagen, D.L., Carson, D.A., 1993. Methylthioadenosine phosphorylase deficiency in human non-small cell lung cancers. Cancer Res. 53, 1098–1101.

Nozaki, T., Ali, V., Tokoro, M., 2005. Sulfur-containing amino acid metabolism in parasitic protozoa. Adv. Parasitol. 60, 1–99.

Oliveira, E.F., Cerqueira, N.M.F.S.A., Fernandes, P.A., Ramos, M.J., 2011. Mechanism of formation of the internal aldimine in pyridoxal 5′-phosphate-dependent enzymes. J. Am. Chem. Soc. 133, 15496–15505.

Ott, P.A., Carvajal, R.D., Pandit-Taskar, N., Jungbluth, A.A., Hoffman, E.W., Wu, B.W., Bomalaski, J.S., Venhaus, R., Pan, L., Old, L.J., Pavlick, A.C., Wolchok, J.D., 2013. Phase I/II study of pegylated arginine deiminase (ADI-PEG 20) in patients with advanced melanoma. Invest New Drugs 31, 425–434.

Park, I.S., Kang, S.W., Shin, Y.J., Chae, K.Y., Park, M.O., Kim, M.Y., Wheatley, D.N., Min, B.H., 2003. Arginine deiminase: a potential inhibitor of angiogenesis and tumor growth. Br. J. Cancer 89, 907–914.

Peterson, R.E., Ciegler, A., 1969. L-asparaginase production by various bacteria. Appl. Microbiol. 17, 929.

Poirson-Bichat, F., Goncalves, R.A.B., Miccoli, L., Dutrillaux, B., Poupon, M.F., 2000. Methionine depletion enhances the antitumoral efficacy of cytotoxic agents in drug-resistant human tumor xenografts. Clin. Cancer Res. 6, 643–653.

Qiu, F., Chen, Y.R., Liu, X., Chu, C.Y., Shen, L.J., Xu, J., Gaur, S., Forman, H.J., Zhang, H., Zheng, S., Yen, Y., Huang, J., Kung, H.J., Ann, D.K., 2014. Arginine starvation impairs mitochondrial respiratory function in ASS1-deficient breast cancer cells. Sci. Signal 7, ra31.

Qiu, F., Huang, J., Sui, M., 2015. Targeting arginine metabolism pathway to treat arginine-dependent cancers. Cancer Lett. 364, 1–7.

Ramani, K., Yang, H., Kuhlenkamp, J., Tomasi, L., Tsukamoto, H., Mato, J.M., Lu, S.C., 2010. Changes in the expression of methionine adenosyltransferase genes and S-adenosylmethionine homeostasis during hepatic stellate cell activation. Hepatology 51, 986–995.

Reczkowski, R.S., Ash, D.E., 1992. EPR evidence for binuclear manganese (II) centers in rat liver arginase. J. Am. Chem. Soc. 114, 10992–10994.

Reeds, P.J., 2000. Dispensable and indispensable amino acids for humans. J. Nutr. 130, 1835S–1840S.

Richards, N.G.J., Kilberg, M.S., 2006. Asparagine synthetase chemotherapy. Ann. Rev. Biochem. 75, 629–654.

Rogers, Q.R., Visek, W.J., 1985. Metabolic role of urea cycle intermediates: nutritional and clinical aspects: introduction. J. Nutr. 115, 505–508.

Rytting, M., 2010. Peg-asparaginase for acute lymphoblastic leukemia. Exp. Opin. Biol. Ther. 10, 833–839.

Sarquis, M.I., Oliveira, E.M., Santos, A.S., Costa, G.L., 2004. Production of L-asparaginase by filamentous fungi. Mem. Inst. Oswaldo Cruz 99, 489–492.

Sato, D., Nozaki, T., 2009. Methionine gamma-lyase: the unique reaction mechanism, physiological roles, and therapeutic applications against infectious diseases and cancers. IUBMB Life 61, 1019–1028.

Sato, D., Yamagata, W., Harada, S., Nozaki, T., 2008. Kinetic characterization of methionine gamma-lyases from the enteric protozoan parasite *Entamoeba histolytica* against physiological substrates and trifluoromethionine, a promising lead compound against amoebiasis. FEBS J. 275, 548–560.

Savaraj, N., You, M., Wu, C., Wangpaichitr, M., Kuo, M.T., Feun, L.G., 2010. Arginine deprivation, autophagy, apoptosis (AAA) for the treatment of melanoma. Curr. Mol. Med. 10, 405–412.

Savoca, K.V., Abuchowski, A., Van Es, T., Davis, F.F., Palczuk, N.C., 1979. Preparation of a nonimmunogenic arginase by the covalent attachment of polyethylene glycol. Biochim. Biophys. Acta 578, 47–53.

Schellekens, H., 2002. Immunogenicity of therapeutic proteins: clinical implications and future prospects. Clin. Therap. 24, 1720–1740.

Scolnick, L.R., Kanyo, Z.F., Cavalli, R.C., Ash, D.E., Christianson, D.W., 1997. Altering the binuclear manganese cluster of arginase diminishes thermostability and catalytic function. Biochemistry 36, 10558–10565.

Sharma, B., Singh, S., Kanwar, S.S., 2014. L-methionase: a therapeutic enzyme to treat malignancies. Biomed. Res. Int. 2014, 506287.

Shen, L.J., Shen, W.C., 2006. Drug evaluation: ADI-PEG-20: a PEGylated arginine deiminase for arginine-auxotrophic cancers. Curr. Opin. Mol. Therap. 8, 240–248.

Shen, L.J., Lin, W.C., Beloussow, K., Shen, W.C., 2003. Resistance to the antiproliferative activity of recombinant arginine deiminase in cell culture correlates with the endogenous enzyme, argininosuccinate synthetase. Cancer Lett. 191, 165–170.

Simon-Reuss, I., 1953. Arginase, an antimitotic agent in tissue culture. Biochim. Biophys. Acta 11, 396–402.

Smiraglia, D.J., Rush, L.J., Fruhwald, M.C., Dai, Z., Held, W.A., Costello, J.F., Lang, J.C., Eng, C., Li, B., Wright, F.A., Caligiuri, M.A., Plass, C., 2001. Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. Hum. Mol. Genet. 10, 1413–1419.

Smith, D.W., Ganaway, R.L., Fahrney, D.E., 1978. Arginine deiminase from *Mycoplasma arthritidis*: structure-activity relationships among substrates and competitive inhibitors. J. Biol. Chem. 253, 6016–6020.

Spector, E.B., Rice, S.C., Cederbaum, S.D., 1983. Immunologic studies of arginase in tissues of normal human adult and arginase-deficient patients. Ped. Res. 17, 941–944.

Stone, E.M., Glazer, E.S., Chantranupong, L., Cherukuri, P., Breece, R.M., Tierney, D.L., Curley, S.A., Iverson, B.L., Georgiou, G., 2010. Replacing Mn(2+) with Co(2+) in human arginase i enhances cytotoxicity toward L-arginine auxotrophic cancer cell lines. ACS Chem. Biol. 5, 333–342.

Story, M.D., Voehringer, D.W., Stephens, L.C., Meyn, R.E., 1993. L-asparaginase kills lymphoma cells by apoptosis. Cancer Chem. Pharm. 32, 129–133.

Subhi, A.L., Diegelman, P., Porter, C.W., Tang, B., Lu, Z.J., Markham, G.D., Kruger, W.D., 2003. Methylthioadenosine phosphorylase regulates ornithine decarboxylase by production of downstream metabolites. J. Biol. Chem. 278, 49868–49873.

Sugimura, K., Ohno, T., Kimura, Y., Kimura, T., Azuma, I., 1992. Arginine deiminase gene of an AIDS-associated mycoplasma, *Mycoplasma incognitus*. Microbiol. Immunol. 36, 667–670.

Sun, X., Yang, Z., Li, S., Tan, Y., Zhang, N., Wang, X., Yagi, S., Yoshioka, T., Takimoto, A., Mitsushima, K., Suginaka, A., Frenkel, E.P., Hoffman, R.M., 2003. In vivo efficacy of recombinant methioninase is enhanced by the combination of polyethylene glycol conjugation and pyridoxal 5′-phosphate supplementation. Cancer Res. 63, 8377–8383.

Swain, A.L., Jaskolski, M., Housset, D., Rao, J.K.M., Wlodawer, A., 1993. Crystal-structure of *Escherichia-coli* L-Asparaginase, an enzyme used in cancertherapy. Proc. Natl. Acad. Sci. USA 90, 1474–1478.

Syed, N., Langer, J., Janczar, K., Singh, P., Lo Nigro, C., Lattanzio, L., Coley, H.M., Hatzimichael, E., Bomalaski, J., Szlosarek, P., Awad, M., O'neil, K., Roncaroli, F., Crook, T., 2013. Epigenetic status of argininosuccinate synthetase and argininosuccinate lyase modulates autophagy and cell death in glioblastoma. Cell Death Dis. 4, e458.

Takaku, H., Takase, M., Abe, S., Hayashi, H., Miyazaki, K., 1992. Invivo antitumor-activity of arginine deiminase purified from *Mycoplasma arginini*. Int. J. Cancer 51, 244–249.

Takaku, H., Matsumoto, M., Misawa, S., Miyazaki, K., 1995. Antitumor-activity of arginine deiminase from *Mycoplasma arginini* and its growth-inhibitory mechanism. Jap. J. Cancer Res. 86, 840–846.

Tan, Y., Xu, M., Guo, H., Sun, X., Kubota, T., Hoffman, R.M., 1996. Anticancer efficacy of methioninase in vivo. Anticancer Res. 16, 3931–3936.

Tan, Y., Xu, M., Tan, X., Tan, X., Wang, X., Saikawa, Y., Nagahama, T., Sun, X., Lenz, M., Hoffman, R.M., 1997. Overexpression and large-scale production of recombinant L-methionine-alpha-deamino-gamma-mercaptomethane-lyase for novel anticancer therapy. Protein Expr. Purif. 9, 233–245.

Tan, Y., Sun, X., Xu, M., An, Z., Tan, X., Han, Q., Miljkovic, D.A., Yang, M., Hoffman, R.M., 1998. Polyethylene glycol conjugation of recombinant methioninase for cancer therapy. Protein Expr. Purif. 12, 45–52.

Tan, Y.Y., Sun, X.H., Xu, M.X., Tan, X.Z., Sasson, A., Rashidi, B., Han, Q.H., Tan, X.Y., Wang, X.E., An, Z.L., Sun, F.X., Hoffman, R.M., 1999. Efficacy of recombinant methioninase in combination with cisplatin on human colon tumors in nude mice. Clin. Cancer Res. 5, 2157–2163.

Tan, Y.Y., Xu, M.X., Hoffman, R.M., 2010. Broad selective efficacy of recombinant methioninase and polyethylene glycol-modified recombinant methioninase on cancer cells in vitro. Anticancer Res. 30, 1041–1046.

Thivat, E., Durando, X., Demidem, A., Farges, M.C., Rapp, M., Cellarier, E., Guenin, S., D'incan, M., Vasson, M.P., Chollet, P., 2007. A methionine-free diet associated with nitrosourea treatment down-regulates methylguanine-DNA methyl transferase activity in patients with metastatic cancer. Anticancer Res. 27, 2779–2783.

Toney, M.D., 2005. Reaction specificity in pyridoxal phosphate enzymes. Arch. Biochem. Biophys. 433, 279–287.

Tosa, T., Sano, R., Yamamoto, K., Nakamura, M., Ando, K., Chibata, I., 1971. L-asparaginase from proteus-vulgaris. Appl. Microbiol. 22, 387.

Traweek, S.T., Riscoe, M.K., Ferro, A.J., Braziel, R.M., Magenis, R.E., Fitchen, J.H., 1988. Methylthioadenosine phosphorylase deficiency in acute leukemia: pathologic, cytogenetic, and clinical features. Blood 71, 1568–1573.

Vockley, J.G., Jenkinson, C.P., Shukla, H., Kern, R.M., Grody, W.W., Cederbaum, S.D., 1996. Cloning and characterization of the human type II arginase gene. Genomics 38, 118–123.

Vrooman, L.M., Supko, J.G., Neuberg, D.S., Asselin, B.L., Athale, U.H., Clavell, L., Kelly, K.M., Laverdiere, C., Michon, B., Schorin, M., Cohen, H.J., Sallan, S.E., Silverman, L.B., 2010. Erwinia asparaginase after allergy to *E. coli* asparaginase in children with acute lymphoblastic leukemia. Ped. Blood Cancer 54, 199–205.

Wheatley, D.N., Campbell, E., 2003. Arginine deprivation, growth inhibition and tumor cell death: 3. Deficient utilisation of citrulline by malignant cells. Br. J. Cancer 89, 573–576.

