MSc

2.º CICLO

FCUP
2022

U.PORTO

Novos Antídotos para a Toxina Oxidase de Aminoácidos-L do Veneno de Víbora

Miguel Tavares Lacerda

FC

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

Novos Antídotos para a Toxina Oxidase de Aminoácidos-L do Veneno de Víbora

Miguel Tavares Lacerda

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Bioinformática e Biologia Computacional

2022

U.PORTO

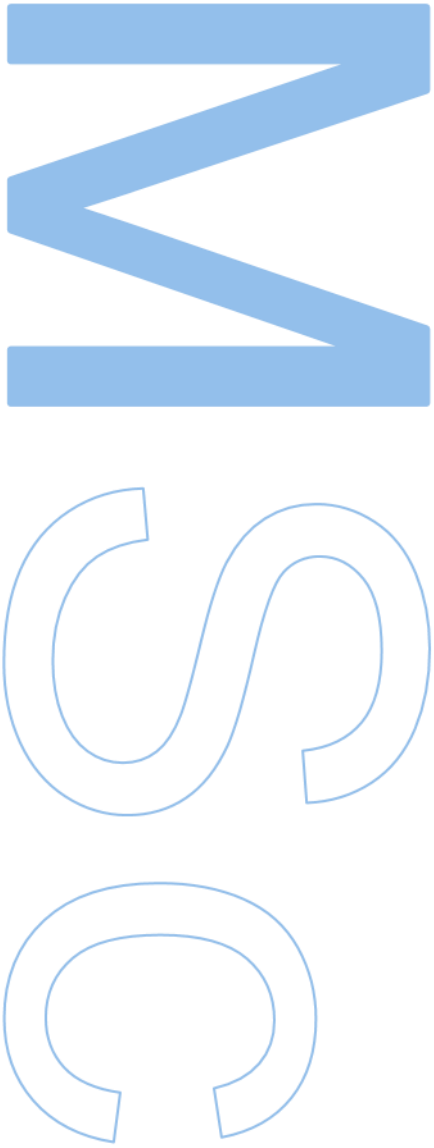
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

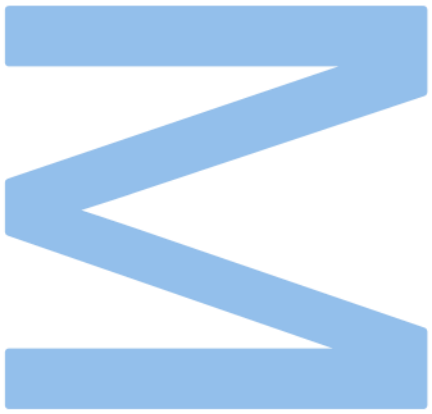
Novos Antídotos para a Toxina Oxidase de Aminoácidos-L do Veneno de Víbora

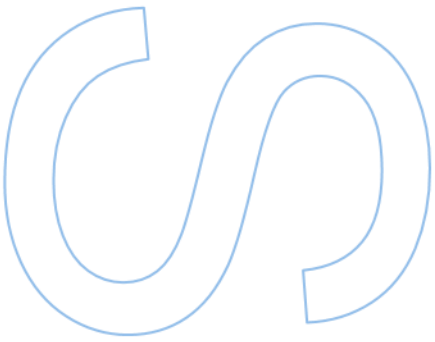
Miguel Tavares Lacerda
Mestrado em Bioinformática e Biologia Computacional
Departamento de Química e Bioquímica (DQBQ)
2022

Orientador
Professor Doutor Pedro Alexandrino Fernandes, Professor Associado,
Faculdade de Ciências de Universidade do Porto
Coorientador
Professora Doutora Maria João Ribeiro Nunes Ramos, Professora Catedrática,
Faculdade de Ciências da Universidade Porto
Professora Doutora Ana Rita Pires Gaio, Professora Auxiliar,
Faculdade de Ciências da Universidade do Porto




U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO







Declaração de Honra

Eu, Miguel Tavares Lacerda, inscrito(a) no Mestrado em Bioinformática e Biologia Computacional da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação/relatório de estágio/projeto "Novos Antídotos para a Toxina Oxidase de L-Aminoácidos do Veneno de Víbora" reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação/relatório de estágio/projeto "Novos Antídotos para a

Toxina Oxidase de L-Aminoácidos do Veneno de Víbora" declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação/relatório de estágio/projeto "Novos Antídotos para a Toxina Oxidase de L-Aminoácidos do Veneno de Víbora" quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Porto, 30 de setembro de 2022

Agradecimentos

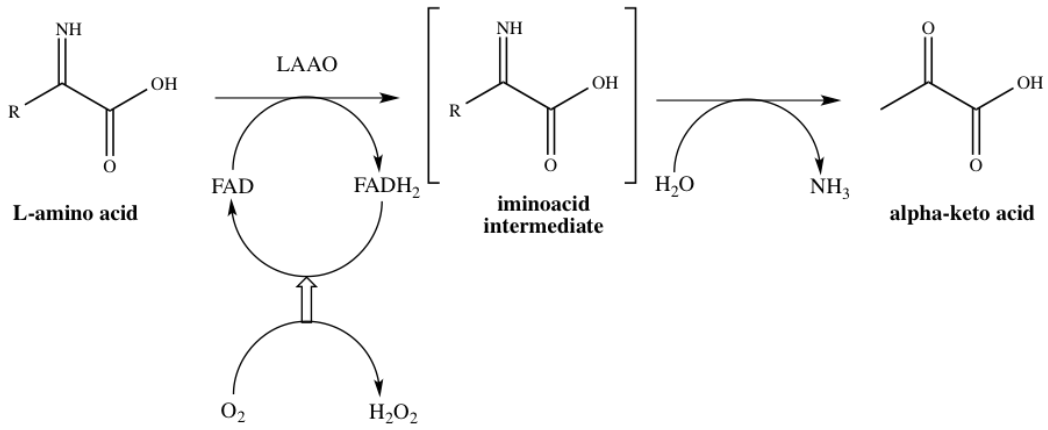
E assim uma fantástica travessia se finda. Gostaria de agradecer a quem esteve comigo durante esta viagem, tanto nos altos como nos baixos. Aos meus amigos mais chegados aqui do Porto, por me ajudarem sempre a desanuviar sempre quando precisava. À minha família por me apoiar de todas a maneiras possíveis e mais algumas. Sem vocês não teria sido possível. Aos meus gatos, que apesar de não serem o meu animal de estimação de preferência aprendi a adorá-los como cães. Ao artista que aprendi a adorar muito por causa da pandemia, e que por muito tempo foi o meu companheiro de isolamento, Marc Rebillet.

Não menos importante, gostaria também de deixar uma imensa, gigantesca, incomensurável, palavra de apreço a toda a malta do grupo de química teórica da Faculdade de Ciências da Universidade do Porto, não só por me terem acolhido ao longo destes 3 anos e me terem feito sentir sempre em casa (ou fora dela, no caso do isolamento!), mas também pela vossa ajuda inestimável neste trabalho, que em seguida apresento. Queria deixar um agradecimento em especial à Doutora Ana Oliveira, pelo seu trabalho incansável e pela sua ajuda inestimável na realização deste projeto. Queria também agradecer ao professor Doutor Pedro Alexandrino por tudo o que me ajudou a atingir durante esta viagem. Por fim, gostaria também de agradecer à professora Doutora Maria João Ramos, que em conjunto com o professor Doutor Pedro Alexandrino me concederam esta oportunidade.

Abstract

L-amino acid oxidases (LAAO) are FAD-dependent oxidoreductases enzymes. They are present in many organisms such as insects, fungi, bacteria, and snakes. For the latest, venomic studies showed that LAAO toxins can represent up to 20% of snake venom composition. Nonetheless, the abundance of this toxin depends on the species, gender, age, habitat, etc.

LAAO catalyzes the stereospecific oxidation of amino acids. The catalytic mechanism, proposed for the substrate hydrolysis in the LAAO family, occurs in a two-step reaction: a semi-oxidation followed by a semi-reduction. First, the NH_3^+ group of the amino acid is attacked by a His residue, converting it to an iminoacid. Then, the FAD molecule is protonated, with a hydrogen initially bonded to the His. Finally, FAD is re-oxidized, when the electrons from an oxygen present at the environment are transferred to the isoalloxazine group. Lastly, the iminoacid is converted to α-ketoacid, producing hydrogen peroxide and ammonia.



The LAAO toxins show harmful effects, as they can induce apoptosis, edema, aggregation and/or inhibition of platelets; they have hemorrhagic and anticoagulant properties. On the other side, these proteins also exhibit anti-bacterial, antiviral or leishmadicidal activities.

The dynamics and mechanism of LAAO are still unknown. After placing the substrate in the catalytic cavity of the LAAO making use of molecular docking calculations, we conducted Molecular dynamics studies to elucidate key structural aspects. First, we wanted to understand the flexibility of the protein, then we aimed to observe permanent

and/or transient cavities that could difuse the substrate or its products, or even alberg a ligand; finally, are goal was to determine the stability of the substrate in the catalytic pocket.

Thereafter we aim to determine the enzymatic mechanism, which will assist the rational development of drug candidates based on the transition state. These kinds of inhibitors will retard or even completely stop its occurrence.

The characterization of any enzymatic mechanism is about setting the energetic profile that describes the reaction. This can be achieved by the calculation of both the reaction's and Gibbs' free energy, for the reactants, transition states and intermediaries. These calculations can be carried by QM/MM methods, that apply both quantum mechanical (QM) methods, to describe parts of the protein directly involved on the chemical reaction (by electronic switches and conformational rearrangements) and Molecular Mechanics (MM) methods (used to describe the part of the enzyme that doesn't directly participate in the reaction). The calculated values can be compared to those experimentally obtained, to determine if a hypothesized reactional step is prone to occur.

In this project, the tridimensional structure of the LAAO from the *Bothrops atrox* (PDB entry: 5TS5*)* was completed using homology modeling techniques and the templates from *Bothrops atrox* and *Calloselasma rhodostoma*. Then the substrate (L-Phe) binding position within the catalytic pocket was estimated by molecular docking calculations. This step was followed by molecular dynamics simulations to verify the stability of the complex and the readjustments of the enzyme when in presence of the substrate ((zwi)-L-Phe)-HID223/FAD and the inimoacid ((imi)-L-Phe)-HIP223/FADH. These simulations were also important to select structures that were used as input to the QM/MM calculations. The last, will supply the energetic profile of the reaction.

Understanding the stability of the catalytic cavity and the key residues that are implied in the substrate binding and catalysis allowed us to select, using a Virtual screening of repurposing drugs, 11 compounds that will be tested in vitro by our experimental collaborators.