Wiswell, O.B., 1951. Effects of intraperitoneally injected arginase on growth of mammary carcinoma implants in the mouse. Proc. Soc. Exp. Biol. Med. 76, 588–589.

Woo, M.H., Hak, L.J., Storm, M.C., Evans, W.E., Sandlund, J.T., Rivera, G.K., Wang, B., Pui, C.H., Relling, M.V., 1998. Antiasparaginase antibodies following E. coli asparaginase therapy in pediatric acute lymphoblastic leukemia. Leukemia 12, 1527–1533.

Yamamoto, N., Gupta, A., Xu, M.X., Miki, K., Tsujimoto, Y., Tsuchiya, H., Tomita, K., Moossa, A.R., Hoffman, R.M., 2003. Methioninase gene therapy with selenomethionine induces apoptosis in bcl-2-overproducing lung cancer cells. Cancer Gene Ther. 10, 445–450.

Yang, Z., Wang, J., Yoshioka, T., Li, B., Lu, Q., Li, S., Sun, X., Tan, Y., Yagi, S., Frenkel, E.P., Hoffman, R.M., 2004a. Pharmacokinetics, methionine depletion, and antigenicity of recombinant methioninase in primates. Clin. Cancer Res. 10, 2131–2138.

Yang, Z.J., Sun, X.H., Li, S.K., Tan, Y.Y., Wang, X., Zhang, N., Yagi, S., Takakura, T., Kobayashi, Y., Takimoto, A., Yoshioka, T., Suginaka, A., Frenkel, E.P., Hoffman, R.M., 2004b. Circulating half-life of PEGylated recombinant methioninase holoenzyme is highly dose dependent on cofactor pyridoxal-5 ’-phosphate. Cancer Res. 64, 5775–5778.

Yang, Z.J., Wang, J.H., Lu, Q., Xu, J.B., Kobayashi, Y., Takakura, T., Takimoto, A., Yoshioka, T., Lian, C.G., Chen, C.M., Zhang, D.D., Zhang, Y., Li, S.K., Sun, X.H., Tan, Y.Y., Yagi, S., Frenkel, E.P., Hoffman, R.M., 2004c. PEGylation confers greatly extended half-life and attenuated immunogenicity to recombinant methioninase in primates. Cancer Res. 64, 6673–6678.

Yang, T.S., Lu, S.N., Chao, Y., Sheen, I.S., Lin, C.C., Wang, T.E., Chen, S.C., Wang, J.H., Liao, L.Y., Thomson, J.A., Wang-Peng, J., Chen, P.J., Chen, L.T., 2010. A randomized phase II study of pegylated arginine deiminase (ADI-PEG 20) in Asian advanced hepatocellular carcinoma patients. Br. J. Cancer 103, 954–960.

Yoon, C.Y., Shim, Y.J., Kim, E.H., Lee, J.H., Won, N.H., Kim, J.H., Park, I.S., Yoon, D.K., Min, B.H., 2007. Renal cell carcinoma does not express argininosuccinate synthetase and is highly sensitive to arginine deprivation via arginine deiminase. Int. J. Cancer 120, 897–905.

Yoshioka, T., Wada, T., Uchida, N., Maki, H., Yoshida, H., Ide, N., Kasai, H., Hojo, K., Shono, K., Maekawa, R., Yagi, S., Hoffman, R.M., Sugita, K., 1998. Anticancer efficacy in vivo and in vitro, synergy with 5-fluorouracil, and safety of recombinant methioninase. Cancer Res. 58, 2583–2587.

Zhang, L., Liu, M., Jamil, S., Han, R., Xu, G., Ni, Y., 2015. PEGylation and pharmacological characterization of a potential antitumor drug, an engineered arginine deiminase originated from *Pseudomonas plecoglossicida*. Cancer Lett. 357, 346–354.

Zingg, J.M., Jones, P.A., 1997. Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation, and carcinogenesis. Carcinogenesis 18, 869–882.

View publication stats

Jenny Stanford Series
on Biocatalysis
Volume 4

Pharmaceutical Biocatalysis

Fundamentals, Enzyme Inhibitors, and Enzymes in Health and Diseases

edited by Peter Grunwald

Chapter 8

Enzymatic Amino Acid Deprivation Therapies Targeting Cancer

Carla S. S. Teixeira, Henrique S. Fernandes, Sérgio F. Sousa, and Nuno M. F. S. A. Cerqueira

UCIBIO@REQUIMTE, BioSIM, Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Alameda Professor Hernâni Monteiro 4200-319, Porto, Portugal

nunoscerqueira@med.up.pt

8.1 Introduction

Proteins are large and complex molecules that play critical roles in the cells. They are involved in virtually all cell functions, including their regulation and structure. Any alteration on the concentration or composition of the proteins will disturb the function of the cells and can compromise cellular viability. In severe cases, this can ultimately lead to the death of the organism.

Proteins are made up of hundreds or thousands of smaller units called amino acids, which are attached to one another through peptide bonds in long chains. There are 20 different types of

Pharmaceutical Biocatalysis: Fundamentals, Enzyme Inhibitors, and Enzymes in Health and Diseases
Edited by Peter Grunwald
Copyright © 2019 Jenny Stanford Publishing Pte. Ltd.
ISBN 978-981-4800-61-7 (Hardcover), 978-0-429-29503-4 (eBook)
www.jennystanford.com

natural amino acids that can be combined and used by the cells to produce proteins. It is, therefore, important that these molecules are available in sufficient amount in the cells to maintain their normal function.

Most of the mammals, and particularly humans, have long ago ceased the capability to produce efficiently 10 of the 20 natural amino acids. This means that for the healthy development, growth, metabolic, and physiological functions of their cells, these compounds must be obtained from the diet. Otherwise, they would starve to death. These amino acids are termed essential and include isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine. This class of amino acids can also include histidine and arginine. Both histidine and arginine are non-essential in healthy adults but become essential when their requirement increases and they need to be obtained from the diet. This usually occurs during the periods of growth, illness, or metabolic stress. For this reason, they are called semi-essential amino acids. All the other amino acids are termed non-essential and are efficiently synthesized by the enzymes present in the cells (alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine, and tyrosine).

The metabolism of the amino acids has often been disregarded for clinical purposes. However, recent studies have shown that the amino acid can represent a great clinical opportunity, still poorly explored, for the treatment of several types of cancer that require high concentrations of one or more amino acid residues that are essential in their development, survival, and spread. In order to overcome these problems, tumours rely on the extracellular pool of these amino acids to satisfy protein biosynthesis demands and continue to grow without limitations. This means that if the concentration of those amino acids decreases, the development of tumour cells can be impaired or even annihilated. At the same time, healthy cells remain almost unchanged since they are less demanding (in the case on the essential amino acids) or can synthesize these compounds (non-essential or semi-essential amino acids) in sufficient amounts by alternative pathways.

The amino acid deprivation cancer therapy relies on these principles, and its main objective is to starve tumours to death

while keeping normal cells unharmed (Fung and Chan, 2017). The optimal conditions for this therapy thus occur when the tumours are sensitive to one of the non-essential or semi-essential amino acid residues. However, a simple diet is not enough to achieve a therapeutically relevant amino acid deprivation for cancer therapy, and therefore, other strategies are used for this purpose.

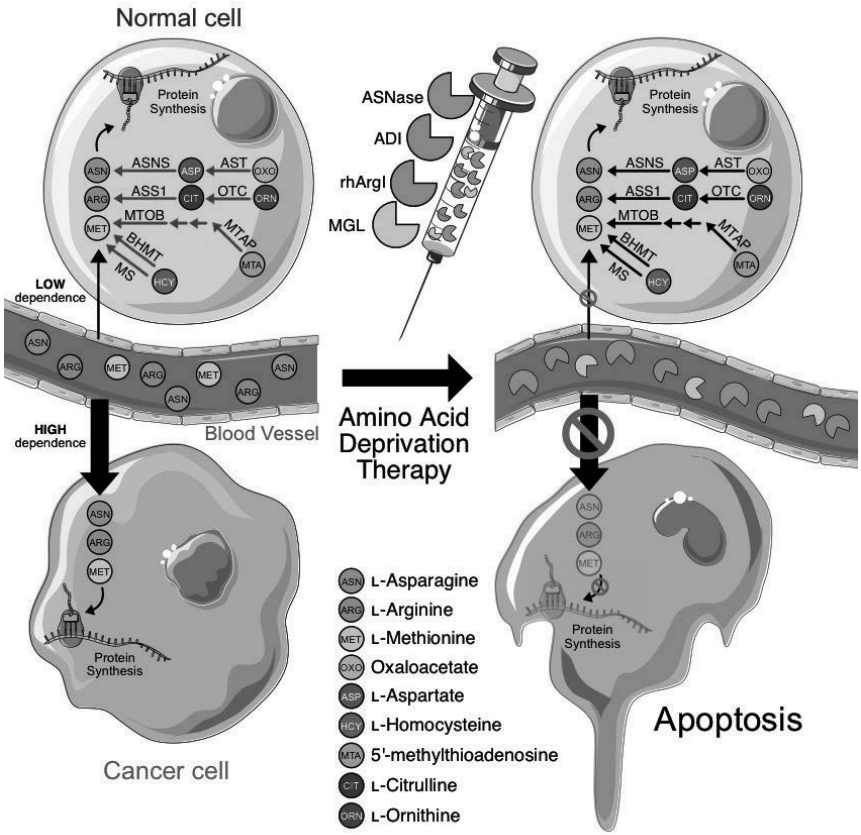


Figure 8.1 Schematic representation of the metabolic mechanisms involving the amino acid depletion therapies, using methionine-γ-lyase (METase), asparaginase (ASNase), arginase (rhARGI), and L-arginine deiminase (ADI).

During the last 50 years, enzymatic therapies have been found to be quite efficient for this purpose. These treatments involve the use of heterologous enzymes (from other organisms,

such as bacteria), recombinant, or engineered human enzymes to decrease the concentration of a specific amino acid in the bloodstream that is essential for the development of a tumour (Pokrovsky et al., 2017).

There are successful examples where the enzymatic amino acid deprivation therapies are used to treat cancer. One example involves the use of arginine deiminase (ADI) in hepatocellular carcinoma (HCC), melanoma, and other urea cycle-deficient cancer cells that have a high demand for L-arginine (semi-essential amino acid). Another example, involves the use of the enzyme methionine-γ-lyase (MGL) in the treatment of several types of tumours, such as, colon, breast, prostate, ovary, lung, brain, kidney, stomach, and bladder cancers, as well as larynx melanoma, sarcoma, leukaemia and lymphomas that require a high concentration of L-methionine (essential amino acid) (Fernandes et al., 2017a). The non-essential amino acids can also be used in enzymatic amino acid depletion therapies. This only happens in tumours where the normal metabolism of one non-essential amino acid in the cells is disrupted or when there is some sort of defect in the ability to use a certain amino acid (DeBerardinis and Chandel, 2016). This is the case of childhood acute lymphoblastic leukaemia (ALL), and to a lesser extent, of non-Hodgkin's lymphoma, whose tumours require a high concentration of L-asparagine. In these cases, the use of L-asparaginase (ASNase) has been found to be remarkably successful in the treatment of these tumours and this method has already been approved by the FDA.

Taking into account the therapeutic potential of these enzymes in cancer therapies, they have been extensively studied in the last two decades. These therapies have been showing very promising results, albeit some challenges have been reported in clinical trials. In the following sections, the structure, function, and catalytic mechanism of therapeutic heterologous enzymes involved in L-ARG, L-MET, and L-ASN degradation will be revised. Particular attention will be given to the heterologous enzymes L-arginase (rhARGI), L-arginine deiminase (ADI) (Ni et al., 2008), L-methionine-γ-lyase (MGL) (Tan et al., 1996), and L-asparaginase (ASNase) (Rytting, 2010) taking into account their current therapeutic potential.

8.2 Amino Acid Deprivation Enzymes

Eukaryotic cells control the concentration of the amino acids in the cell through a tight regulation of their concentration. From this control results a delicate state of dynamic equilibrium that is known as proteome homeostasis. A crucial step in the maintenance of the proteome homeostasis is the synthesis of proteins that are required for cell growth and proliferation. Therefore, any perturbation in this equilibrium may affect permanently cell metabolism and even its survival. The amino acid deprivation therapies have, therefore, a very narrow window to work with, since the decrease in the concentrations of any amino acid in the bloodstream needs to be carefully addressed in order to prevent any catastrophic effect on the normal cells and in the full organism. For this reason, the amino acid deprivation therapy cannot be used in all types of cancers that are autotrophic for a certain amino acid, but only in the ones where a small decrease in that amino acid can lead to a therapeutic effect while keeping the other cells safe. The following sections describe cases where these assumptions have been fully addressed.

8.2.1 Arginine

Arginine (L-ARG) is in different ways, an important amino acid for all living organisms. Besides its common role as a building block for protein synthesis, in the case of mammals, it is also a key precursor/mediator of a series of metabolic, immunological, neurological and signalling pathways in the cells (Barbul et al., 1977, 1983; Billiar, 1995; Rogers and Visek, 1985).

For adult mammals, L-ARG is considered a semi or a conditionally essential amino acid. This means that although for healthy individuals arginine is not an essential dietary amino acid, the emergence of a catabolic stress, such as an inflammation or infection, raises the metabolic demands to levels that cannot be supplied by the endogenous synthesis, turning it into an essential amino acid (Daly et al., 1988; Morris Jr., 2007; Reeds, 2000).

A healthy individual obtains the metabolic required L-ARG from the diet, endogenous de novo synthesis and cellular breakdown of proteins (Morris Jr., 2006). When deprived of L-ARG, healthy

cells undergo a cell cycle arrest at Go/G1 phase and become quiescent. This effect is reversible for the majority of the cells who recover to their normal proliferation status if reinstated with L-ARG [16].

In the case of tumour cells, its metabolism is so faulty that when deprived of arginine they continue the cell cycle, undergoing unbalanced growth, which eventually leads to the activation of apoptotic pathways (Patil et al., 2016). This particularity, common to all tumour cells, has been widely explored in the development of new anti-cancer therapies against some cancers with auxotrophy toward L-ARG (Dhankhar et al., 2018).

The susceptibility of some cancers to L-ARG deprivation was observed for more than 60 years (Bach and Swaine, 1965) although, by the time, the mechanism behind it was unknown. To date, there is evidence of two enzymes with proven ability to efficiently deprive L-ARG auxotrophic tumour cells from its essential amino acid: a Mycoplasma arginine deiminase (ADI) and a human recombinant arginase (rhArg). They have different catalytic mechanisms and are effective in cancer cells with different characteristics.

8.2.2 Arginine Deiminase

In adult mammals, 5 to 15% of the plasma arginine is obtained by de novo biosynthetic pathway, which involves the conversion of citrulline to L-ARG catalysed by argininosuccinate synthase (ASS1) and argininosuccinate lyase (ASL). ASS1 catalyzes the condensation of citrulline and aspartate to form argininosuccinate that is split into fumarate and L-ARG by ASL. ASS1 is a ubiquitous enzyme whose expression, localization, and regulation differ significantly depending on the tissue-specific needs for L-ARG (Haines et al., 2011).

Some cancers such as metastatic melanoma (MM), prostate carcinomas, hepatocellular carcinoma (HCC) (Dillon et al., 2004) non-Hodgkin's lymphoma, Hodgkin's lymphoma, pancreatic carcinoma, osteosarcoma, and malignant pleural mesothelioma and some breast tumours (Qiu et al., 2015) have an elevated requirement for L-ARG due to the deficient expression of ASS1 that results in L-ARG auxotrophy (Feun et al., 2008, 2012, 2015). Arginine deiminase (ADI,

EC 3.5.3.6) is an enzyme with high affinity to L-ARG that catalyses the conversion of L-ARG into citrulline and ammonia. The cancer cells who do not express ASS1 are unable to re-synthesize L-ARG from citrulline, which culminates in the full deprivation of an amino acid that is imperative for their survival (Shen et al., 2003).

ADI, together with ornithine transcarbamylase (OTC, EC 2.1.3.3) and carbamate kinase (CK, EC 2.7.2.2), belongs to the arginine deiminase pathway (ADI pathway), which is used by many L-ARG-dependent microorganisms to catabolize the amino acid–generating ATP. It has been identified, purified, and characterized in bacteria, archaea and some anaerobic eukaryotes (Horn, 1933; Knodler et al., 1994, 1995; Leopoldini et al., 2009), but there is no evidence of its existence in higher eukaryotes. It was demonstrated that ADI only possesses anti-tumour activity in ASS1 deficient tumours, being ineffective when ASS1 is restored (Ensor et al., 2002; Qiu et al., 2014). Further evidence suggests that ADI mediates the inhibition of tumour growth not only by exhausting the supplies of L-ARG, but also by its anti-angiogenic activity via suppression of NO generation (Beloussow et al., 2002; Park et al., 2003; Yoon et al., 2007).

The first evidence of the in vitro anti-tumour activity of ADI was reported in 1981 by Jones, who applied ADI isolated from *Pseudomonas putida* on murine leukaemia lymphoblast (Ni et al., 2008). A decade later, in 1990, Miyazaki and colleagues demonstrated that the addition of a small dose (5 ng/ml) of Mycoplasma ADI to the culture media depleted all the available L-ARG, which in turn lead to the inhibition of various human cancer cell lines (Miyazaki et al., 1990). From there onwards, several studies confirmed the potential use of Mycoplasma ADI as an anti-tumour agent in vitro and in vivo (Ashikaga et al., 1994; Sugimura et al., 1992; Takaku et al., 1992).

Although there are studies with other ADI sources (Zhang et al., 2015), the most commonly used ADI protein is derived from *M. arginini*. It is composed of two identical subunits with a molecular weight of 45 kDa each and an isoelectric point of 4.7. It is stable at neutral pH and exerted maximal enzyme activity at pH 6.0–7.5 and at 50°C (Takaku et al., 1992).

Since ADI from *M. arginini* has high immunogenicity and its half-life time is too short (approximately 5 h) to be directly used

in vivo, it was conjugated with polyethylene glycol (PEG) via a succinimidyl succinate amide bond (Takaku et al., 1993). A formulated ADI-PEG20 (molecular weight: 20 kDa) showed significantly reduced antigenicity and an increased half-life time from 5 h to 7 days in mice and rabbit. In addition, it was also demonstrated that it efficiently degraded plasma arginine to undetectable levels (<5–7 μmol/L) (Ensor et al., 2002; Holtsberg et al., 2002) leading to the L-ARG deprivation in humans (Izzo et al., 2004).

Several studies showed that ADI is a potentially better therapeutic agent for the treatment of leukaemia than ASNase, which has been used for the treatment of acute leukaemia for over 20 years (Ensor et al., 2002; Durden and Distasio, 1980). ADI inhibited the growth of cultured leukaemia cells at concentrations about 20–100 times lower and does not have the serious side effects associated with ASNase (Gong et al., 2000; Muller and Boos, 1998).

The anti-tumour activity of ADI-PEG20 has been further demonstrated through laboratory studies in HCC (Dillon et al., 2004), pancreatic (Bowles et al., 2008), prostate (Kim et al., 2009), small cell lung (Kelly et al., 2012), head and neck (Huang et al., 2012), and breast cancers (Qiu et al., 2014), lymphoma (Delage et al., 2012), myxofibrosarcomas (Huang et al., 2013), melanoma (Dillon et al., 2004; Feun et al., 2012), and glioblastoma (Syed et al., 2013).

8.2.2.1 Therapeutic uses

By 2018, there are records of more than 20 clinical trials in different phases (phase I, II, and III) and different recruiting status whose general objective is to evaluate the safety and therapeutic efficacy of ADI-PEG20 in patients suffering from different cancers (ClinicalTrials.gov. Available at: https://clinicaltrials.gov).

The first clinical trials were completed 15 years ago and showed that the plasma L-ARG levels decrease from approximately 130 μmol/L to below the level of detection (<2 μmol/L) and remain low for more than 7 days after the first intramuscular injection of 160 IU/m^2 of ADI-PEG20 in both HCC (NCT00056992) and MM patients (NCT00029900). The response rates of these studies are in the range 25–47%, and they demonstrated good safety and tolerability (Izzo et al., 2004; Ascierto et al., 2005).

All the posterior ADI-PEG20 clinical trials completed confirmed that monotherapy with ADI-PEG 20 in patients with both MM and HCC is generally well tolerated.

In a HCC trial with patients with metastatic or inoperable histologically confirmed HCC, it was observed that the L-ARG levels remained below baseline for 50 days and had an overall mean survival of 15.8 months. The authors also observed a low rate of significant immunogenic events (2.5%) suggesting good tolerability of ADI (Glazer et al., 2010).

Another phase II study in Asian patients with HCC confirmed that weekly treatment with 160 IU/m^2 was sufficient to deplete circulating L-ARG for 7 days. Additionally, patients with depletion of circulating L-ARG for <4 weeks had a trend towards better survival than those with depletion of arginine for <4 weeks (Yang et al., 2010).

Based on these encouraging results, a phase III trial (NCT01287585) was conducted with HCC patients who had failed prior chemotherapy versus placebo. The results were not as encouraging as expected. ADI-PEG20 monotherapy at the dose of 18 mg/m^2 proved to be ineffective in prolonging overall survival in the selected HCC patients. However, it was also observed that the patients submitted to prolonged L-ARG depletion from ADI-PEG20 showed a relatively superior overall survival. Based on these results, the authors are now focused on the development of new strategies to enhance prolonged L-ARG depletion and synergize the effect of ADI-PEG20 (Abou-Alfa et al., 2018).

The results from recent clinical trials concerning MM are not so encouraging as the ones from HCC. Feun and colleagues performed a clinical trial with ocular and cutaneous melanoma patients (NCT00450372) whose results demonstrated that the response to L-ARG-depleting therapy was indeed correlated with lack of ASS expression. However, contrary to previous results, they also observed that it was necessary to raise the dose of ADI-PEG20 from 160 to 320 IU/m^2 per week to observe any patient response (Feun et al., 2012). In another contemporary clinical trial with unresectable stage III or stage IV melanoma patients (NCT00520299), it was also observed that ADI-PEG20 has limited clinical activity as a single agent at doses of 160 IU/m^2 or less.

This study also revealed an inverse correlation between mean levels of ADI-PEG20 and ADI-PEG20 antibody levels which is suggestive of some immunogenicity associated with ADI-PEG20. Unexpectedly, their results suggested that uveal melanoma might be more sensitive to L-ARG deprivation than cutaneous melanoma (Ott et al., 2013). A recently completed Phase II clinical trial with malignant pleural mesothelioma patients (NCT01279967) showed that ADI-PEG20 significantly improves progression-free survival, and possibly the overall survival (Szlosarek et al., 2017).

There are also several on-going phase I and II clinical trials using ADI-PEG20 monotherapy in patients with different tumours, such as relapsed sensitive or refractory small cell lung cancer (SCLC) (NCT01266018), prostate cancer and non-small cell lung cancer (NCT01497925), acute myeloid leukaemia (NCT01910012), advanced non-Hodgkin's lymphoma (NCT01910025).

Other phase-I and -II clinical trials testing ADI-PEG20 in combination with existing chemotherapy are completed or ongoing for advanced gastrointestinal cancer (NCT02102022), HER2 negative metastatic breast cancer (NCT01948843), advanced pancreatic cancer (NCT02101580), malignant pleural mesothelioma (NCT02709512), soft tissue sarcoma (NCT03449901), HCC (NCT02101593), and MM (NCT01665183), among others.

8.2.2.2 Structure and reaction mechanism

As already discussed, most studies performed to date used ADI isolated from *Mycoplasma arginini*. Based on the Protein Data Bank existing crystallographic structures and on their crystallographic structures, Das et al. (2004) proposed a catalytic mechanism for ADI. Some of its the most important structural features (Das et al., 2004; Galkin et al., 2005; Lu et al., 2004; Lu et al., 2006; Smith et al., 1978) are represented in Fig. 8.2 (PDB code 1S9R) (Das et al., 2004).

The analysis of existing data suggests that L-ARG enters into the enzyme's active site and is oriented to its more favourable catalytic position by a network of polar interactions between its guanidino moiety and the side chain carboxylates of ASP161 and ASP271 residues (Fig. 8.2).

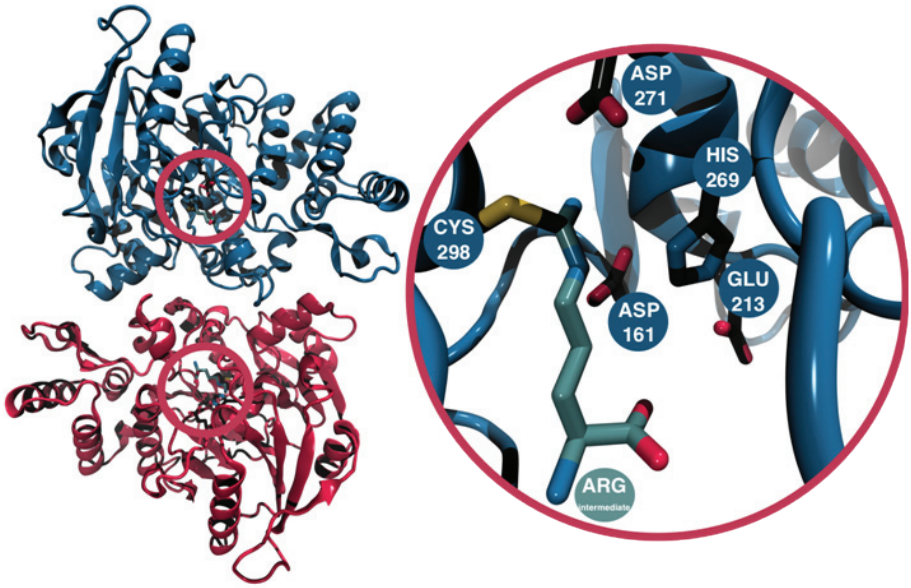


Figure 8.2 (Left) Quaternary structure of ADI homodimer (PDB code: 1S9R) coloured by subunit; (right) close-up view of the active site of ADI.