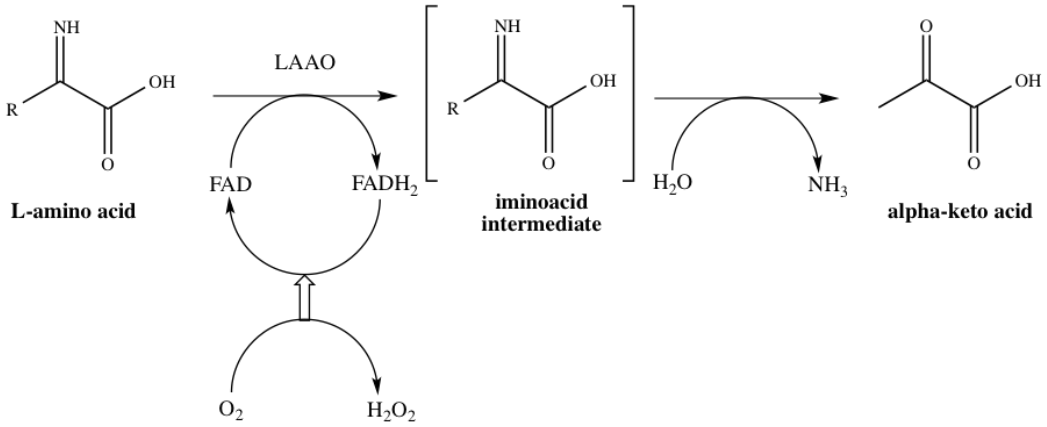
Keywords: Viperidae venoms, L-amino acid oxidases, Catalytic Cycle, QM/MM methods

Resumo

As oxidases de L-aminoácidos (LAAO) são enzimas que pertencem à superfamília das oxidases dependentes de FAD. Elas estão presentes em vários organismos, como insetos, fungos, bactérias e cobras. Para estas, estudos de venómica demonstram que estas toxinas podem representar até 20 % da composição total do veneno de uma cobra. No entanto, esta abundância depende de fatores como a espécie, género, habitat, etc.

As LAAO são responsáveis pela oxidação estereoespecífica de aminoácidos. O mecanismo catalítico, proposto para a hidrólise do substrato para enzimas desta família, é uma reação composta por duas etapas – uma reação semi-oxidativa seguida de um passo de semi-redução.

Primeiro, o grupo NH_3^+ do aminoácido é atacado por um resíduo de Histidina, levando à sua conversão em iminoácido. Seguidamente, a molécula de FAD é protonada, com o hidrogénio inicialmente presente no substrato. No final, o FAD é reoxidado, quando eletrões de uma molécula de O_2 reativo presentes no meio são transferidos para o anel de isoaloxazina. Por último, o iminoacido é convertido num α-cetoácido, produzindo peróxido de hidrogénio e amónia.



As toxinas LAAO possuem efeitos prejudiciais, uma vez que elas podem induzir a apoptose, edema e agregação/inibição de plaquetas. Elas possuem propriedades hemorrágicas e anticoagulantes. Por outro lado, estas proteínas também exibem atividades antibacterianas, antivirais e leishmidicidas.

A dinâmica desta proteína, assim como o seu mecanismo catalítico, ainda se encontram desconhecidos. Após a montagem do substrato no centro ativo, por técnicas de Docking

Molecular, foram conduzidos estudos de Dinâmica Molecular a fim de elucidar sobre aspetos-chave estruturais. Primeiro, foi definido como objetivo entender a flexibilidade da proteína; foi também proposto a observação de *pockets* permanentes e/ou transientes que pudessem difundir o substrato ou os seus produtos, ou até albergar um ligando. Finalmente foi também determinada a estabilidade do substrato no pocket catalítico.

Após este passo, foi proposto como objetivo a determinação do mecanismo catalítico, o que irá auxiliar no desenvolvimento racional de potenciais novos fármacos, baseados no estado de transição. Estes tipos de inibidores serão responsáveis por retardar ou mesmo interromper a ocorrência do mecanismo.

A caracterização de uma reação enzimática prende-se com a descrição energética do perfil reacional. Isto pode ser atingido através do cálculo de energias, tanto de reação, como a energia livre de Gibbs, para o reagentes, intermediários e produtos de reção. Estes cálculos podem ser feitos através de métodos QM/MM, que aplicam métodos de Mecânica Quântica, para descrever partes da proteína diretamente envolvidas na reação química (por trocas eletrónicas, e rearranjos conformacionais), e métodos de Mecânica Molecular para descrever partes das enzimas que não participam diretamente na reação. Estes valores calculados podem depois ser comparados com aqueles experimentalmente obtidos, a fim de determinar se um passo racional hipotético é suscetível de acontecer.

Neste projeto, a estrutura tridimensional da LAAO da espécie *Bothrops atrox* (entrada PDB: 5TS5) foi completa usando técnicas de Modelação por Homologia, com templates de *Bothrops atrox* e *Calloselasma rhodostoma*. Depois, a geometria mais favorável para ligação do substrato (L-Phenylalanine) dentro do *pocke*t catalítico foi estimada através de cálculos de *Docking* Molecular. Este passo foi seguido por simulações de Dinâmica Molecular para verificar a estabilidade do complexo e reajustamentos da enzima, quando na presença do substrato((zwi)-L-Phe) /HID223/FAD e o intermediário ((imi)-L-Phe) /HIP223/FADH. Estas simulações foram também importantes na escolha de estruturas para serem usadas como input para cálculos QM/MM. Estes últimos irão fornecer o perfil energético da reação.

Uma melhor perceção da estabilidade e atividade catalítica de resíduos-chave relacionados à ligação do substrato e na catálise, permitiu a seleção, através de um protocolo de *Virtual Screening* por *Drug Repourpusing*, 11 compostos que irão posteriormente ser testados *in vitro* por os colaboradores experimentalistas.

Palavras-Chave: Venenos de Víbora, Oxidases de L-aminoácidos, Ciclo Catalítico, Métodos QM/MM

Index

Introduction 1
1.Background 1
1.1Venom from a biological point of view 1
1.2The development of venomous toxins in snakes 2
2.Proteomic profile of the *Viperidae* Species 3
3. The *Viperidae* and the *Bothrops* genus 5
3.1 Geographical Distribution and Habitat 5
3.2 Bothrops atrox 5
3.2.1 Morphological Features 6
3.3 Venom and Pharmacological Effects 6
4. Conventional Treatment and Pitfalls 7
4.1The snake bite global initiative 8
5. L-Amino Oxidases 10
5.1 Classification 10
5.2 Overall Structure 10
5.3 Catalytic Mechanism 12
6.Computer Aided Drug Discovery Techniques 16
Aims 18
Methods and Materials 19
1.Construction of the Initial Model 19
1.1 Importance of the Initial Structure for Computational Studies 19
1.2 Homology Modeling 19
1.1.2 Construction of the Initial Model 21
1.2 Molecular Docking 22
1.2.1 Background 22

1.2.2 Construction of the Initial Complex 27
2. Parametrization of two systems 28
2.1 Setup of the Initial Structures 31
2.1.1 FADox 31
2.1.2 FADH 33
2.2 Geometry Optimization 34
3.Strutural Analysis of *Bothrops Atrox's* LAAO 36
3.1 Molecular Dynamics (MD) 36
3.2 MD Analysis 38
3.2.1 Root-Mean-Squared deviation Calculations 38
3.2.3 Root Mean Squared Fluctuation Calculation 39
3.2.4 Distance and Angles Analysis 41
3.2.4 Dihedral Angle Analysis 42
4. Study of the catalytic mechanism 43
4.1 Quantum Mechanics 43
4.2 Schrödinger Equation 44
4.3 *Ab initio* approaches 45
4.3.1 Hartree-Fock 45
4.3.2 Density Functional Theory (DFT) 45
4.4. Quantum Mechanics/Molecular Mechanics (QM/MM) studies 48
4.2 Initial Structures Preparation 52
4.3 Calculations Setup 52
4.4 Geometry Optimization 52
4.5 Determination of Stationary Points 53
5. Virtual Screening (VS) 54
5.1 Structure-Based Virtual Screening (SBVS) 54
5.1.1 Repurposing of small-molecule therapeutics (SMTs) 55

Results and Discussion 58

- Creating the 3D-LAAO model 58
- How does the substrate “sits” in the catalytic cavity? 59
- The System Parameterization and Geometry Optimization 60
- The Complex’s Motions 60
- Stability of the Complexes: RMSD Analysis 61
- RMSF Analysis 64
- Distances 65

Quantum Mechanics/Molecular Mechanics Studies 84

- Selection of the Initial Structure 84
- Preparation of the System – Definition of Layers and Freezing 85
- Geometry Optimization 86
- Step 1 – Deprotonation of L-Phe 87
- Key Aspects and Future Perspectives 88

The searching for novel LAAO inhibitors 90

- Structure-Based Virtual Screening 90
- Selection of the VS pipeline and scoring function 90
- Clustering Analysis 91
- Favorable Interactions 97
- Cross-Referencing 99

Conclusions 101

References and Bibliography 103

List of Tables

Table 1 - Comparison of incidence and mortality rates between tropical diseases formally recognized by WHO and snake bite envenoming..... 9
Table 2 -Types of scoring functions for the docking algorithm (Left) with examples (Right)..... 27

List of Figures

Figure 1 - Presence of venoms in many species and kingdoms, such as fungi, insects, bacteria and animals (left); displaying antibacterial, antifungical and leishmidicidal properties (right). 1
Figure 2 - Geographical distribution of Snakes..... 3
Figure 3 - Proteomic profile of Bothrops Atrox..... 4
Figure 4 – Reported geographical distribution of B.atrox..... 5
Figure 5 - Bothrops atrox, displaying a characteristic skin pattern for this species- 6
Figure 6 - Tertiary Structure of LAAO from Bothrops atrox (PDB code: 5TS5[1]). α-helices represented in light blue, β-strands represented in purple, and loops represented in cream. 11
Figure 7 - Tridimensional representation of LAAO's active site, and its respective residues. The substrate (L-Phe) is represented in black. Image made with Pymol software. 12
Figure 8 - General catalytic mechanism for LAAO, using L-Phe as substrate. (Up) – Semi-reduction step. (Down) – Semi-oxidation step. 13
Figure 9 - Representation of the conformational changes experienced by the “gateway” residues. (A) – The Zwitterionic form of the substrate prompts the His to adopt conformation A that favors the protonation of the substrate (L-Phe), while blocking the entrance of O_2. The aliphatic sidechain of Arg322 (conformation A) makes hydrophobic interactions with the aromatic part of the substrate, blocking its exit. (B) – Conformation B adopted by Arg322 makes the hydrophobic interaction becomes less pronounced, allowing the substrate to leave the active site. 14
Figure 10 - Stabilization interactions made with the substrate (L-Phe). Hydrogen bonds represented in dashed lines and hydrophobic interactions represented with surfaces. Image made with Pymol software. 15
Figure 11 - Representation of the proposed triad composed of Lys326-WAT-FAD(N5), in licorice. Image made with Pymol software. 16
Figure 12 - General scheme for a Homology Modeling protocol. A failure in the Model Validation step implicates the repetition of previous steps of the protocol, aimed at the refinement of the model. This process implies de repetition of the Template Selection, Multiple Sequence Alignment, and the Model Generation. Image made with BioRender[2] 20
Figure 13 - Modeller script for the creation of 6 models, using an alignment file (1), the templates (2), and the target (3)..... 22

Figure 14 - Schematic representation of the pose search strategies. (1) – Flexible receptor. (2) – Rigid receptor. 23
Figure 15 - GOLD script used for the generation of 10 poses. (1) – Flag for the definition of the anchor for the ligand. (2) – Flag for the definition of the data file to be used as ligand. (3) – Flag to prevent early termination of the docking, allowing for more conformational diversity in the solutions. (4) – Flag for the definition of the scoring function. (5) – Flag for the definition of the water molecules to be used. (6) – Flag for the definition of the receptor. 28
Figure 16 - Schematics of all the terms accounted in a force field energy calculation. 29
Figure 17 - pdb4amber script used for the initial preparation of .pdb files to be used in AMBER suite of programs 31
Figure 18 - Script used for the assembly of the complex with the fully oxidized form of FAD, the zwitterionic state of L-Phe, and the protein..... 32
Figure 19 - Script used for the generation of libraries pf parameters, topologies, coordinates, and 3D structures. (1) Script used for the iminoacid. (2) Script used for the complex. 34
Figure 20 - Scripts used for the energy minimization of the system. (1)- Water minimizations. (2) - Hydrogens minimizations. (3)- All system minimization. 35
Figure 21 - Script instructing SANDER module to conduct all the minimizations. 36
Figure 22 - Scripts used for the energy minimization of the system. (1) – Heating phase. (2) – Equilibration phase. (3) – Production phase. 38
Figure 23 - Script for the retrieval of the average structure over the last 10ns of simulation, and RMSD values for the protein, the substrate, and the cofactor. (1) - Complex with the fully oxidized FAD. (2) - Complex with semi-reduced FAD molecule. 39
Figure 24 - Script for the retrieval of the RMSF values for the protein, the substrate, and the cofactor. Analysis conducted using CPPTRAJ software..... 40
Figure 25 - Script used in the calculation of the distances and angles throughout the simulation. Analysis conducted using the CPPTRAJ software. 42
Figure 26 - Script for the retrieval of the dihedral angles throughout the simulation. Analysis conducted using CPPTRAJ software. 43
Figure 27 - Protein system treated over a QM/MM scheme, where the part of the system playing an active part on the reaction mechanism (in orange sticks) is treated under a Quantum Mechanics formalism, with the rest of the system (in blue lines) is treated with a Molecular Mechanics formalism. 49

Figure 28 - (Left) - Superimposition of the initial target system (in green) and the selected generated model (in gray). (Right) - Evaluation metrics of the 6 generated models with the chosen one highlighted in blue. 59
Figure 29 - Representation of the protein active site with the docked substrate (in green). Important residues for the stabilization of L-Phe a labelled in black, as noted by Anwar Ullah. 60
Figure 30 - RMSD values displayed by the Protein (up), Substrate (low), and FAD (low), for the 5 replicates over the course of 100ns, for the oxidized and semi-reduced system. 63
Figure 31 - RMSFs along the MD measured for the 5 replicates, for the oxidized and semi-reduced system. High RMSF zones highlighted in colors. 64
Figure 32 - Tridimensional representation of the protein, with high RMSF zones highlighted in colors. 65
Figure 33 - Close-up of the active center of the protein, highlighting the residues important for the reaction, either by stabilization (Arg90, Tyr372, Gly464) or by a direct participation (FAD and His223). Residues labelled in black Distances measured are represented in dashed. Atoms considered for the measurements labelled in red. 66
Figure 34 - (Up) Distances between the His223 and the N of L-Phe, calculated for the 5 replicates, for oxidized and the semi-reduced System. Distance measured for the NE2 atom of His223 represented in lines. Distance measured for the ND1 atom of His223 represented in dashed. (Down) Distances between the Cα of the L-Phe and N5 of the FAD molecule, calculated for the 5 replicates, for oxidized and the semi-reduced System. 67
Figure 35 - Distances between the NE, NH1 and NH2 of Arg90 and O and OXT of L-Phe, over the course of the simulations, for the oxidized and the semi-reduced System. (Up) – NE – O. (Mid) – NH1 – OXT. (Down) – NH2 – OXT. 69
Figure 36 - Distance between Gly 464 (OXT) and the L-Phe (N), for the Oxidized and the Semi-reduced system. 70
Figure 37 - Distance between Tyr372 (OH) and L-Phe (O and OXT) for the Oxidized and the Semi-reduced system. (Up) – OH-O. (Down) – OH – OXT. 71
Figure 38 - Distances between Arg322 (NH1, NH2, and NE) and L-Phe (OXT and O), along the simulations, for the oxidized system. (Up) – NH1 – OXT and NH1 – O, respectively. (Mid) – NH2 -OXT and NH2 – O, respectively. (Down) – NE – OXT and NE – O, respectively. 72

Figure 39 - Distances between Arg322 (NH1, NH2, and NE) and L-Phe (OXT and O), along the simulations, for the semi-reduced system. (Up) – NH1 – OXT and NH1 – O, respectively. (Mid) – NH2 -OXT and NH2 – O, respectively. (Down) – NE – OXT and NE – O, respectively. 73
Figure 40 - Distance between Asp224 (OD1 and OD2) and His223 (ND1 and NE2 for the semi-reduced system) along the simulations and for the oxidized and semi-reduced system. (UP – OD1 – ND1/OD1 – NE2). (Down) – OD2 – ND1/OE2 – NE2. 75
Figure 41 - Close-up of the active center of Bothrops atrox's LAAO, highlighting dihedrals that were measured (in lines). Dihedral angles labeled in red. Residues considered labeled in black. PS: Same labels for the dihedrals between Arg90 and Arg322. 76
Figure 42 - Dihedral angles displayed by the imidazole ring of His223, highlighting possible conformation changes, for the oxidized system and the semi-reduced System. (Up) – Psi dihedral. (Down) – Phi dihedral. 77
Figure 43 - Observed changes in dihedral angles for His223 and Arg 322, pointing for a change in conformation. (White) – Residues belonging to the first frame for the simulation of the oxidized system. (Green) - Last frame of the simulation of the semi-reduced system. Superimposition 78
Figure 44 - Dihedral angles displayed by the aliphatic chain of Arg322, highlighting possible conformation changes, for the oxidized system. 79
Figure 45 - Dihedral angles displayed by the aliphatic chain of Arg322, highlighting possible conformation changes, for the semi-reduced system. 80
Figure 46 - Dihedral angles exhibited by Arg 90 along the simulations, for the oxidized system. (Up) – C-CA and CB-CG dihedrals. (Down) – CB-CG and CG-CD dihedrals. 81
Figure 47 - Dihedral angles exhibited by Arg 90 along the simulations, for the semi-reduced system. (Up) – C-CA and CB-CG dihedrals. (Down) – CB-CG and CG-CD dihedrals. 82
Figure 48 - Dihedral angles exhibited by Tyr372 along the simulations, for the oxidized and the semi-reduced systems. (Up) - N-CA-CB-CG dihedral. (Down) – CA – CB – CG – CD2 dihedral. 83
Figure 49 – Superposition of the minimized average structure of the oxidized (in green) and the semi-reduced (in white) systems, highlighting the possible conformational changes. Residues belonging to the oxidized system labelled in black. Residues belonging to the semi-reduced system labelled in red. 84
Figure 50 - Analysis of distances and torsional angles between the FAD (N5) and (zwi)-L-Phe (Hα), comprised of the Cα and Hα atoms of the L-Phe, plus the N5 atom for the

isoalloxazine group of the FAD molecule, from structures arising from the production phase of MD. Considered atoms for the analysis are labelled in red. Considered residues are labelled in black. Measured distances and torsional angles are labelled in black. . 85
Figure 51 - The two analyzed strategies for the composition of the HL, with 151 atoms (A) and 115 atoms (B). 86
Figure 52 - Highlight of the system's HL after ME+EE. Catalytic important residues and distances labeled in black. catalytic important atoms labeled in red. 87
Figure 53 - Energetic profile of the Potential Energy Surface (PES) for the deprotonation step of the catalytic mechanism. The structure corresponding to the transition state is highlighted in green. 88
Figure 54 - (Left) - Distance between HID223 (ND1) and Asp224 (OD1) for all the simulations, for the oxidized system. (Right) – Tridimensional representation of the HID223 establishing a hydrogen bond with Asp224. 89
Figure 55 - Number of clusters generated for the top250 ranked compounds, when using a non-hierarchical clustering algorithm, with Tanimoto and Dissimilarity parameters of 0.6 and 0.8, respectively. 92
Figure 56 - Clusters' population, when using Ward's hierarchical clustering algorithm. (A) – 50 generated clusters. (B) – 100 generated clusters. 93
Figure 57 - Number of clusters generated with respective populations, when using a Diverse Set clustering algorithm. (A) – 50 generated clusters. (B) – 100 generated clusters. 95
Figure 58 - PCAs for the different clustering algorithm applied - Ward (left) and Diverse Set (right). (A) & (C) - 50 generated clusters. (B) & (D) - 100 generated clusters. 96
Figure 59 - Chemical diversity comparison between two sets of data, displayed over a violin plot and an example table highlighting minimum, average and maximum calculated dissimilarity values (A) - when 50 clusters were generated. (B) - when 100 clusters were generated. 97
Figure 60 - (A) -Representation of the protein with highlight on the more observed hydrogen bonds, represented in surface. Image made with Pymol software. (B) - Table with the absolute frequency of the top 5 more observed residues. (C) - Relative frequency the top 5 most observed residues (D) – Relative frequency of the number of "relevant" interactions made by the docked ligands. 98
Figure 61 - Absolute frequency of the residues with 3+ frequent interactions (Hydrogen Bonds (Imp.Res)), displaying Pi-stacking interactions Pi-Stacking interactions), and

observed Pi-Stacking interactions and residues with 3+ frequent hydrogen bond interactions (Both)..... 99
Figure 62 - Final Compound selection for the cross-referencing step, for Ward Clustering with 50 clusters and Diverse Setting with 50 clusters. 100

List of Abbreviations

3FTxs	Three-finger Toxins
AMBER	Assisted Model Building with Energy Refinement
ARG	Arginine
ASN	Asparagine
ASP	Astex Statistical Potential
B3LYP	Becke-3-parameter Lee-Yang-Parr
CAAD	Computer Aided Drug Discovery
CRISP	Cysteine-rich Secretory Proteins
CTL/SNALECS	C-type Lectins/Snalecs
DAAO	D-Aminoo Acid Oxidase
DASOX	D-Aspartate Oxidase
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
ECPF	Extended Conectivity Pharmacophore Fingerprint
EE	Eletrostatic Embedding
FAD	Flavin Adenine Dinucleotide
FMN	Flavin Mononucleotide
GAFF	General Amber Force Field
GGA	Generalized Gradient Approximation
GLY	Glycine
GTO	Gaussian Type Orbitals
HF	Hartree-Fock
HIS	Histidine
HITS	High Troughput Screening
HL	High Layer
ILE	Isoleucine
INT	Intermidiary
IRC	Intrinsic Reaction Cordinate
LAAO	L-Amino Oxidases

LBDD	Ligand Based Drug Discovery
LBVS	Ligand Bases Virtual Screening
LDA	Local Density Aproximation
LL	Low Layer
LYS	Lysine
MD	Molecular Dynamics
ME	Mechanical Embedding
MET	Methionine
MP	Metaloproteases
NCBI	National Center for Biotechnology Information
NP	Natriuretic Peptides
ONIOM	Our own n-layered Integrated molecular Orbital and Molecular
PDB	Protein Data Bank
PES	Potential Energy Surface
PF	Pharmacophore Fingerprint
PHE	Phenylalanine
PLA2	Type II Phospholipases
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative Structure–Activity Relationship
RESP	Restrained ElectroStatic Potential
RMN	Reciprocal Neighrest Neighbor
RMSD	Root Mean Squared Deviation
RMSF	Root Mean Squared Fluctuaction
RNA	Ribonucleic Acid
SANDER	Simulated Aneling with NMR-Derived Energy Restrains
SBDD	Structure Based Drug Discovery
SBVS	Structure Bases Virtual Screening
SMTs	Small-Molecule Therapeutics
SP	Serine Proteases
STO	Slater Type Orbitals
SV	Snake Venom

TRP	Triptofane
TS	Transition State
TYR	Tyrosine
VS	Virtual Screening
WHO	World Health Organization

Introduction

1.Background

1.1Venom from a biological point of view

The role of natural selection in the influence of gene evolution, namely regarding the venom evolution, has always fascinated the researchers.