The proposed mechanism for the enzymes is represented in Fig. 8.3 and comprises six sequential steps.

The ADI mechanism starts with a nucleophilic attack from CYS398 to Cζ carbon of the substrate and a series of simultaneous proton transfers, from CYS398 to Nη2 nitrogen of the substrate, from Nη2 to HIS2696 and from HIS269 to GLU213. At the end of the first step, a tetrahedral intermediate is generated with the Cζ carbon of the substrate covalently bonded to CYS398 side chain (Step 1).

In the second step, HIS269 gives a proton to nitrogen Nη2, generating an ammonia molecule. Simultaneously, HIS269 receives a proton from GLU213. At the end of this step, the ammonia molecule is released from the active site, and the reaction intermediate remains covalently bonded to CYS398 (Step 2).

The next step of the catalytic process requires the intervention of a water molecule that is in close proximity to the ASP271 and HIS269 residues. This water molecule is involved in

a nucleophilic attack to carbon Cζ of the reaction intermediate and simultaneously donates a proton to HIS269 that, by its turn, donates a proton to GLU21. The structural data suggests that this reaction may be endorsed by the close proximity of ASP161 that establishes two hydrogen bonds with the –NH groups of the intermediate that turns Cζ carbon more electrophilic. At the end of this reaction, another tetrahedral intermediate is generated covalently bonded to CYS398. Contrary to the first step (Step 1), this one has a hydroxyl group attached to Cζ carbon instead of an amino group (Step 4).

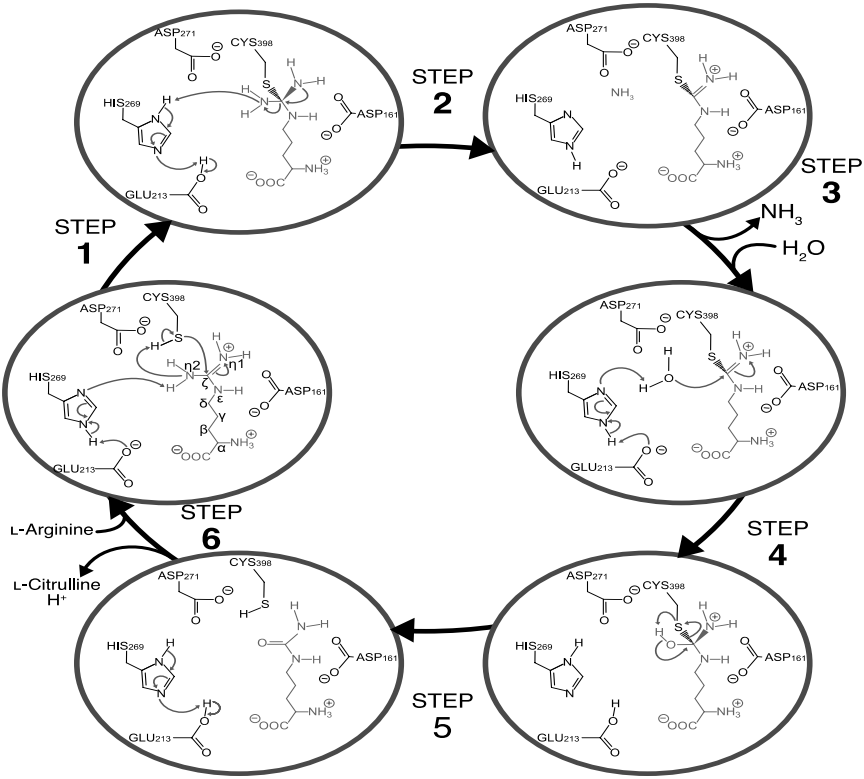


Figure 8.3 Schematic representation of the current proposal of the catalytic mechanism of ADI.

In the following step, a proton is transferred from the hydroxyl group of the reaction intermediate to CYS398, with the consequent cleavage of the covalent bond between CYS398 and

the substrate. Simultaneously, there is a proton transfer from GLU213 to HIS269. This reaction generates a molecule of L-citrulline that is ready to leave the active site (Step 5).

The last step is the enzymatic turnover in which HIS398 loses a proton to the solvent or an unknown neighbouring active site residue and the enzyme becomes ready for a new catalytic cycle (Step 6).

8.2.3 Arginase

Arginase (EC 3.5.3.1) was the first enzyme used to deprive cancer cells of L-ARG. It is found in organisms from the five kingdoms and is one of the very few enzymes that specifically requires a spin-coupled Mn^{2+}–Mn^{2+} cluster for catalytic activity in vitro and in vivo (Maharem et al., 2018; Reczkowski and Ash, 1992).

There are two genetically distinct isozymes named types I and II. They share approximately 60% amino acid sequence homology and differ in their expression, regulation, tissue distribution, subcellular locations in mammals, immunologic reactivity, physiologic function, and certain enzymatic properties (Dizikes et al., 1986; Herzfeld and Raper, 1976; Jenkinson et al., 1996; Spector et al., 1983; Vockley et al., 1996).

Arginase I (ArgI) is the only of the two isozyme types with demonstrated anti-cancer therapeutic potential. It is a cytosolic enzyme found predominantly in liver hepatocytes, where it catalyses the hydrolysis of L-ARG to L-ornithine and urea. This is the final cytosolic step of the urea cycle, which detoxifies ammonia in mammals (Krebs, 1973). L-ornithine can be further metabolized to polyamines via ornithine decarboxylase (ODC). Then, it can generate L-proline via ornithine aminotransferase (OAT), or it can be converted to citrulline by ornithine transcarbamylase (OTC) (Morris Jr., 2002).

Recent studies show that tumour inhibition mediated by ArgI is only effective in OTC deficient cells and is independent of ASS expression (Cheng et al., 2007). OTC is an enzyme highly expressed in liver and small intestine, but its expression in cancer and other normal tissues is mostly down-regulated (Delers et al., 1984).

A metabolic functional cell possessing OTC will avoid the intracellular depletion of arginine by converting the ArgI product

ornithine into citrulline, which can be converted back to L-ARG via ASS. An OCT-deficient cell when deprived of L-ARG is not capable to generate the intermediate metabolite citrulline, which is the substrate of ASS, culminating in the complete depletion of intracellular L-ARG. In the absence of its substrate, the ASS presence in the OTC-deficient cells is irrelevant. A recent study showed that this simple explanation might not be fully correct and applicable to all cancer cells. In particular, the results showed that some OTC and, partially ASS, deficient cancer cells exhibit high resistance to L-ARG deprivation in vitro (Bobak et al., 2010). This means that the inhibitory effects of L-ARG depleting agents on tumour growth require further investigation in order to be fully understood.

The individual evaluation of the expression of both OTC and ASS in each cancer seems to be a useful request before deciding which of the two enzymes should be used as in the l-ARG depriving therapy.

The administration of ArgI for the treatment of tumours has been reported since the 1950s (Simon-Reuss, 1953; Wiswell, 1951). Although the in vitro experiments were very promising at the time, the scale-up to in vivo tests was discouraging due to ArgI low affinity for L-ARG (K_M: 6 mmol/L at physiologic pH), its alkaline optimum pH (pH 9.6), and its very short circulatory half-life time (approximately 30 min) (Savoca et al., 1979).

The construction of a recombinant human arginase I (rhArgI), which was only possible after the advent of recombinant DNA technology, brought a renewed interest in the use of arginase as an anti-tumour agent.

In the last years a pegylated cobalt-modified enzyme was developed. The substitution of the Mn^{2+} metal centre in human ArgI with Co^{2+} (Co-rhArgI) results in an enzyme that displays 10-fold higher catalytic efficiency for l-ARG hydrolysis, 12–15-fold reduction in the IC_{50} towards a variety of malignant cell lines and a half-life time of 22 h in serum. The enhanced pharmacological properties of Co-hArgI translate into a 12–15-fold improvement in the killing of melanomas and hepatocellular carcinomas in vitro (Stone et al., 2010). Later, the same group reported that the modification of Co-hArgI with PEG-5K NHS esters increased the retention of the enzyme in circulation by about 2 orders of magnitude (Glazer et al., 2011).

This modified version of bioengineered Co-rhArgI–PEG was proven to be cytotoxic in HCC (Cheng et al., 2007), melanoma (Lam et al., 2011), pancreatic carcinoma tumour xenografts (Glazer et al., 2011), acute myeloid leukaemia (Tanios et al., 2013), and glioblastoma cells (Khoury et al., 2015).

Beyond the already discussed particularities between rhArgI and ADI sensitive cancers, there are other differences that must be mentioned. One is the fact that contrary to ADI, rhArgI is a human enzyme and consequently non-immunogenic. The other is the greater toxicity of the rhArgI relative to ADI, which according to some authors may be due the lack OTC in normal tissues and the consequent impossibility to rescue L-ARG from ornithine. The authors proposed that providing citrulline supplementation to patients undergoing rhArgI mediated L-ARG deprivation could lower the treatment toxicity and eventually improve the therapeutic index, particularly in patients with ASS-negative metastatic cancers (Mauldin et al., 2012).

8.2.3.1 Therapeutic uses

By 2018, the number of clinical trials using rhArgI is quite lower than the number of clinical trials with ADI. Also, to our knowledge, to date there is no phase III clinical trial completed. There are at least four clinical trials completed, involving phase I and II. One of them aimed to determine whether rhArgI-PEG was safe and effective in the treatment of advanced HCC (NCT01092091). It was concluded that rhArgI-PEG has a manageable safety profile and is potentially a superior L-ARG depleting agent than ADI in the treatment of HCC due to its low toxicity profile, apparently inexistence of immunogenicity and sustainable l-ARG depletion. The optimal therapeutic dose of rhArgI-PEG was defined as 1600 IU/kg in advanced HCC patients. In contrast to ADI-PEG20, rhArgI-PEG was administrated intravenously (Yau et al., 2013). This clinical study followed a first one performed by the same research group, where pegylated recombinant human ArgI was administrated in 8 escalation doses alone and in combination with standard doses of doxorubicin (NCT00988195).

Two other completed studies evaluated the safety and efficacy of rhArgI-PEG in combination therapy in patients with advanced HCC (NCT02089633) and the safety of increasing dose

levels of Co-rhArgI-PEG in patients with relapsed or refractory acute myeloid leukaemia or myelodysplastic syndrome (NCT02732184).

There are at least two clinical trials recruiting patients: One of them intends to evaluate the safety of rhArgI-PEG to treat patients with L-ARG auxotrophic advanced solid tumours (melanoma, renal cell carcinoma, prostate cancer, and hepatocellular carcinoma), and that have progressed after receiving approved or established therapies (NTC02285101). The other one aims to study the safety of increasing dose levels of Co-rhArgI-PEG in patients with advancer cancers (NCT02561234). One phase II clinical trial is active but not recruiting. Its objective is to evaluate the efficacy of rhArgI-PEG in patients with relapsed or refractory acute myeloid leukaemia in terms of remission rate (NCT02899286). Finally, there is an international phase I/II trial that is not yet recruiting and whose objective is to evaluate the safety and activity of rhArgI-PEG in children and young people with relapsed/refractory leukaemia, neuroblastoma, sarcoma, and high-grade gliomas (NCT03455140).

8.2.3.2 Structure and reaction mechanism

hArgI is an oligomer with 107 kDa with optimum pH at 9.3 and a K_M for l-ARG of about 10.5 mM (Beruter et al., 1978). The enzyme has some particular structural features that includes spin-coupled Mn^{2+}–Mn^{2+} cluster and an S-shaped tail, 19 amino acids long, located at the subunit-subunit interface. This S-shaped tail-mediated 54% of all the intersubunit interactions (Lavulo et al., 2001).

There are several proposals for the catalytic mechanism of hArgI, (Leopoldini et al., 2009; Christianson and Cox, 1999; Cox et al., 2001; Khangulov et al., 1998), although many aspects remain elusive. The analysis of the crystallographic structure of the human ARGase in complex with L-ornithine (PDB code: 3GMZ) (Fig. 8.4) (Ilies et al., 2011) reveals some clues about the important catalytic residues, and their possible role in the catalytic mechanism. GLU277 is one of the residues that stands out. It is an active site residue that is apparently implicated in the substrate binding through the formation of a strong electrostatic interaction with the guanidium group. This interaction places

the substrate in a favourable position with respect to the metal-bound hydroxide anion with whom it must react (Cavalli et al., 1994; Kanyo et al., 1996; Scolnick et al., 1997). Additionally, GLU277, along with a hydroxide anion and ASP232 and ASP128 are coordinated with the Mn^{2+} ions.

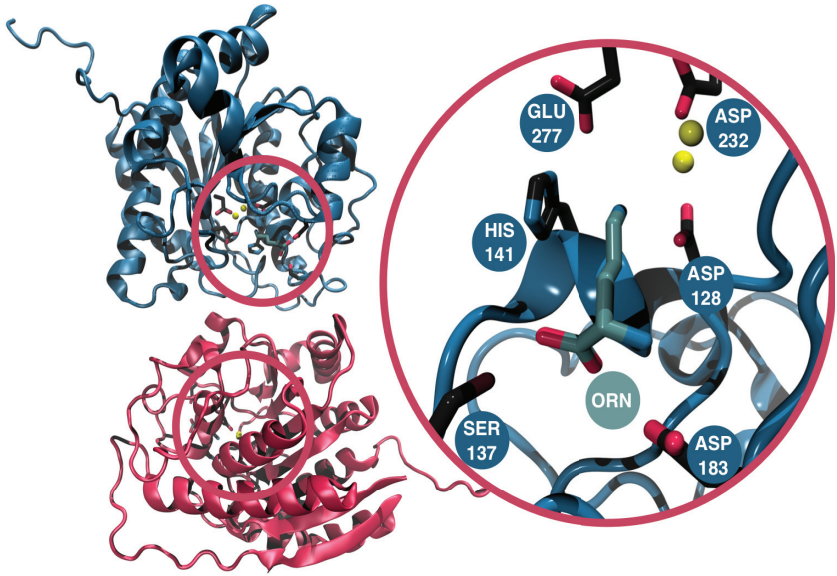


Figure 8.4 (Left) Quaternary structure of hArgl homodimer (PDB code: 3GMZ) coloured by subunit; (right) close-up view of the active site of the enzyme.

The proposed mechanism for this reaction is represented in Fig. 8.5 and comprises six sequential steps.

The reaction starts with the binding of the L-ARG's guanidine group to Mn^{2+} ions. After its correct positioning, L-ARG undergoes a nucleophilic attack from the bridging hydroxy anion to its carbon Cζ (Step 1) originating a tetrahedral intermediate. In the second step, ASP128, which is thought to be deprotonated (Leopoldini et al., 2009), acts as an acid/base catalyst, abstracting a proton from the hydroxyl group and shuttling it to nitrogen Nε of the substrate (Step 2). In the third step, the bond between nitrogen Nε and Carbon Cζ from the substrate breaks, generating the reaction products: L-ornithine and urea (Step 3). The next

steps will promote product dissociation from the enzyme. It requires the entry of a water molecule that will bridge the binuclear Mn^{2+} cluster, causing the urea product to move to a terminal coordination site on Mn^{2+} (Step 4). Then, HIS141 residue acts as a proton shuttle from the bulk solvent to nitrogen Nε of ornithine. This proton transfer facilitates product release and regenerates the nucleophilic metal-bridging hydroxide anion (Step 5). Finally, ornithine dissociates and the enzymatic turnover takes place preparing the enzyme for a new catalytic cycle (Step 6).

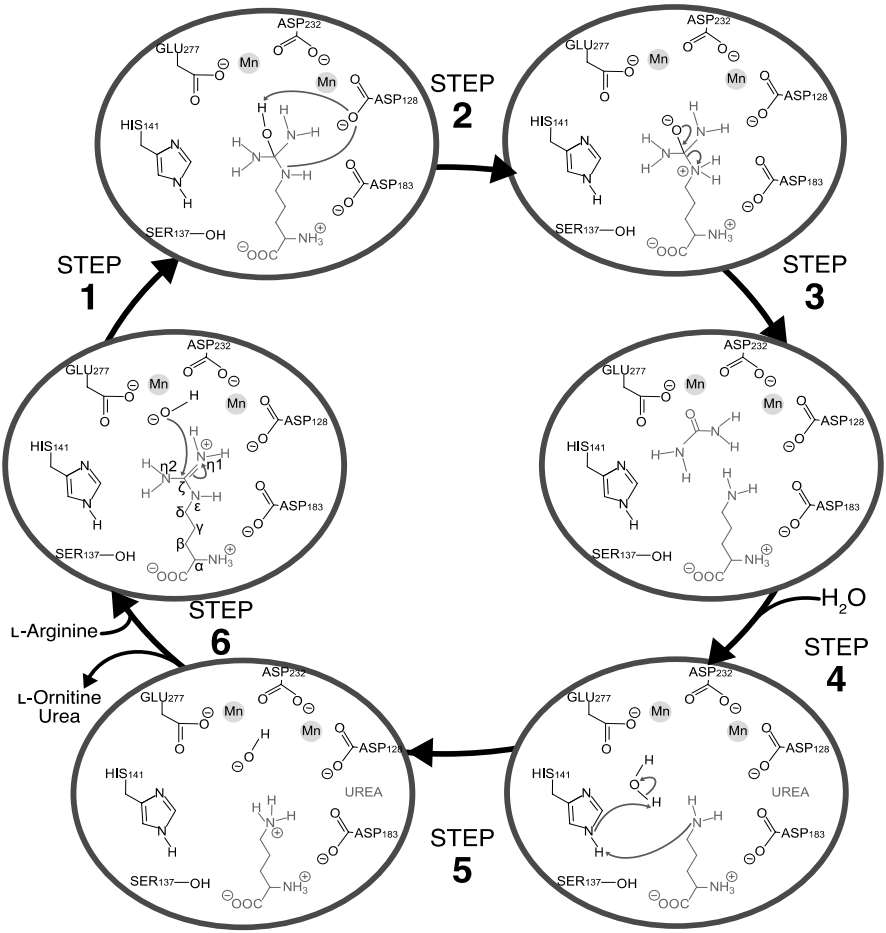


Figure 8.5 Schematic representation of the Current proposal of the catalytic mechanism of hArgI.

8.2.4 Asparagine

L-Asparagine (L-ASN) is a non-essential amino acid that is synthesized in humans by the asparagine synthetase enzyme (Lomelino et al., 2017). However, L-ASN is indispensable for the synthesis of several proteins, in particular, the ones that are glycosylated through N-glycosylation (Cherepanova and Gilmore, 2016) Moreover, this amino acid is involved in the control of some cell functions in the nervous system, and also in the metabolism of ammonia to enable the elimination of this toxic metabolic by-product (Erecinska et al., 1991; Chen and Chen, 1992).

In humans, the asparagine synthetase catalyses the production of L-ASN through an ATP-dependent reaction, where L-aspartate and L-glutamine are converted to L-ASN and L-glutamate (Lomelino et al., 2017; Richards and Kilberg, 2006).

Although cells can produce their own supply of L-ASN, some types of cancer, such as leukaemia (Shahriari et al., 2017; Wetzler et al., 2007), breast (Knott et al., 2018; Yang et al., 2014), gastric (Yu et al., 2016), and lung (Xu et al., 2016) cancers, are impaired in the pathways related to the synthesis of L-ASN. Consequently, these tumours become dependent on an external supply of L-ASN, which offers a specific opportunity window to target cancer using L-ASN deprivation therapies.

L-Asparaginase (ASNase) is an enzyme available in bacteria, plants, and fungi that catalyses the hydrolysis of L-ASN to L-aspartate and ammonia (Gesto et al., 2013). The adequate administration of ASNase in the bloodstream can reduce the plasma levels of L-ASN and promote tumour growth slowdown by starving the cancer cells.

Kidd observed, for the first time, the therapeutic potential of ASNase as an anti-tumour approach in two different types of lymphoma. (Kidd, 1953, 1953a) However, at the time, the role of ASNase was unknown because Kidd only reported that the gynae pig serum was able to regress the lymphomas in murine. Later on, Broome and Dolowy were able to identify and isolate ASNase from the gynae pig serum, and then, they understood that the anti-tumour effect was due to the ASNase (Broome, 1961, 1963; Dolowy, 1966). These studies were pioneer to point out ASNase as an enzyme with great therapeutic potential against cancer.

ASNase (EC 3.5.1.1) is an enzyme available in several bacteria genus, namely *Bacillus*, *Pseudomonas* (Saeed et al., 2018), *Serratia*, *Aerobacter*, *Photobacterium, Xanthomonas*, (Peterson and Ciegler, 1969) *Proteus* (Tosa et al., 1971), *Streptomyces* (DeJong, 1972), *Vibrio* (Kafkewitz and Goodman, 1974), and *Aspergillus* (Sarquis et al., 2004). Besides the broad spectrum of microorganism that has ASNase in their proteome, ASNases derived from *Escherichia coli*, and *Erwinia chrysanthemi* showed the most promising results in anti-cancer therapy. The *E. coli* form is the main source of ASNase because of its capability to produce the enzyme in great quantities. The *E. chrysanthemi* form is an alternative when hypersensitivity reactions occur in patients treated with the *E. coli* ASNase. In fact, the immunogenicity is the biggest drawback of the use of these non-human enzymes in anti-cancer therapy.

ASNase from *E. coli* presents two different forms, ASNase-I and ASNase-II, which are different regarding its cellular location and affinity for the substrate. ASNase-I is present in the cytosol, and it has a lower affinity for the substrate (L-ASN) when compared with ASNase-II that is located in the periplasmic space (region between the inner cytoplasmic membrane and the bacterial outer membrane) (Yao et al., 2005). Due to the higher affinity of ASNase-II for the substrate, this enzyme has been extensively studied for application in anti-cancer therapy.

8.2.4.1 Therapeutic uses

ASNase-II has been used as a general therapy agent for the treatment of acute lymphoblastic leukaemia, and other related blood cancers (Shrivastava et al., 2016). The administration of ASNase causes a decrease in the circulating pool of L-ASN, starving the susceptible cancer cells whose cell cycle becomes arrested in the G1-phase, which causes apoptosis (Lee et al., 1989; Keating et al., 1993).

Besides the excellent therapeutic efficiency, ASNase has a short half-life time in the bloodstream, causing a higher demand for injections to keep the therapeutic level. Consequently, this procedure leads to some side effects, namely allergies, exacerbated immune responses, and even anaphylactic shock (Mashburn and Wriston, 1964; Moola et al., 1964; Soares et al., 2002; Henriksen et al., 2015).

Similarly to what happens with other enzymes used in therapy, pegylation is one of the most appropriate manners to improve de plasma half-life and reduce the immunogenicity of the ASNase. Pegylation is a process in which the enzyme is modified with 5 kDa units of monomethoxypolyethylene glycol (PEG) to reduce de exposition of immunogenic patterns in the enzyme's surface (Rytting, 2010). The pegylated ASNase keeps its enzymatic activity, but the half-life time in the bloodstream increases about five times, and the immunogenicity is also lower (Zeidan et al., 2008).

Currently, pegylated ASNase is being used as first-line therapy in the USA, UK, Australia, and New Zealand, and as second-line therapy in Europe, where it is used when the patients are known to have an allergy to the free ASNase (Fu and Sakamoto, 2007). The application of ASNase was approved by the FDA in 2006 and it is commercialized under the name Oncaspar®. Ten years later, in 2016, EMA approved Oncaspar® to be used in European Union.

Although all these formulations make possible the use of ASNase as an anticancer therapeutic target, the formulations currently available are far from perfect. The major side effect continues to be the immunogenicity derived from the bacterial source of the ASNase. There are well-known side effects caused by the administration of ASNase in patients with cancer, namely hypersensitivity reactions (Avramis and Tiwari, 2006), coagulation disorders (Hernandez-Espinosa et al., 2006), pancreatitis, hyperglycaemia (Raja et al., 2012), and hepatotoxicity (Bodmer et al., 2006). These side effects give rise to an intense search for alternatives in order to identify other isoforms, and enzymes from other organisms, that show, at least, the same efficacy, and attenuated side effects (Thomas et al., 2010; Shrivastava et al., 2012; Mendes et al., 2012).

8.2.4.2 Structure and reaction mechanism

ASNase-II from *E. coli* is a homotetramer whose subunits have 326 amino acids each (Fig. 8.6). The pocket where the enzyme binds the substrate and actively catalyses the reaction is composed of amino acid residues from two adjacent subunits: THR12, TYR25, SER58, GLN59, THR89, ASP90, and LYS162 from one subunit, and ASN248 and GLU283 from the adjacent subunit. These residues

establish a strong set of hydrogen bond interactions that includes a conserved water molecule (Fig. 8.6). There were several plausible proposals for the catalytic mechanism of ASNase-II, but the symmetry of the location of THR12 and THR89 residues, which enclose the C2 carbon of the substrate, suggests a direct involvement in the catalytic mechanism (Swain et al., 1993).