The venoms are present in many organisms, spanning throughout multiple kingdoms, (such as Fungi, Animalia, and Monera) and lineages, like arthropods (bees, wasps, caterpillars, certain moths, scorpions, spiders, or centipedes), cnidarians [3] (e.g., jellyfish, stingrays, or surgeonfish), fishes, amphibians, mammals, and reptiles[4]. In fact, it was even proposed their presence in certain species of extinct animals, such as dinosaurs [5] and mammals[6].

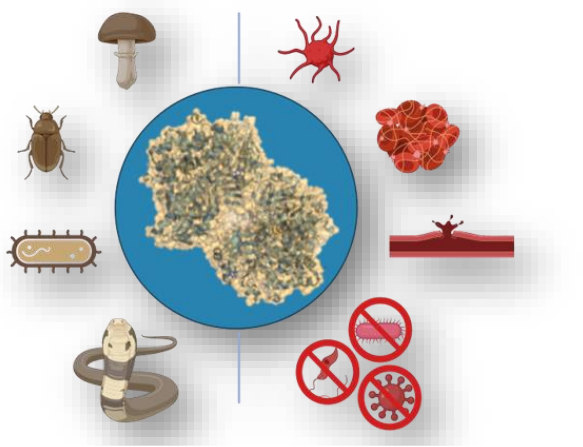


Figure 1 - Presence of venoms in many species and kingdoms, such as fungi, insects, bacteria and animals (left); displaying antibacterial, antifungical and leishmidicidal properties (right).

As a biological tool, venom is often employed for both predatory and defense reasons, with a similar objective – To immobilize and/or cripple the prey/attacker. Venoms are a mix of biologically active proteins, peptides, salts, and organic molecules, secreted through a specialized gland, resulting in an affliction of a wound, consequently triggering a multitude of physiological and biochemical processes. Venoms from snakes and other reptiles works by targeting several biological important molecules, culminating in severe local and systemic malfunctions in many cells and organs[7]. For instance, Viper's venom can cause local pain, swelling, tissue damage, coagulopathy, and for several species, damage to the kidneys, adrenals, and even the pituitary gland[8]. This can be triggered by several enzymes, ultimately leading to tissue necrosis, amputations or, in the worst cases, death.

Venom encoding genes have originated from the recruitment of body proteins or physiological protein encoding genes, into the venom gland [9]. Multiple copies of those genes were generated, over extensive duplication and positive selection [4, 10]. Some of those genes degenerated into pseudogenes, whereas others experienced a process of neofunctionalization [8], granting them of new biochemical functions. Since then, the observed diversification in different organisms can be explained through birth-and-death model of evolution - under the influence of a positive Darwinian selection [11], where some authors empathize that this biological diversification is due to gene duplication and rapid evolution [6].

1.2 The development of venomous toxins in snakes

As an many people worldwide suffer from snakes-bite envenoming every year, the study of the diversity among the various venom toxins in snakes represents an important subject among various fields, namely in the biomedical field or in evolutionary phylogenetics.

Snakes are carnivorous reptiles, belonging to the suborder of living snakes (Serpentes). This suborder comprises more than 3.900 species [12], even though venomous snakes belong to the species of advanced snakes (*Caenophidian*), composed of 2650 species, with 650 of them being considered venomous by the conventional anthropocentric definition[8], whereas currently, the medically important venomous snakes, deemed front-fanged snakes, all belong to 3 major classes – the *Viperidae (340 species)*, the *Elapidae (360 species)* and the *Atractaspidae (69 species)* [13].

The *Viperidae species* can be further divided into two clades: the *Viperinae* (True Vipers, 101 species) and *Crotalinae* (Pit Vipers, 239 species).

The venom glands of caenophidian (advanced) snakes are homologous, and it is suggested that the 3 families of front fanged snakes evolved from non-front fanged venomous snakes [10].

Geographically speaking, snakes are present in every continent except for Antarctica, with their distribution spanning from the Scandinavia all the way through Australia, being a species perfectly acquainted to all terrains and climates, including fossorial, arboreal, terrestrial, and aquatic environs, being able to live in climates ranging from arid deserts to the open ocean (Figure 2).

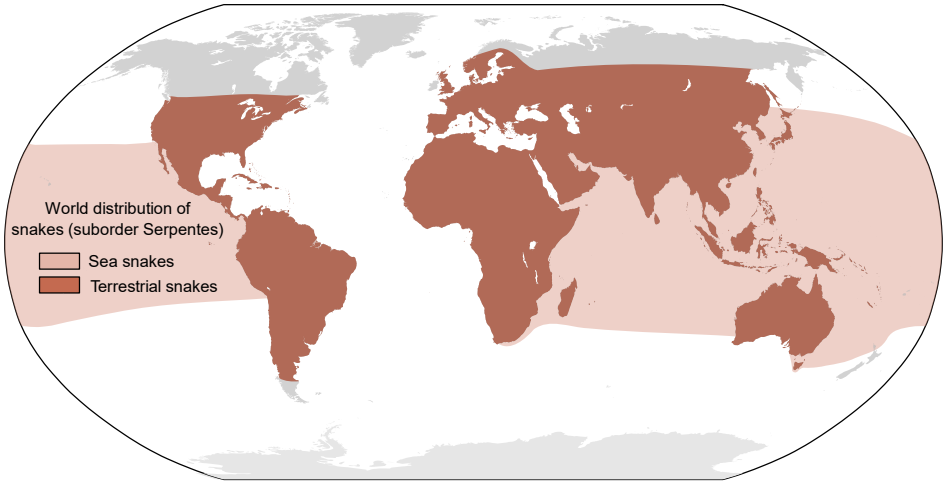


Figure 2 - Geographical distribution of Snakes.

In snakes, as for other specimen of the *Squamata* (reptiles), it has been proposed that Venom genes evolved from a common ancestor of *Toxicofera* (clade of snakes and lizards), consequently evolving for more complex and diverse observed in snakes and lizards, due to further toxin recruiting events. The further observed toxin diversity is obtained from unique gene mutations where functionally important toxin types are reinforced through adaptative evolution. The likelihood for neofunctionalization is enforced through random mutation, gene conversion and unequal crossing over, with derived activities emerging because of mutations of the surface chemistry. [8,10]

2.Proteomic profile of the *Viperidae* Species

Snake venoms are cocktails made of several dozens of highly effective proteins and several low molecular weight compounds. They can act synergistically and can sometimes form complexes [3]. Their venom composition is very diverse and varies from species to species. Elapidae venoms are composed by eight protein superfamilies, whereas viperines have eleven protein families and ten protein families were identified for crotalines.

The most common proteins in the venom are Type II Phospholipases (PLA2), the metalloproteases (SVMP), serine proteases (SVSP) and Three-finger toxins (3FTxs). There are also 6 families of proteins deemed secondary, given the proportion of the venom that they can achieve: cysteine-rich secretory proteins (CRISP), L-aminoacid

oxidases (LAAO), Kunitz peptides, C-type lectins/snalecs (CTL/SNALECS), desintegrins and natriuretic peptides (NP).

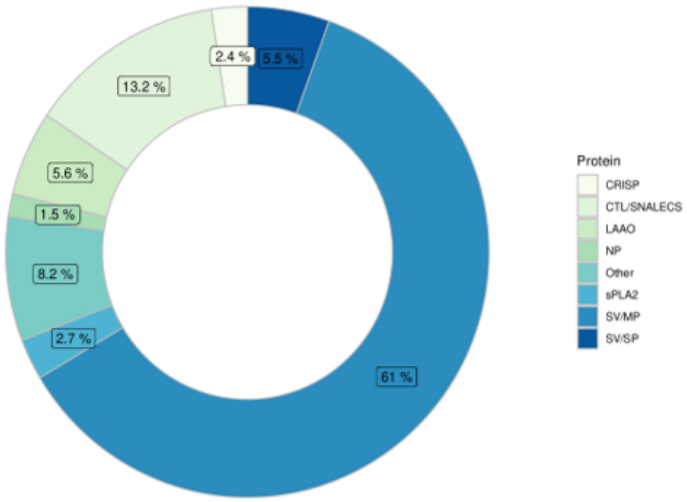


Figure 3 - Proteomic profile of Bothrops Atrox.

The main difference between the referred species of venomous snakes is the overall distribution and number of each protein family. For instance, the main difference between elapids and vipers is the presence of 3FTx in elapid venoms and the virtual absence in viper venoms. Elapids also tend to be less diverse in the range and number of protein families, largely consisting of PLA2 and 3FTx. [10]

Particularly in the *Viperidae* species, PLA2 represents the most widespread protein family, where it can reach a venom's proportion of 51% (*Agkistrodon contortrix*) for crotalines, and 65% (*Vipera nikolskii*) for viperines. On the other hand, metalloproteases represent the major protein family in viper venoms, where is present in virtually all the species, reaching values of 72% for viperines (*Echis ocellatus*), and 85% for crotalines (*Bothrops atrox*). L -amino acid oxidases are the most observed enzymes in crotalines, where it can reach up to of 20% (in *Rhinocerophis cotiara*) of the snakes' total venom. [10]

3. The *Viperidae* and the *Bothrops* genus

3.1 Geographical Distribution and Habitat

The *Bothrops* genus is particularly prevalent in Latin America, where they are responsible for most of the registered snake bite incidents. In Brazil, for instance, vipers, and more particularly the Bothrops genus are responsible for 85% of the ophidian envenomization. [14] In fact, of the 47 documented species allocated to the *Bothrops* family, 12 of them inhabit Brazil. [15] In Equator, another country largely constituting the amazon, there are currently 17 documented species of the *Bothrops* genus [16], where *Bothrops atrox* constitutes the most prominent species in terms of snakebite envenomization.