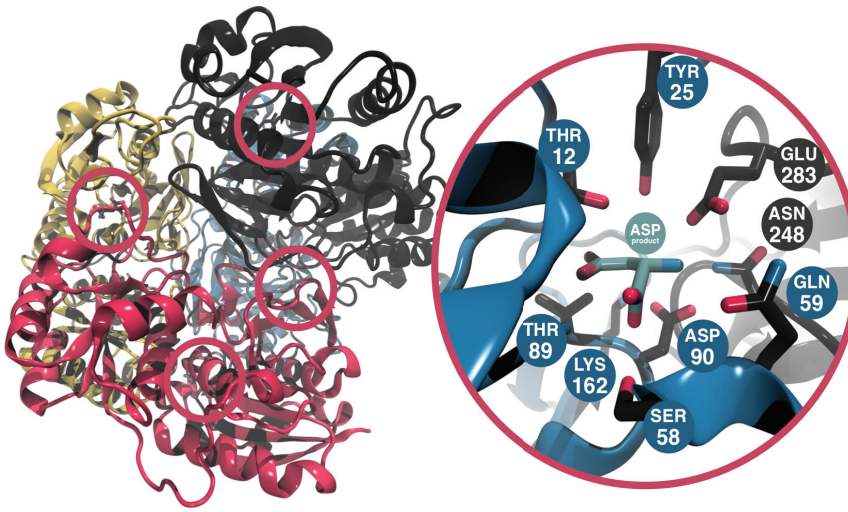


Figure 8.6 (Left) Quaternary structure of the ASNase (PDB code: 3ECA) coloured by subunit. (Right) Representation of the important amino acid residue, and the product in the active site of the ASNase. The molecules are tagged by residue name and residue index.

Currently, the catalytic mechanism of ASNase-II is known and supported by several computational and experimental studies (Gesto et al., 2013; Borek et al., 2014; Schalk et al., 2016).

The catalytic process starts with the formation of a tetrahedral intermediate after the nucleophilic attack of the water molecule to carbon Cα of the L-ASN (substrate) (Fig. 8.7, Step 1). Since the water molecule establishes a new covalent bond with the substrate, one of its protons is passed to the amino group of LYS162 that becomes positively charged.

In the second step, LYS162 residue protonates the amino group of the substrate in a proton transfer mediated by the THR89 residue (Fig. 8.7, Step 2). Afterwards, the amino group is eliminated as ammonia, and the aspartic acid is formed (Fig. 8.7,

Step 3). Finally, both products are released, and the enzyme becomes available to bind a new substrate molecule to initiate another catalytic cycle (Fig. 8.7, Step 4) (Gesto et al., 2013).

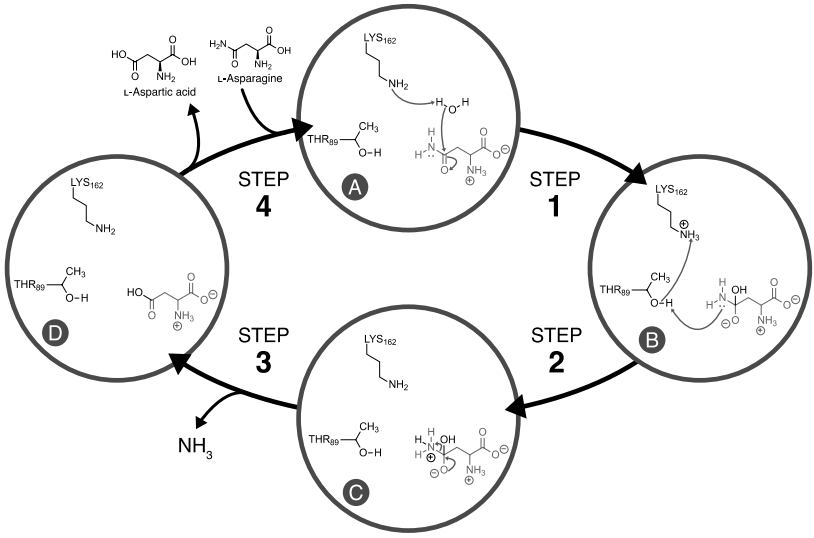


Figure 8.7 A schematic representation showing a proposal for the catalytic mechanism of the ASNase. A–D are reaction intermediates.

8.2.5 Methionine

L-Methionine (L-MET) is one of two sulphur-containing proteinogenic amino acids. This essential amino acid for humans (Young, 1994) acts as an essential precursor for the synthesis of cysteine (through the production of S-adenosylmethionine), taurine, glutathione (a vital tripeptide that reduces reactive oxygen species (ROS), protecting cells from oxidative stress), and polyamines spermidine and spermine (crucial for nuclear and cell division (Cellarier et al., 2003; Davis and Uthus, 2004; Ho et al., 2011; Thivat et al., 2007). Although all these roles are essential for the cell wellbeing, the most important role of L-MET is as a source of methyl units for the DNA methylation through its intermediate S-adenosylmethionine.

It is well known that the DNA methylation is one of the most significant epigenetic signalling markers, controlling the gene expression in several processes such as X-chromosome

inactivation, genomic imprinting, repression of repetitive elements, and carcinogenesis (Geiman and Muegge, 2010; Ramani et al., 2010). Regarding the carcinogenesis, in some types of cancer, e.g. endometrial (Fiolka et al., 2013), breast (Chimonidou et al., 2011), lung (Powrózek and Małecka-Massalska, 2016), and myeloma (Wong and Chim, 2015) cancers, the tumourigenic process has been associated with the hypermethylation of the promoter regions of tumour-suppressor genes (Kulis and Esteller, 2010). Considering the dependence of cells of L-MET for the DNA methylation event, it was suggested that strategies leading to L-MET deprivation could be used as therapy against cancer. Since, L-MET is a crucial amino acid in the synthesis of almost all proteins of the cell, this means that the enzyme cannot be applied in long treatments and must be inactivated.

Since L-MET is an essential amino acid, the L-MET deprivation could be harmful for the human body, but it was proved that long-term nutritional deprivation of L-MET does not compromise the human life because cells have several pathways to ensure the L-MET recycle. L-MET can be recovered through the re-methylation of homocysteine by the L-methionine synthase (MS) or, in the liver, by the betaine-homocysteine methyltransferase (BHMT) (Cellarier et al. 2003; Hoffman, 1984; Guo et al., 1993). Additionally, methylthioadenosine phosphorylase (MTAP) is also able to catalyse the production of L-MET from 5'-methylthioadenosine.

Although all of these salvage pathways can restore the required levels of L-MET, several malignant cell lines, including breast, lung, colon, kidney and bladder cancers, glioblastoma and melanoma, have suppressed these salvage pathways (Cavuoto and Fenech, 2012). Consequently, the cancer cells are dependent on an exogenous supply of L-MET, manifesting a growth inhibition when this supply is suppressed (Jeon et al., 2016; Chaturvedi et al., 2018).

The high demand for L-MET in the tumour cells as a result of their great protein synthesis and the lack of L-MET salvage pathways make L-MET essential for the tumour growth and proliferation. Therefore, the L-MET restriction can be exploited as a cancer therapy, in particular for these L-MET-dependent tumours (Chaturvedi et al., 2018).

In the literature, two different approaches have been documented with therapeutic promising results using the dietary

deprivation of L-MET and the use of L-MET depletion enzymes. The nutritional restriction of L-MET showed promising results in animal studies, but this diet showed a limited efficiency in clinical trials (Chaturvedi et al., 2018; Thivat et al., 2009). For example, a phase I clinical trial reported a poor reduction of L-MET in plasma (58%) after 17.3 weeks of L-MET dietary restriction (Epner et al., 2002).

Due to the lack of effectiveness in reducing the serum levels of L-MET through nutritional restriction, methioninase or methionine γ-lyase (MGL) has been studied to achieve the desired level of reduction. Several studies reported a fast and effective depletion of L-MET from serum, showing a vast efficiency against human tumours graphs implanted in mice. A pilot study was conducted in three patients showing a drastic reduction in plasma L-MET from 23.1 μM to 0.1 μM (reduction of approximately 99.6%) when the enzyme was administrated by a 10 h intravenous infusion (Tan et al., 1996a). Based on this report, MGL corroborated its efficiency in reducing serum L-MET levels. Moreover, a slowdown in the tumour growth and spread necrosis in tumour regions were observed in either the MHL administered alone or in combination with other drugs (Yoshioka et al., 1998; Tan et al., 1999; Murakami et al., 2017, Kawaguchi et al., 2017, Kokkinakis et al., 2001).

Beyond the depletion effect of MGL, this enzyme can also be administered in combination with the prodrug selenomethionine. The MGL can produce methyl-selenol from the selenomethionine, which causes oxidation of thiol groups generating toxic superoxide molecules that cause cell death. By this way, MGL can be fused with a specific antigen to target the tumour cell and, then, the selenomethionine is administrated, being active only in the area where the MGL is located (Zhao et al., 2006; Guillen et al., 2014).

8.2.5.1 Therapeutic uses

MGL has a broad range of application in cancer therapies. Since this enzyme is absent in mammals, MGL was initially isolated from *Clostridium sporogene* (Sharma et al., 2014). Later on, the enzyme from *Pseudomonas putida* was chosen for clinical purposes, due to its greater stability and affinity for the substrate

(low K_M) (Esaki and, Soda, 1987). Therefore, the MGL gene from *P. putida* was isolated and expressed in *Escherichia coli* to produce the recombinant enzyme in great quantities for administration in humans (Hori et al., 1996).

The use of an enzyme from another organism raises complications related to the immunogenicity of the recombinant enzyme and its low stability in the bloodstream. In order to overcome this problem, the recombinant MGL was modified with polyethylene glycol (PEG). The pegylated enzyme presented low immunogenicity, a longer half-life time, and greater resistance against degradation by other enzymes in the bloodstream (Tan et al., 1999; Kudou et al., 2007, Takakura et al., 2006). Pegylation can overcome these issues but causes a decrease in the affinity for the PLP cofactor which decreases the efficiency and specificity of the MGL. Fortunately, this issue was easily solved through the supplementation of the patients with PLP and oleic acid (Takakura et al., 2006).

Although all these promising results, there are only one phase I clinical trials addressing the viability to administered intravenously MGL to deplete plasma L-MET. The pegylated MGL was tested in three healthy people reporting no significant toxicity and a considerable capacity to decrease serum L-MET level. In sum, this pilot phase I clinical trial reported that the administration of MGL is safe and able to effectively reduce L-MET level in the bloodstream (Tan et al., 1996; Tan et al., 1997). However, the clinical trials must continue to corroborate the effectiveness of MGL to treat cancer.

Since the recombinant MGL presents strong applicability for the cancer treatment, the molecular details, in particular, the catalytic mechanism can provide valuable insights to enhance its efficiency.

8.2.5.2 Structure and reaction mechanism

MGL (EC 4.4.1.11) is a pyridoxal-5′-phosphate (PLP)-dependent enzyme that catalyses the reversible conversion of L-MET into methanethiol, ammonia, and 2-oxobutanoate (Kreis and Hession, 1973). The enzyme is a homotetramer with 398 amino acid residues in each subunit, and it has four active sites placed in each interface of two adjacent subunits (Fukumoto et al., 2012). Each

one of the four active sites binds a PLP cofactor molecule, which is required for binding the substrate (L-MET) and triggering the reaction (Kudou et al., 2007, Motoshima et al. 2000; Tone, 2005).

Although MGL belongs to the γ-family of PLP-dependent enzymes, there is an important cysteine residue (CYS116) in the active site that is absent in the other enzymes of the same family. Mutagenic studies reported that the CYS116 residue is crucial for the enzymatic activity (Kudou et al., 2008) and substrate specificity (Kudou et al. 2007, Inoue et al., 2000). Beyond the CYS116 residue, there are five more residues with a key role in the mechanism based on structural, enzymatic, and mutagenesis studies: TYR59, ARG61, TYR114, ASP186, LYS211, LYS240, and ASP241 (Fig. 8.8).