Differences in the venom composition of the Bothrops genus have been linked to several

factors such as ontogeny, geographical distribution, prey preferences and sex.



Figure 4 – Reported geographical distribution of *B.atrox*.

3.2 Bothrops atrox

Bothrops atrox is a species of pit viper that is widely distributed in the northern region of South America. In fact, they are responsible for most of the snake mobility and mortality in the Amazon [17].

They tend to inhabit humid places, like evergreen lowlands and foothill forests, savannas, plantations of all sorts (i.e., bananas, cassava, African palm, and rice fields), pastures, rural gardens, even though they can venture into human dwellings [13].

3.2.1 Morphological Features

Bothrops atrox (also known as common lancehead) varies in length, depending on the sex. For instance, females have a mean length of 174.3 cm, from the snout to the tip of the tail, whereas males tend to be shorter, presenting a mean length of 135.3 cm. [13] Their color of skin ranges from gray, brown, olive, or green tonalities and the skin patterns vary heavily for camouflage. This species is also characterized for margined triangles with points that meet on the dorsal line. Its head is triangular in shape ("lance-shaped") and pointed. Their belly is either white, cream, or yellowish gray, with increasing amounts of peppering in the posterior parts, ranging from gray to black, that may fade again under the tail. [14].



Figure 5 - Bothrops atrox, displaying a characteristic skin pattern for this species-

Currently there are no registered subspecies of B*othrops atrox* [18], however, its morphological features can sometimes be erroneously confused with other species of B*othrops*, such as B*othrops asper, Bothrops brazili, Bothrops leucurus* or *Bothrops moojeni*. [13]

3.3 Venom and Pharmacological Effects

As stated, *Bothrops atrox* is a venomous species and represents the principal cause of snakebite mortality and morbidity in the Amazon region. These statistics can be

explained by its perfect camouflage and prevalence in agricultural areas, allied with its aggressive behavior.

They have a LD_{50} of 1.1–4.9 mg/kg in senior specimens [13], whereas the venom of juveniles is more inflammatory, lethal, hemorrhagic, and kills more quickly than that of adults. [19, 20]

Symptoms form this snake envenoming lead to intense pain, swelling, bruising, bleeding, blistering, coagulopathy, nausea, vomiting, numbness, loss of consciousness, soft-tissue infections, and necrosis. In worst case scenarios, it can even cause amputations (induced by tissue necrosis and bacterial infections at the site of the bite), permanent complications and disabilities, and even death (in 1–2% of cases) [21, 22] Critically envenomated victims die from intracranial hemorrhage, acute renal failure, hemorrhagic shock, acute pulmonary oedema (swelling caused by an excess of fluid in body tissues), or respiratory arrest. [19]

4. Conventional Treatment and Pitfalls

Intravenous administration of antivenom is the only therapy to counteract snake-bite envenoming. [13] Analgesics, fluid therapy, hemodialysis and antibiotic therapy are also used. Additionally, the use of dexamethasone as a preemptive measure could be useful to improve some of the local symptoms caused by the envenomation. [23]

The antivenoms commercially available are the complete molecule or F (ab') 2 fragments of immunoglobulins purified from the plasma of horses hyperimmunized with different mixtures of snake venoms. These venoms are usually selected based in the geographical distribution of the medically important snakes while studies on immunogenicity. [24]

There are many facilities in producing antivenoms in Central and South America, namely in Mexico, Costa Rica, Venezuela, Colombia, Peru, Bolivia, Brazil, and Argentina. The production is either monospecific or polyspecific bothropic antivenoms [25], depending on whether if the antivenom is comprised of a cocktail of antigens targeting several species, or solely one. For instance, polyespecific antivenoms produced for Bothrops' venom manufactured by several laboratories in Brazil, uses a mixture of the venoms *of Bothrops jararaca, B. jararacussu, B. moojeni, B. neuwiedii*, and *B. alternatus* as antigens.[26]

The production of a mixture of an antivenom that combines antibodies against several viper species allowed the development of a broader effective treatment. This achievement only was possible thanks to the great sequence and structural similarity

among the different *Bothrops* species. This phenomenon is called cross-immunization – where an antivenom from a different species is able to retard or completely reverse the venom´s action. Indeed, in the 1980s, a study based on cross neutralization between specific *Bothrops* venoms from 10 different species and antivenoms guided for the formulation of immunization pools in Brazil.[27] However, it excluded the venom of *B. atrox*. This issue is mitigated by the great similarity in the composition of *B. atrox* venom with other species included in the immunization pool. [28]

Indeed trials conducted with antivenoms produced in Brazil, Colombia, Ecuador and Costa Rica, confirmed the efficiency of such antivenoms for the treatment of *B.atrox* bitten patterns, while antivenoms venoms synthesized for other species were equally efficient in mitigating and or reversing all sign of envenoming detected both clinically and in laboratory. [28]

Concerns on how interspecific and intraspecific venom variation associated with the ontogeny or geographical distribution of snakes affects the effectiveness of therapeutic antivenoms against the Amazon *Bothrops* venom prevail. Issues like this arise given that the presence/abundance different families of enzymatic toxins may help to explain the eventual lack of neutralization of certain activities of antivenoms.

Several studies highlight the efficiency of an immunizing mixture of antivenoms in neutralizing *B. atrox* venoms form different regions, also pointing for the quantitative differences in effective dosses of the various antivenoms. However, despite differences in different venoms' toxicity, pro-coagulant activity has been neutralized by antivenoms. [29]

However, the major pitfall of the antivenom strategy is the logistical part, particularly in the more remote and/or impoverished regions around the globe, where access to proper medical care is sometimes scarce or non-existent. In addition, antivenoms are costly, often scarce, poorly distributed in areas where they are most needed and may require cold-chain transport and storage.

4.1The snake bite global initiative

Snake bite is a neglected tropical disease that is recurrent in regions such as Africa, Asia, or Latin America. These regions tend to have a lot of low-income communities where snake bites contribute to yet another occupational problem and environmental risk. Every year, 5.4 to 5.5 million people are bitten by snakes, resulting in about 400 000

amputations, and between 81 000 and 138 000 deaths. These number does not fall into comparison to other tropical diseases (table 1); however, this problem was so neglected that the World Health Organization (WHO) only added it to its list of neglected tropical diseases in April 2009 [30].

Neglected tropical disease	*Deaths per year*
Snake-bite envenoning	*81.000-138.000*
Chagas disease	*14.000*
Cholera	*4.000*
Dengue haemorragic fever	*19.000*
Leishmaniosis	*51.000*
Japanese encephalitis	*14.000*
Schistosomiasis	*15.000*
Yelow fever	*60-100*

Table 1 - Comparison of incidence and mortality rates between tropical diseases formally recognized by WHO and snake bite envenoming.

While efforts are being made to try to ease this health problem, aspect such as availability of specific anti-venoms, a fast treatment procedure, or even the existence of qualified personnel will always be an issue inherent to poverty regions around the globe. The collective efforts of the world toxicology community and individuals working in the specialties of injury, disability, and health economics led to the launch of a global snake bite initiative at the inaugural conference in Melbourne, VIC, Australia, in November 2008.[31] This initiative consists of an interdisciplinary group of people focused on solving the problems addressing this tropical disease. The project was immediately endorsed by the International Society on Toxinology. [31]

The snake-bite initiative as several guidelines, one of which is to develop antivenoms that are safe, while improve their efficacy and affordability. [31] This objective can be achieved by creating antivenoms with an active principle that is specific to most of the indigenous region's snakes. By closely studying the enzyme underneath many of the venoms and more specifically its active site, one can aim to, in the future, develop a drug that works on many of snake inflicting wounds, thus contributing to a reduction on its price while augmenting its availability.

5. L-Amino Oxidases

Detected in 1944 by Zeller and Maritiz[32] , who studied the venom of *Vipera aspis*, L-amino acid oxidases (LAAO) are enzymes that catalyze the stereospecific oxidation of an L-amino acid, with the consequent formation of an α keto-acid and the concomitant release of hydrogen peroxide (H_2O_2) and ammonia (NH_4^+).

Snake venom LAAOs (SV-LAAOs) are present in virtually all of *Bothrops*' species, although in varying percentages. They are naturally homodimer with covalently linked cofactor, either Flavin Adenine Dinucleotide (FAD) or Flavin Mononucleotide (FMN), whereas in their denatured, inactive form, they are exhibited a monomer. In fact, their inactivation is caused trough a continuous exposure to low temperatures, resulting in an alteration of the overall display of the active site. [33]

They tend to exhibit catalytic specificity to long chain hydrophobic and aromatic amino acids and are characterized for being active in a wide range of pHs [34, 35] and their tridimensional structure, molecular mass, and isoelectric point quite varied.

Like other deemed medically important protein families, LAAO spark interest within the biomedical/biotechnological community as they are capable of inducing apoptosis in various cell lines, and the aggregation/inhibition of platelets, giving them hemorrhagic and anticoagulant properties. They also exhibit antimicrobial, antiparasitic and leishmidicidal properties. [35] Additionally, the existence of LAAOs may translate as a tool of protection against natural agents, parasites, and bacteria. [36]

The venoms displaying these enzymes tend to display a yellow color, related to the presence of the riboflavin present in the cofactors, namely in their oxidized form.

5.1 Classification

LAAO belong to the family FAD-dependent oxidoreductases. Other members of this family include Glycerol-3-Phosphate dehydrogenase, Sarcosine oxidase. D-amino acid oxidase (DAAO) or D-aspartate oxidase (DASOX) [35].

5.2 Overall Structure

LAAO is a glycoprotein with a molecular mass between 120 and 150 k_{Da}, in native (or dimeric) form, and 50-66 k_{Da} when denatured (monomeric form) [34]. Its primary structure is formed by 503 to 516 amino acid residues in its precursor or zymogen form, and 485 to 498 residues in the mature form [37]. Their primary and secondary structures are mainly conserved throughout the protein, with substantial differences in

the amino acid sequences being observed in the N and C-terminal parts of SV-LAAO. Sequence analysis indicates a high sequence identity (>83%) among Sv-LAAO, comparatively lower sequence identity with Pig kidney DAAO (<51%) and very low sequence identity with bacterial LAAO [35].

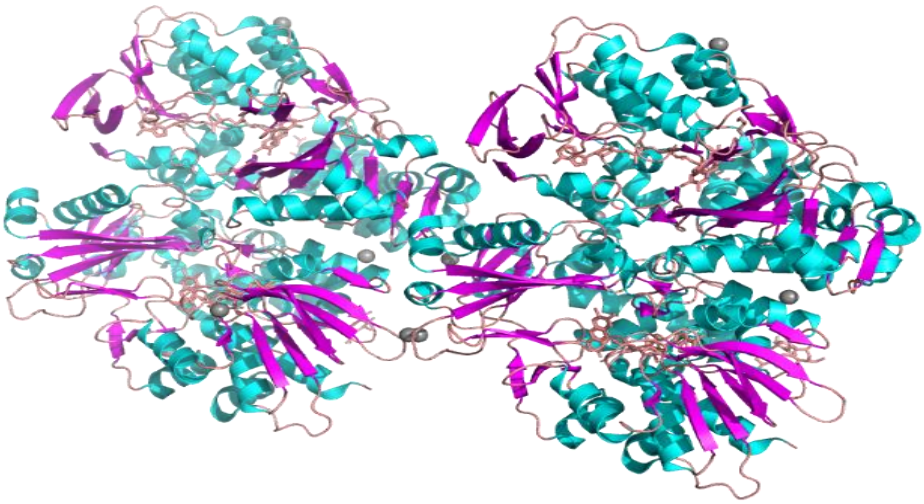


Figure 6 - Tertiary Structure of LAAO from Bothrops atrox (PDB code: 5TS5[1]). α-helices represented in light blue, β-strands represented in purple, and loops represented in cream.

The global folding of this enzyme is composed of seventeen alpha-helices, twenty-two beta strands and several loops, assembled in three domains:

1. FAD-binding domain.
2. Substrate binding domain; and
3. Helical domain. [35, 38]

Between the substrate and the helical domains there is a well-defined tunnel, which provides the access of the substrate to the active site [39]. Other key characteristics are the existence of two disulfide bridges and nature of the amino acids in the loops. These residues vary both in composition and in length, hereby altering the electric charges of the surface, which may impact the specificity to the substrates.

Moreover, another key structural aspect is the existence of several glycosylated sites., normally containing 3-4% (up to 12%) of carbohydrates [35]. These glycosylation sites (normally two sites present in LAAO – Asn 172 and Asn 361, with one reported for *Bothrops atrox*, located in the residue Asn172) are conserved and might assist the exit of the hydrogen peroxide [35, 38]. Some studies pin point the role of zinc ions [35], in the

LAAO's from *Vipera ammodytes ammodytes* and *B. atrox,* to stabilize the enzyme dimerization interface.

The active site is located between the FAD-binding domain and the Helical domain, and is formed by the FAD molecule, and residues of Arg90, His223, Phe227, Lys326, Tyr 372 and Trp375, plus a conserved water molecule near Lys326 residue. [35]

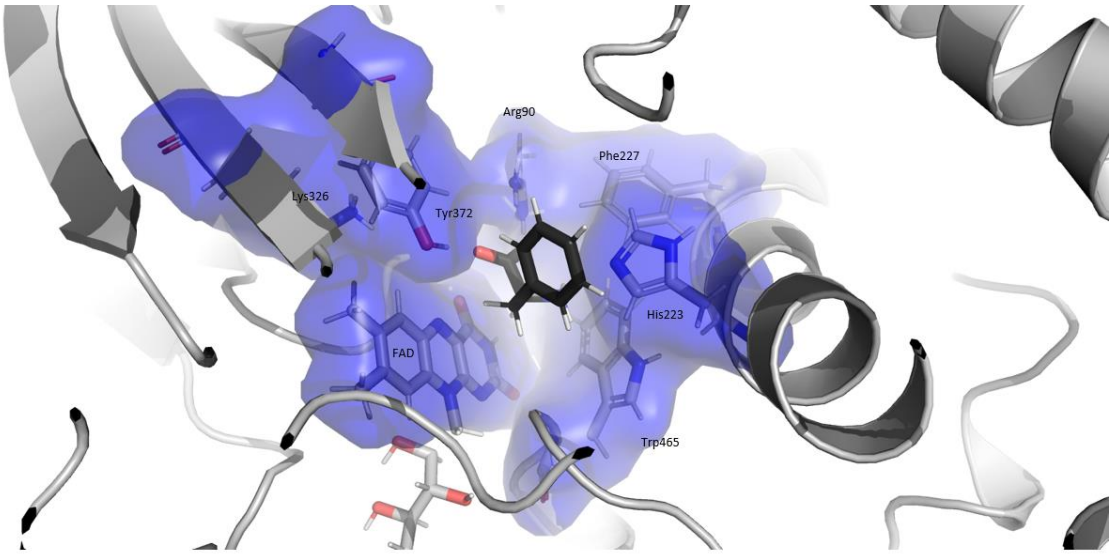


Figure 7 - Tridimensional representation of LAAO's active site, and its respective residues. The substrate (L-Phe) is represented in black. Image made with Pymol software.

Because LAAO convert a L-amino acid into α-keto acid, at some point during the catalysis, it needs to be close to O_2 molecules. Other key aspect for this enzyme is the presence of a second tunnel to aid the inlet of O_2, and the outlet of H_2O_2 molecules. [38,42] This exit route leads to the proximity a glycan moiety, that is thought to be responsible for the anchoring of the enzyme to the host cell [40].

5.3 Catalytic Mechanism

The overall reaction can be summarized by two half-reactions:

(1) An initial half reduction step, where the catalytic His receives a proton from the substrate: the zwitterionic amino acid; Then, the charge at the Cα of substrate, attacks the FAD molecule, and transfers the Hα to the FAD forming FADH an iminoacid. At this step the fully oxidized molecule of FAD is semi-reduced to FADH.

(2) An intermediate semi oxidation stage, where the FADH acts as an acceptor of H^+ and, in the presence of a reactive O_2 and water, is fully reduced to $FADH_2$, with the formation of hydrogen peroxide. Additionally, this stage sees the concomitant production

of an α-keto acid through a non-enzymatic hydrolysis of the iminoacid and the release of ammonia. [33, 40]

Indeed, for the full catalytic cycle to occur, the enzyme must accommodate two substrates - the amino acid and the O_2. The amino acid participates in the reductive semi reaction, whereas the O_2 takes part in the oxidative semi reaction. [40]

Although initially proposed as a different catalytic mechanism from DAAOs (D-Amino acid Oxidase), subsequent studies by Massey and Curtis [41] implied that the LAAO shared a similar mechanism. [39, 42, 43] Nonetheless, the atomistic details and the energetic catalytic profile are still unknown.

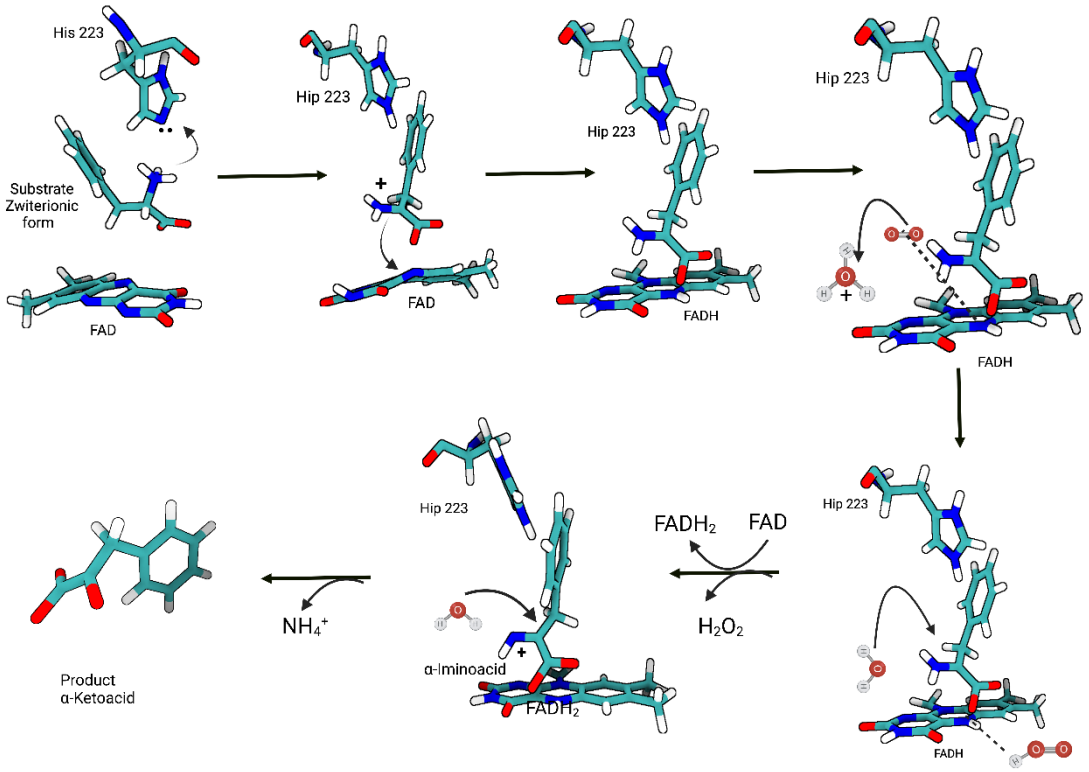


Figure 8 - General catalytic mechanism for LAAO, using L-Phe as substrate. (Up) – Semi-reduction step. (Down) – Semi-oxidation step.

The catalysis commences with the entrance of the substrate, in zwitterionic state, through a funnel-shaped tunnel, defined between the substrate-binding domain and the helical domain, leading to the active site. The substrate is trapped in a region comprised by the residue His223 and Arg322. It has been proposed that, upon the entrance of the substrate in the catalytic pocket, these residues undergo a conformational change that prevents the amino acid to leave to active site, while blocking the entrance of O_2, almost

acting as a gateway [40]. Indeed, the protonated state of the substrate influences these residues, prompting an alteration in their side chains (see figure 9).