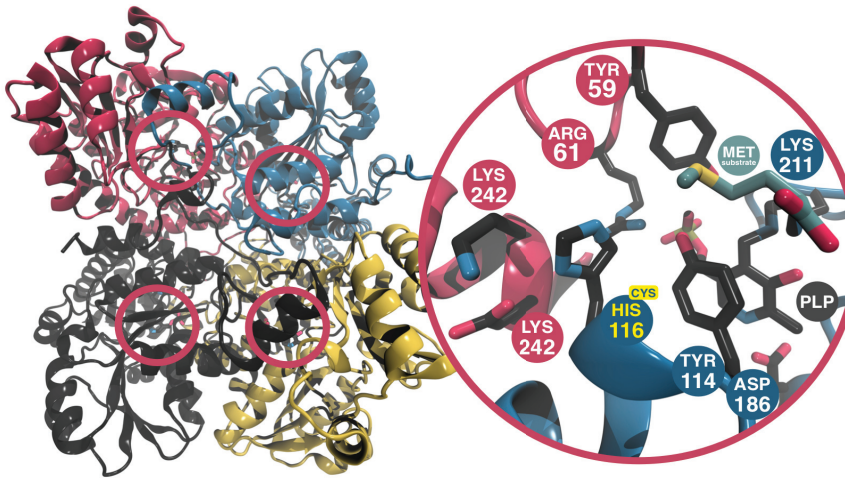


Figure 8.8 (Left) Quaternary structure of the MGL (PDB code: 3VK3) coloured by subunit. (Right) Representation of the important amino acid residues, the PLP cofactor, and the substrate in the active site of the MGL. The molecules are tagged by residue name and residue index. The CYS116 was mutated by a HIS residue to allow the co-crystallization of the MGL with the natural substrate (L-MET).

The proposed catalytic mechanism for MGL is based on these studies and the available knowledge for other PLP-dependent enzymes (Kudou et al. 2007, Cerqueira et al., 2011; Oliveira et al., 2011; Fernandes et al., 2017b, 2018).

All PLP-dependent enzymes wait for their substrate in a steady state where the PLP is covalently bonded to a lysine residue (LYS211) forming an internal aldimine (Fig. 8.9A). When a molecule of substrate enters the active site, its amino group makes a nucleophilic attack (Fig. 8.9, Step 1) to the C8 of the PLP cofactor generating a gem-diamine intermediate (Fig. 8.9B). Immediately after, the bond between the LYS211 residue and the PLP cofactor is broken, and the substrate (L-MET) becomes covalently bonded to the PLP cofactor (Fig. 8.9, Step 2), originating an intermediate commonly called external aldimine (Fig. 8.9C).

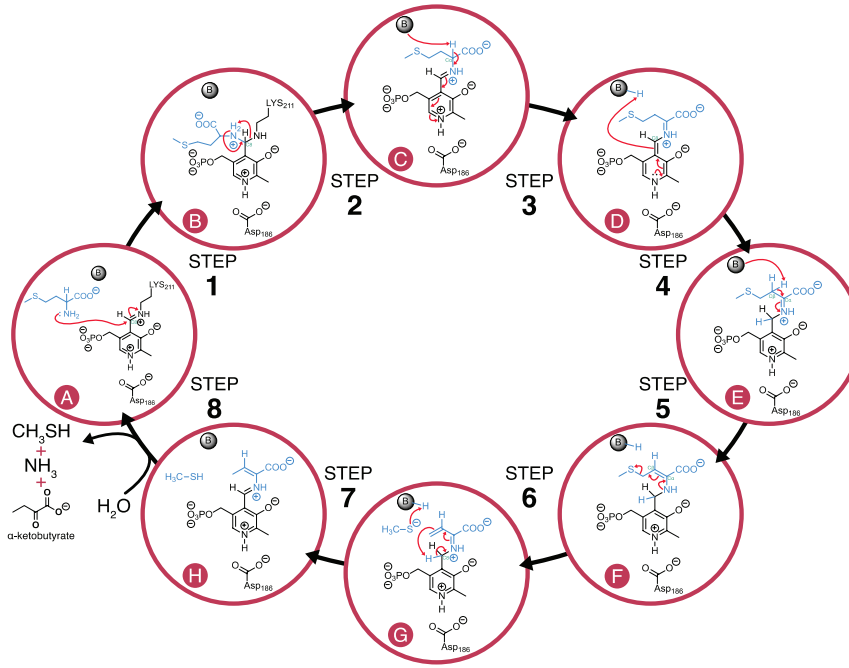


Figure 8.9 A schematic representation showing a proposal for the catalytic mechanism of the MGL. A–H are reaction intermediates, and the "B" sphere represents a generic molecule in the active that can behave as a base.

In the next step, the proton bonded to the Cα of the L-MET is abstracted from an unknown base from the active site (Fig. 8.9, Step 3). Based on the crystallographic structure and the catalytic mechanism of other PLP-dependent enzymes, the LYS211 residue

could be the base involved in this reaction. The resulting electron pair is stabilized across the PLP conjugated system forming a quinonoid intermediate (Fig. 8.9D).

In the fourth step, the proton abstracted in the previous step is transferred to carbon C8 of the PLP cofactor, which increases the acidity of the protons bonded to the Cβ of the L-MET. Consequently, in the fifth step, the base can get one of these protons.

The abstraction of the proton bonded to carbon Cβ atom triggers the elimination of the methanethiol group (γ-elimination) (Fig. 8.9, Step 6), which is rapidly protonated by the base in the active site. In the meantime, the obtained amino-crotonate intermediate (Fig. 8.9H) undergoes hydrolysis of its amino group. In this eighth step, the amino group is eliminated as an ammonia molecule, and a water molecule replaces that position as a carbonyl group.

In the end, the LYS211 becomes covalently bonded to the PLP cofactor, and the enzyme restores its initial configuration where the internal aldimine (Fig. 8.9A) is ready to catalyse a new reaction as soon as the substrate becomes available.

8.3 Conclusions

Amino acids are no longer regarded solely as building materials but also as sensing markers that regulate important signalling pathways that can be used for therapeutic purposes.

In the last decades, research on amino acid depletion in cancer cells using enzymes has provided valuable insights on the potential impact of metabolic control and regulation in the tumour microenvironment. These studies have shown that decreasing the concentration of specific amino acids in the bloodstream, which are required in high demands by tumour, can be an effectiveness way to treat the tumour.

The arginine dependence of some tumour cells, although known for more than 60 years, has failed to reunite by 2018 all the safety and efficacy conditions to become approved for use in therapy.

Although arginase was initially pointed as a potential new medicine against cancer cells, its inherent limitations, insurmountable at the time, turned the attentions to ADI. In 2018, there are

more than 20 clinical trials ongoing or completed whose aim is to evaluate the potential of ADI as a safe and effective anti-cancer treatment target. The results from a recently finished phase III clinical trial (NCT01287585) were not as encouraging as expected but did not withdraw the researcher's attention from ADI.

In the meanwhile, new strategies to improve arginase potential have arouse and at the moment there is a completed clinical trial (NCT01092091) were it was shown that rhArgI-PEG has a manageable safety profile and is potentially a superior arginine depleting agent than ADI in the treatment of HCC. At the moment there are about 10 clinical trials in different stages evaluating the effectiveness and safety of arginase as a potential anti-cancer therapeutic targets. The combination of the different arginine depleting enzymes with existing therapy is also being tested. The observation that patients submitted to prolonged arginine depletion showed a relative superior overall survival is a valid reason to keep the exploitation of arginine deprivation as a potential therapy against some cancers.

The L-ASN depletion has already been used in therapy after some modifications to overcome some of the most common side effects as immunogenicity. Therefore, the pegylated ASNase is used as first-line therapy in the USA, UK, Australia, and New Zealand, and as second-line therapy in Europe against acute lymphoblastic leukaemia and other related blood cancers. Although the evolution of these formulations, the research on improvements that can increase the efficacy and suppress the side effects goes on, in particular, to overcome some cases of hypersensitivity reactions, pancreatitis, hyperglycaemia, and hepatotoxicity. Contrary to he L-ASN, the L-MET depletion therapies against cancer are taking the first steps. The bacterial origin of the MGL enzyme is the source of major complications in the use of this enzyme in therapy. Several cases of immunogenicity reactions have been reported, and although the pegylation strategy suppressed the immunogenicity and low half-life time of the enzyme in the bloodstream, this modification causes a decrease in the affinity of the enzyme for the substrate. Thus, the amino acid deprivation strategy to treat cancer targeting the serum L-MET supply is currently in an early stage of development comparatively to L-ASN, for example.

As any new therapy targeting cancer, amino acid deprivation therapies must be explored with caution in order to eliminate potential negative impacts that can be caused by the side effects observed when the metabolism of normal cells is also affected. It is thus important to understand at the atomic-level the underlying mechanisms of metabolic function of the enzymes that are used in these therapies. At the same time, this knowledge can be used to guide the development of new and more specific therapies alone or in combination with other drugs that can decrease the number of side effects, and turn these treatments more effective.

References

Abou-Alfa, G. K., et al., *Ann. Oncol*., **29** (2018), 1402–1408.

Ascierto P. A., et al., *J. Clin. Oncol*., **23** (2005), 7660–7668.

Ashikaga T., et al., *Biosci. Biotechnol. Biochem*., **58** (1994), 839–842.

Avramis V. I., Tiwari P. N., *Int. J. Nanomedicine*, **1** (2006), 241–254.

Bach S. J., Swaine D., *Br. J. Cancer*, **19** (1965), 379–386.

Barbul A., et al., *Surgical Forum*, 28 (1977), 101–103.

Barbul, A., et al., *Am. J. Clin. Nutr*., 37 (1983), 786–794.

Beloussow K., et al., *Cancer Lett*., 183 (2002), 155–162.

Beruter J., Colombo J. P., Bachmann C., *Biochem. J*., 175 (1978), 449–454.

Billiar T. R., *Ann. Surgery*, **221** (1995), 339–349.

Bobak Y. P., et al., *Cell. Biol. Int*., **34** (2010), 1085–1089.

Bodmer M., et al., *Digestion*, **74** (2006), 28–32.

Borek D., et al., *FEBS J*., **281** (2014), 4097–4111.

Bowles T. L., et al., *Int. J. Cancer*, **123** (2008), 1950–1955.

Broome J. D., *Nature*, **191** (1961), 1114–1115.

Broome J. D., *J. Experim. Med*., **118** (1963), 99–120.

Cavalli R. C., et al., *Biochemistry*, **33** (1994), 10652–10657.

Cavuoto P., Fenech M. F., *Cancer Treat. Rev*., **38** (2012), 726–736.

Cellarier E., et al., *Cancer Treat. Rev*., **29** (2003), 489–499.

Cerqueira N. M., Fernandes P. A., Ramos M. J., *J. Chem. Theory Comput*., **7** (2011), 1356–1368.

Chaturvedi S., Hoffman R. M., Bertino J. R., *Biochem. Pharmacol*., **154** (2018), 170–173.

Chen Z. P., Chen K. Y., *J. Biol. Chem*., **267** (1992), 6946–6951.

Cheng P. N., et al., *Cancer Res*., **67** (2007), 309–317.

Cherepanova, N. A., Gilmore R., *Sci. Rep*., **6** (2016), 20946.

Chimonidou M., et al., *Clin. Chem*., **57** (2011), 1169–1177.

Christianson D. W., Cox J. D., *Annu. Rev. Biochem*., 68 (1999), 33–57.

Cox J. D., et al., *Biochemistry*, **40** (2001), 2689–2701.

Daly J. M., et al., *Ann. Surgery*, **208** (1988), 512–523.

Das K., et al., *Structure*, 12 (2004), 657–667.

Davis C. D., Uthus E. O., *Exp. Biol. Med*. (Maywood), **229** (2004), 988–995.

DeBerardinis R. J., Chandel N. S., *Sci. Adv*., **2** (2016), e1600200.

DeJong P. J., *Appl. Microbiol*., **23** (1972), 1163–1164.

Delage B., et al., *Cell Death Dis*., **3** (2012), e342.

Delers A., et al., *Mol. Cell. Biol*., **4** (1984), 809–812.

Dhankhar R., et al., *Expert Opin. Ther. Pat*., **28** (2018), 603–614.

Dillon B. J., et al., *Cancer*, **100** (2004), 826–833.

Dizikes G. J., et al., *Biochem. Biophys, Res. Commun*., **141** (1986), 53–59.

Dolowy W. C., et al., *Cancer*, **19** (1966), 1813–1819.

Durden D. L., Distasio J. A., *Cancer Res.,* **40** (1980), 1125–1129.

Ensor C. M., et al., *Cancer Res*., **62** (2002), 5443–5450.

Epner D. E., et al., *Nutr. Cancer*, **42** (2002), 158–166.

Erecinska M., et al., *J. Neurochem*., **57** (1991), 491–498.

Esaki N., Soda K., *Methods Enzymol*., **143** (1987), 459–465.

Fernandes H. S., et al., *Expert. Opin. Ther. Pat*., **27** (2017a), 283–297.

Fernandes H. S., Ramos M. J., Cerqueira N., *Chemistry*, **23** (2017b), 9162–9173.

Fernandes H. S., Ramos M. J., Cerqueira N. M. F. S. A., *ACS Catalysis*, **8** (2018), 10096–10110.

Feun L., et al., *Curr. Pharm. Des*., **14** (2008), 1049–1057.