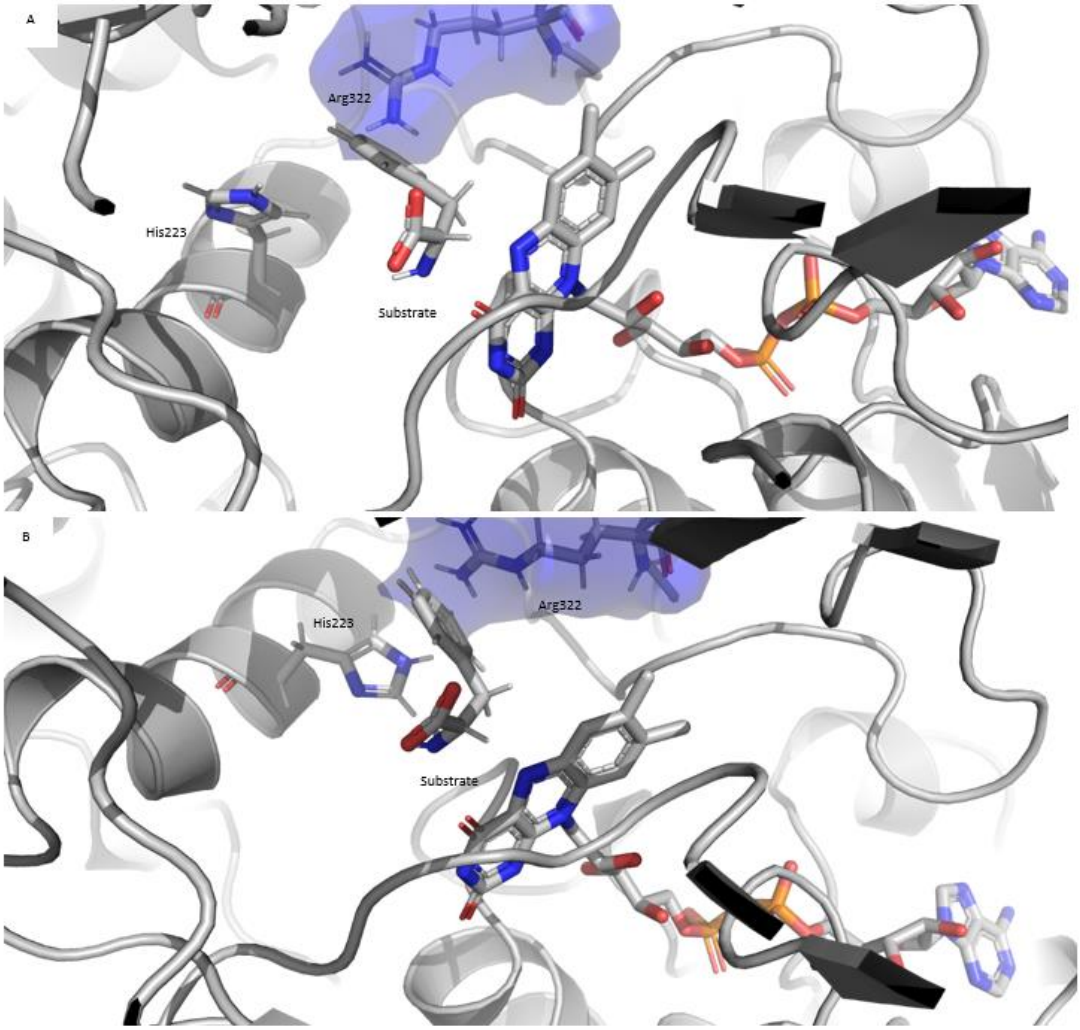


Figure 9 - Representation of the conformational changes experienced by the "gateway" residues. (A) – The Zwitterionic form of the substrate prompts the His to adopt conformation A that favors the protonation of the substrate (L-Phe), while blocking the entrance of O_2. The aliphatic sidechain of Arg322 (conformation A) makes hydrophobic interactions with the aromatic part of the substrate, blocking its exit. (B) – Conformation B adopted by Arg322 makes the hydrophobic interaction becomes less pronounced, allowing the substrate to leave the active site.

According to *Moustafa et al*. [40], the Arg residue in the gateway interacts with the carboxylic group of the substrate, keeping it in a favorable orientation for the catalysis to occur. Then, at the end of the reaction, these residues assume another conformation (conformation B), consequently allowing the release of H_2O_2. The interaction of the substrate in the active site is mediated through a H-bond interaction between the Arg90 side chain and Tyr372 (sidechain). The amino group of the substrate, positively charged,

is stabilized by the carboxylic group of Gly464 backbone. Additionally, the phenyl group of the substrate sits in a hydrophobic region composed by the Ile430, Ile374 and Phe227.

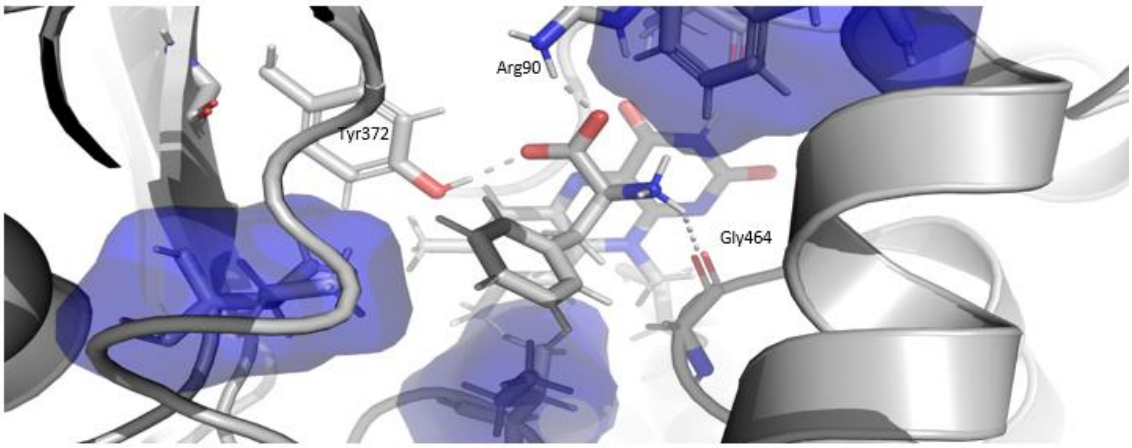


Figure 10 - Stabilization interactions made with the substrate (L-Phe). Hydrogen bonds represented in dashed lines and hydrophobic interactions represented with surfaces. Image made with Pymol software.

To date, two separate mechanisms of action have been proposed for the semi reductive half reaction – The Carbanion mechanism, where the proton is transferred, leaving a negative charge on the alpha carbon, followed by the transfer of two electrons in either one or two steps; The Hydride transfer mechanism, where the hydrogen is transferred simultaneously with two electrons. Studies on structural, mechanistic, and theoretical matters (mainly conducted on DAAO) have corroborated the hydride approach. [42]
This reaction initiates with the oxidation of the substrate, with the His residue working as a base, deprotonating it. This results in excess of positive charges in the molecule, causing a transfer of lone pair of electrons from the amino nitrogen atom to the C_α to form to imine. This is followed by the formation of a Michaelis complex.
Reports suggests that the Michaelis complex of the enzyme might consist in the oxidized enzyme plus the aminoacidic substrate bound in a distorted tetrahedral geometry around the substrate C_α. This conformation, although distorted, is stabilized by surrounding interactions at the active site. A further distortion of this complex from the initial tetrahedral to planar conformation (resulting from the alpha carbon going from an initial sp3 character to sp^2) leads to the oxidation of the amino acid, by approaching a geometry that closely mimics that of the imino product [40]. This is accompanied by a direct hydride transfer between the C_α and the N5 of the isoalloxazine group of the FAD molecule, that acts as a receptor of protons.

The oxidative half reaction utilizes O_2 to regenerate the oxidized factor and produces H_2O_2. Doubts remain on whether the entrance and consequent binding of the O_2 happens after a release of the imino product, or if there is a formation of a tertiary product composed of the imino product, the enzyme and the O_2 molecule [41]. There is a conservated water molecule and a lysine residue near the cofactor, constituting a triad, a structural feature rather normal in this type of flavoenzymes. [40]

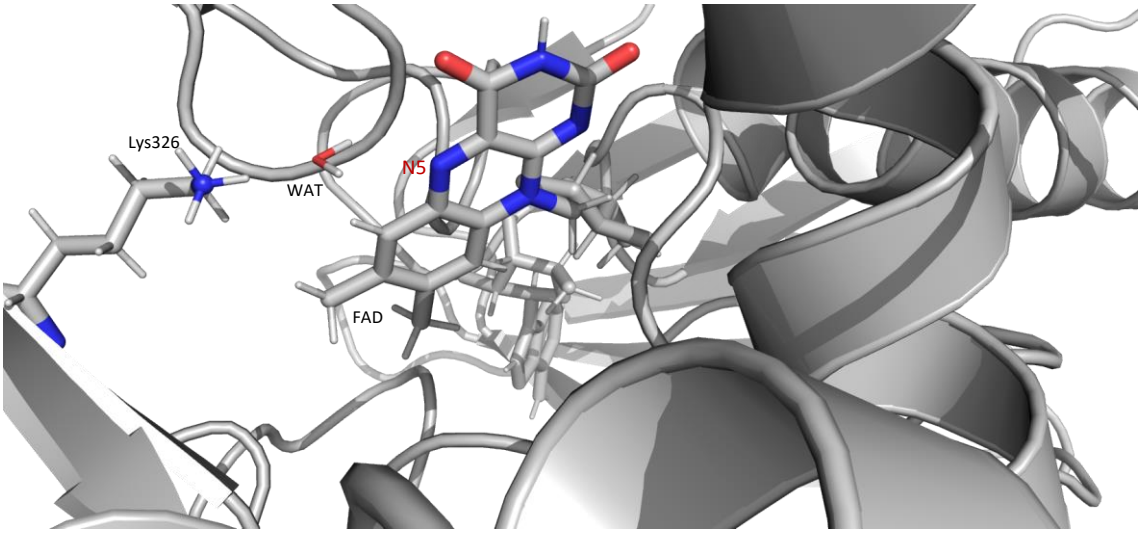


Figure 11 - Representation of the proposed triad composed of Lys326-WAT-FAD(N5), in licorice. Image made with Pymol software.

This water molecule acts as a hydrogen donor to the N5 of FAD and to one oxygen of the carboxylate group of the substrate. Additionally, it accepts a hydrogen from the side chain amino group of a lysine. Thus, the alfa carbon of the substrate experiences an oxidative attack, generating the product [40]. These interactions are disrupted upon the completion of the reductive half reaction.

Alternatively, it is proposed a role for this water molecule in the later production of H_2O_2 [40], from the flavin-hydroperoxy intermediate thought to be formed as an intermediate.

6.Computer Aided Drug Discovery Techniques

The increasing number of high-resolution crystallographic structures that emerged in the last decade, and the discovery of new techniques to determine larger structures, enabled the study of structural features of proteins, or even complexes, with greater detail. However, these techniques are static and do not take into account the dynamic behavior of living organisms and their physiological conditions. These structures are unable to

give sufficient information about transient conformational changes or even to characterize an enzymatic reaction. Moreover, experimental structure techniques require large amounts of very pure protein. To speed up the prediction of how a ligand can bind or not to a target, allowing the discovery of new drugs, researchers make use of Computational Aided Drug Discovery (CADD) methods. CADD comprise several computational tools used for the study of new drugs or for repurposing of existent ones. CADD can be used to identify new drugs based on a known scaffold — ligand-based drug discovery (LBDD) — or for the rational design of compounds based on their chemical complementarity with the target structure — structure-based drug discovery (SBDD) [44]. SBDD can be divided in two classes, depending on the employed strategy.

Indeed, the discovery of a new drugs through SBDD can be carried out using *de novo* design, where molecular docking calculations identify the most favorable chemical groups of a compound to bind their target, and the molecular determinants that assist the binding in a druggable cavity. Or, alternatively, it can be conducted though virtual screening (VS). The last filters, over a chemical library of thousands or even millions of compounds, the ones that are more likely to bind to the target by attributing each compound a binding score.

On the other hand, LBDD applies a similarity criterion of search and inspects close relative molecules based on structural and chemically similarity to known actives (molecules with verified biological activity against a particular target). The rational design of potent drugs can start on QM/MM studies. The characterization of the transition state (TS) of an enzyme reaction allows the design of TS analogs, which could be a very powerful inhibitor [45, 46]. Additionally, LBDD can employ QSAR methods to determine the correlation between structural/electronic properties and biological activity.