Feun L. G., et al., *Br. J. Cancer*, **106** (2012), 1481–1485.

Feun L. G., Kuo M. T., Savaraj N., *Curr. Opin. Clin. Nutr. Metabol. Care*, 18 (2015), 78–82.

Fiolka R., et al., *Oncol. Rep*., **30** (2013), 2878–2886.

Fu C. H., Sakamoto K. M., *Expert. Opin. Pharmacother*., **8** (2007), 1977–1984.

Fukumoto M., et al., *Biosci. Biotechnol. Biochem*., **76** (2012), 1275–1284.

Fung M. K. L., Chan G. C., *J. Hematol. Oncol*, **10** (2017), 144.

Galkin A., et al., *J. Biol. Chem*., **280** (2005), 34080–34087.

Geiman T. M., Muegge K., *Mol. Reprod. Dev*., **77** (2010), 105–113.

Gesto D. S., et al., *J. Am. Chem. Soc*., **135** (2013), 7146–7158.

Glazer E. S., et al., *J. Clin. Oncol*., **28** (2010), 2220–2226.

Glazer E. S., et al., *Translat. Oncol*., **4** (2011), 138–146.

Gong H., et al., *Leukemia*, **14** (2000), 826–829.

Guillen K. P., Kurkjian C., Harrison R. G., *J. Biomed. Sci*., **21** (2014), 65.

Guo H., et al., *Cancer Res*., 53 (1993), 5676–5679.

Haines R. J., L. C. Pendleton L. C., Eichler D. C., *Int. J. Biochem. Mol. Biol*., **2** (2011), 8–23.

Henriksen L. T., et al., *Pediatr. Blood Cancer*, **62** (2015), 427–433.

Hernandez-Espinosa D., et al., *Am. J. Pathol*., **169** (2006), 142–153.

Herzfeld A., Raper S. M., *Biochem. J*., **153** (1976), 469–478.

Ho E., et al., *Adv. Nutr*., **2** (2011), 497–510.

Hoffman R. M., *Biochim. Biophys. Acta*, **738** (1984), 49–87.

Holtsberg F. W., et al., *J. Control. Release*, **80** (2002), 259–271.

Hori H., et al., *Cancer Res*., **56** (1996), 2116–2122.

Horn F., *Hoppe Seylers Z. Physiol. Chem*., **216** (1933), 244–247.

Huang C. C., et al., *Oral. Oncol*., **48** (2012), 1227–1235.

Huang H. Y., et al., *Clin. Cancer Res*., **19** (2013), 2861–2872.

Ilies M., et al., *J. Med. Chem*., **54** (2011), 5432–5443.

Inoue H., et al., *Biosci. Biotechnol. Biochem*., **64** (2000), 2336–2343.

Izzo F., et al., *J. Clin. Oncol*., **22** (2004), 1815–1822.

Jenkinson C. P., Grody W. W., Cederbaum S. D., *Comp. Biochem. Physiol. B Biochem. Mol. Biol*., **114** (1996) 107–132.

Jeon H., et al., *Oncotarget*, **7** (2016), 67223–67234.

Kafkewitz D., Goodman D., *Appl. Microbiol*., **27** (1974), 206–209.

Kanyo Z. F., et al., *Nature*, **383** (1996), 554–547.

Kawaguchi K., et al., *Oncotarget*, **8** (2017), 85516–85525.

Keating M. J., et al., *Leuk. Lymphoma*, **10** Suppl (1993), 153–157.

Kelly M. P., et al., *Br. J. Cancer*, **106** (2012), 324–332.

Khangulov S. V., et al., *Biochemistry*, **37** (1998), 8539–8550.

Khoury O., et al., *J. Neurooncol*., **122** (2015), 75–85.

Kidd J. G., *J. Exp. Med*., **98** (1953), 583–606.

Kidd J. G., *J. Exp. Med*., **98** (1953a), 565–582.

Kim R. H., et al., *Cancer Res*., **69** (2009), 700–708.

Knodler L. A., Edwards M. R., Schofield P. J., *Exp. Parasitol*., **79** (1994), 117–125.

Knodler L. A., Schofield P. J., Edwards M. R., *Microbiology,* **141** (1995), 2063–2070.

Knott S. R. V., et al., *Nature*, **554** (2018), 378–381.

Kokkinakis D. M., et al., *Cancer Res*., **61** (2001), 4017–4023.

Krebs H., *Biochem. Education*, **1** (1973), 19–23.

Kreis W., Hession C., *Cancer Res*., **33** (1973), 1866–1869.

Kudou D., et al., *J. Biochem*., **141** (2007), 535–544.

Kudou D., et al., *Biosci. Biotechnol. Biochem*., **72** (2008), 1722–1730.

Kulis M., Esteller M., *Adv. Genet*., **70** (2010), 27–56.

Lam T. L., et al., *Pigment Cell Melanoma Res*., **24** (2011) 366–376.

Lavulo L. T., Sossong T. M. Jr., Brigham-Burke M. R., et al., *J. Biol. Chem*., **276** (2001), 14242–14248.

Lee, S.-M., Wroble M. H., Ross J. T., *Appl. Biochem. Biotechnol*., **22** (1989), 1–11.

Leopoldini M., N. Russo N., and M. Toscano M., *Chemistry* (Weinheim an der Bergstrasse, Germany), **15** (2009), 8026–8036.

Lomelino C. L., et al., *J. Biol. Chem*., **292** (2017), 19952–19958.

Lu X., et al., *J. Am. Chem. Soc*., **126** (2004), 5374–5375.

Lu X., et al., *Biochemistry*, **45** (2006), 1162–1172.

Maharem T. M., et al., *Int. J. Biol. Macromol*., **108** (2018), 88–97.

Mashburn L. T., Wriston J. C., *Arch. Biochem. Biophys*., 105 (1964), 450–453.

Mauldin J. P., et al., *Translat. Oncol*., **5** (2012), 26.

MendesA. V., et al., *Perspect. Psychiatr. Care*, **48** (2012), 138–148.

Miyazaki, K., et al., *Cancer Res*., **50** (1990), 4522–4527.

Moola Z. B., et al., *Biochem. J*., **302** (1964), 921–927.

Morris S. M., Jr., *Annu. Rev. Nutr*., **22** (2002) 87–105.

Morris S. M., Jr., *Am. J. Clin. Nutr*., **83** (2006), 508S–512S.

Morris, S. M., Jr., *J. Nutr*., **137** (2007), 1602S–1609S.

Motoshima H., et al., *J. Biochem*., **128** (2000), 349–354.

Muller H. J., Boos J., *Crit. Rev. Oncol./Hematol*., **28** (1998), 97–113.

Murakami T., et al., *Oncotarget*, **8** (2017), 35630–35638.

Ni Y., Schwaneberg U., Sun Z. H., *Cancer Lett*., **261** (2008), 1–11.

Oliveira E. F., et al., *J. Am. Chem. Soc*., **133** (2011), 15496–15505.

Ott P. A., et al., *Invest. New Drugs*, **31** (2013), 425–434.

Park, I. S., et al., *Br. J. Cancer*, **89** (2003), 907–914.

Patil M. D., et al., *Oncogene*, **35** (2016), 4957–4972.

Peterson R. E., Ciegler A., *Appl. Microbiol*., **17** (1969), 929–930.

Pokrovsky V. S., et al., *Curr. Med. Chem*., (2017); doi: 10.2174/09298673246 66171006132729.

Powrózek T., Małecka-Massalska T., *Transl. Cancer Res*., **5** (2016), S1531–S1538.

Qiu F., et al., *Sci. Signal.*, **7**(319) (2014), ra31.

Qiu F., Huang J., Sui M., *Cancer Lett*., **364** (2015), 1–7.

Raja R. A., Schmiegelow K., Frandsen T. L., *Br. J. Haematol*., **159** (2012), 18–27.

Ramani K., et al., *Hepatology*, **51** (2010), 986–995.

Reczkowski R. S., Ash D. E., *J. Am. Chem. Soc*., **114** (1992), 10992–10994

Reeds P. J., *J. Nutr*., **130** (2000), 1835S–1840S.

Richards N. G., Kilberg M. S., *Annu. Rev. Biochem*., **75** (2006), 629–654.

Rogers Q. R., Visek W. J., *J. Nutr*., **115** (1985), 505–508.

Rytting M., *Expert. Opin. Biol. Ther*., **10** (2010), 833–839.

Saeed H., et al., *Protein J*., **37** (2018), 461–471.

Sarquis M. I., et al., *Mem. Inst. Oswaldo Cruz*, **99** (2004), 489–492.

Savoca K. V., et al., *Biochim. Biophys. Acta*, **578** (1979) 47–53.

Schalk A. M., et al., *J. Biol. Chem*., **291** (2016), 5088–5100.

Scolnick L. R., et al., *Biochemistry*, **36** (1997), 10558–10565.

Shahriari S., et al., *Iran. J. Public Health*, **46** (2017), 357–367.

Sharma B., Singh S., Kanwar S. S., *Biomed. Res. Int*., **2014**, Article ID 506287.

Shen L.-J., et al., *Cancer Lett*, **191** (2003), 165–170.

Shrivastava A., et al., *Prep. Biochem. Biotechnol*., **42** (2012), 574–581.

Shrivastava A., et al., *Crit. Rev. Oncol. Hematol*., **100** (2016), 1–10.

Simon-Reuss I., *Biochim. Biophys. Acta*, **11** (1953), 396–402.

Smith D. W., Ganaway R. L., Fahrney D. E., *J. Biol. Chem*., **253** (1978), 6016–6020.

Soares A. L., et al., *Int. J. Pharmaceutics*, **237** (2002), 163–170.

Spector E. B., Rice S. C., Cederbaum S. D., *Pediatric Res*., **17** (1983), 941–944.

Stone E. M., et al., *ACS Chem. Biol*., **5** (2010), 333–342.

Sugimura K., et al., *Microbiol. Immunol*., **36** (1992), 667–670.

Swain A. L., et al., *Proc. Natl. Acad. Sci. USA*, **90** (1993), 1474–1478.

Syed N., et al., *Cell Death Dis*., **4** (2013), e458.

Szlosarek P. W., et al., *JAMA Oncol*., **3** (2017), 58–66.

Takaku H., et al., *Int. J. Cancer*, **51** (1992), 244–249.

Takaku H., et al., *Jpn. J. Cancer Res*., **84** (1993), 1195–1200.

Takakura T., et al., *Cancer Res*., **66** (2006), 2807–2814.

Tan Y., et al., *Anticancer Res*, **16** (1996), 3931–3936.

Tan Y., et al., *Anticancer Res*., **16** (1996a), 3937–3942.

Tan Y., et al., *Anticancer Res*., **17** (1997), 3857–3860.

Tan Y., et al., *Clin. Cancer Res*., **5** (1999), 2157–2163.

Tanios R., et al., *Leukemia Res*., **37** (2013) 1565–1571.

Thivat E., et al., *Anticancer Res*., **27** (2007), 2779–2783.

Thivat E., et al., *Anticancer Res*., **29** (2009), 5235–5240.

Thomas X., et al., *Bull. Cancer*, **97** (2010), 1105–1117.

Tone, M. D., Arch. *Biochem. Biophys*., **433** (2005), 279–287.

Tosa T., et al., *Appl. Microbiol*., **22** (1971), 387–392.

Vockley J. G., et al., *Genomics*, **38** (1996), 118–123.

Wetzler M., et al., *Blood*, **109** (2007), 4164–4167.

Wiswell O. B., *Proc. Soc. Experim. Biol. Med*. (New York, N Y), **76** (1951), 588–589.

Wong K .Y., Chim C. S., *Epigenomics*, **7** (2015), 985–1001.

Xu Y., et al., *Cancer Gene Ther*., **23** (2016), 287–294.

Yang, T. S., et al., *Br. J. Cancer*, **103** (2010), 954–960.

Yang H., et al., *Chem. Biol. Drug Des*., **84** (2014), 578–584.

Yao M., et al., *Acta Crystallogr. D Biol. Crystallogr*., **61** (2005), 294–301.

Yau T., et al., *Invest. New Drugs*, **31** (2013), 99–107.

Yoon, C. Y., et al., *Int. J. Cancer*, **120** (2007), 897–905.

Yoshioka T., et al., *Cancer Res*., **58** (1998), 2583–2587.

Young V. R., *J. Nutr.*, **124** (1994), 1517S–1523S.

Yu Q., et al., *Scand. J. Gastroenterol.*, **51** (2016), 1220–1226.

Zeidan A., Wang E. S., Wetzler M., *Expert Opin. Biol. Ther.*, **9** (2008), 111–119.

Zhang L., et al., *Cancer Lett.*, **357** (2015), 346–354.

Zhao R., Domann F. E., Zhong W., *Mol. Cancer. Ther.*, **5** (2006), 3275–3284.