Aims

The primary objective for this work is to understand the dynamics and characterize the catalytic mechanism of LAAO present in *Bothrops atrox.* Additionally, we tried to identify novel LAAO inhibitors by investigating the possibility of binding of drugs in stage 1 to 4. To achieve these goals, several intermediate points were proposed:

1. Build the complete tridimensional structure of *Bothrops atrox* (PDB entry: 5TS5).
2. Construct the tridimensional complex of the enzyme and its natural substrate.
3. Conduct Molecular Dynamics calculations to analyze the flexibility of the protein, the integrity of its active site and substrate, the stability of the FAD(H) cofactor, and the presence of permanent and/or transient pockets.
4. Run a virtual screening protocol over a pre-compiled chemical library of drugs in order to select compounds that could be used to another propose, i.e, block the action of the enzyme in study. The efficiency of the compounds will then be evaluated *in vitro* by our collaborators in South America.
5. Determine the enzymatic mechanism of LAAO, and its energetic profile, making use of the ONIOM ('Our own n-layered Integrated molecular Orbital and Molecular mechanics') QM/MM method implemented in the Gaussian software. The method will allow the calculation of the free energies at the different catalytic steps of the enzyme reaction.

Methods and Materials

1.Construction of the Initial Model

1.1 Importance of the Initial Structure for Computational Studies

The initial structure selected for this work was the L-Amino oxidase toxin present in the *Bothrops Atrox* species (PDB entry: 5TS5 [1]). Since the selection of the initial structure represents a pivot point in any computational study, a modeling protocol was employed to complete the PDB structure determined by X-ray crystallography, followed by a global refinement.

1.2 Homology Modeling

In the absence of an experimentally determined protein structure, there are three methods one can follow in order to predict the 3D structure of the protein: methods *ab initio* [47], used when the protein that we want to study doesn't have homologous structures that would help us determine its structure. It is therefore applied to construct a model from scratch, based on physical principles. However, this technique does not produce very accurate models and its usually only successfully applied in small proteins (up to 120 amino acids). [48]

Comparative/homology modeling is a method than produces results comparable to low-resolution experimentally determined structures. [49] It can generate an accurate 3D model for a protein, using as template one or more experimentally known proteins that have similar sequence. [50]

The construction of a model can be obtained through 3 distinct methods: rigid body assembly, resulting from the overlap of parts from distinct templates to construct a final model; assembly through segment corresponding, based on the partial knowledge of the protein target structure; as well as the division and consequential search for segments that are similar, both sequentially and structurally, to the target's. The model is then built by overlapping the atomic positions of each segment. Modeling by satisfaction of spatial restrains relies on mathematical principles to construct a structure that satisfies spatial (homology and stereochemical) restrains.

In 2021, the *AlphaFold* technique emerged and estimates the 3D structure of proteins by applying artificial intelligence The technique is suitable to predict proteins, with high atomistic accuracy even in cases that there is no homolog structure [51].

These methods produce similar outcomes and overall, homology modeling relies on the fact that homologous proteins are similar in structure and function, even when they share low sequence identity. In addition, proteins from the same family have a more conserved tertiary structure than their primary sequences, [52] because the number of forms or *folds* that a protein can adopt is limited and less variable than its primary structure [50]. This ensures that, when the sequence identity between two proteins is proven, structural similarity is usually implied.

Homology modeling can be synthetized in 5 steps; searching for structures related to our target protein, selecting the templates, aligning the sequences from the templates with our target, building a model and then evaluate the model (Fig. 12).

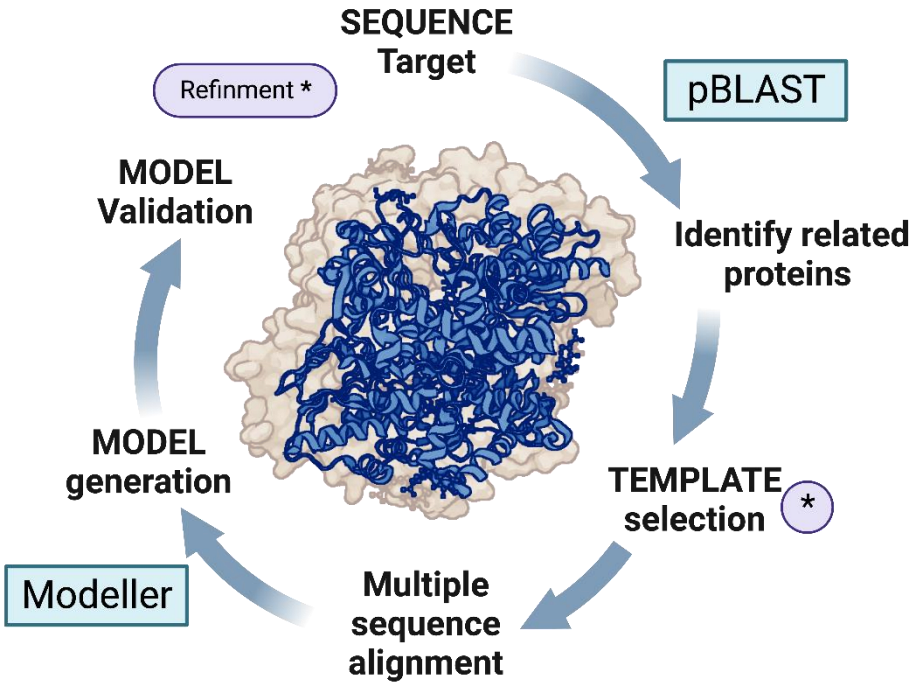


Figure 12 - General scheme for a Homology Modeling protocol. A failure in the Model Validation step implicates the repetition of previous steps of the protocol, aimed at the refinement of the model. This process implies de repetition of the Template Selection, Multiple Sequence Alignment, and the Model Generation. Image made with BioRender[2] .

The evaluation of the resulting model can be performed by looking at the criteria used to build the model (internal evaluation) or by performing an evaluation of the information not used for the construction or the model, also known as external evaluation. Internal validation refers to the evaluation of self-consistency, that checks if model satisfies the restraints used to calculate it.[50] This can be done by the assessment of the

stereochemistry of a model (e.g., bonds, bond angles, dihedral angles, and nonbonded atom–atom distances), using tools such as Procheck [52]and WhatCheck[53] .By contrast, external evaluation relies on the assessment of parameters that were not used in the construction of the model, like the quality of the used templates, the prediction of "unreliable regions" of the models by the calculation pseudo-energy profiles. Nowadays, there are many computer programs and software in the Web that automate the modeling process. Despite being convenient and useful, for more challenging modeling cases, such as problematic alignments, proteins that include cofactors, or modeling of loops, the best results are achieved through the use of sophisticate modeling programs. As the 3D structure of a protein can determine its function, its unenviability represents a huge handicap in many biochemical fields. More precisely, drug discovery can be severely impacted by advances in this field, as structure of the protein would make possible the study of, not only the function the protein, but more importantly its interactions, active sites, or binging activities.

1.1.2 Construction of the Initial Model

The initial structure of *Bothrops Atrox* LAAO was retrieved form protein data bank (PDB entry: 5TS5. We have used the homology modelling technique to complete the target structure of LAAO from Bothrops atrox (PDB ID 5TS5). Specifically, it was missing the N-terminal region (residues 1 to 18) and several residues, in different protein regions, lacked electronic density and therefore the X-ray structure didn't capture the orientation of the side chains of the following residues:

- ASP 1, ASP 2, ARG 3, GLU 7, GLU 8, ARG 11, ASN 31, GLU 77, LYS 78, GLN 107, GLU 116, LYS 134, LYS 162, GLU 165, ARG 168, ARG 169, ASN 193, LYS 230, VAL 267, LYS 268, SER 277, GLU 278, LYS 279, LYS 302, LYS 310, GLU 405, GLU 442, ARG 481, SER 483, ARG 484.

Therefore, using MODELLER software, version 9.11, a refined 3D model, including the missing regions, was constructed using as template the tertiary structures of *Calloselasma rhodostoma (PDB* entry: 2IID) and *Bothrops atrox* (PDB entry: 5TS5).
The preference for completing the structure using a single template is explained by high sequence similarity between the LAAO enzymes of the vipers (87%).

```
from modeller import *
from modeller.automodel import *

env = environ()
a = automodel(env, alnfile='Align_2iid_5ts5.PIR',                    (1)
      knowns=('2iid_A','5TS5_A'),sequence='5TS5',                    (2) (3)
          assess_methods=(assess.DOPE)
)
a.starting_model = 1
a.ending_model = 6
a.make()
```

Figure 13 - Modeller script for the creation of 6 models, using an alignment file (1), the templates (2), and the target (3).

1.2 Molecular Docking

1.2.1 Background

Developed during the 1980's [54], molecular docking is used for the prediction of the geometry of association between a pair of proteins (protein-protein) or between a protein and its ligand. It can therefore be used to infer the ideal structure of association within an allosteric center, as well as geometries for a ligand to bind (pose) in this orthosteric pocket or catalytic cavity. This technique can also be used to predict the association between two proteins or DNA or RNA.

However, and as many tools, this technique presents some drawbacks, namely regarding protein flexibility. In some protocols this feature is disregarded, even though it is reported that protein flexibility is responsible for altering and even increasing the affinity between a protein and the ligand [55].

The protein-ligand docking protocol comprises two steps, where there is an iterative positioning search, normally conducted through a searching algorithm, followed by a scoring step, where to each of the found poses is attributed a specific score, restricted to a scoring function previously set.

The pose search can by conducted using a flexible (slow run) or with a rigid receptor (fast run). The flexible receptor approach follows the "induced fitting" paradigm, whereas the rigid receptor follows the "lock and key" paradigm. [56] (Figure 14)

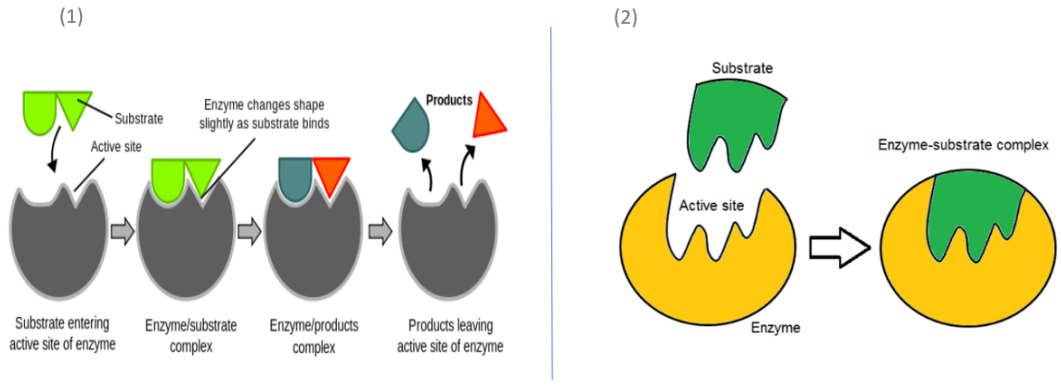


Figure 14 - Schematic representation of the pose search strategies. (1) – Flexible receptor. (2) – Rigid receptor.

The number of degrees of freedom included in the search algorithm represents a key aspect to determine its efficiency [57]. In a real biological system, ideally the system includes the ligand, the receptor, and the solvent molecules. All this complex represents an extensive computational effort, as the computed degrees of freedom makes it impeditive. This is normally addressed by the exclusion of the solvent molecules from the problem, or by the implicit modeling of them in the scoring functions or including only few water molecules that are present in the ligand cavity. However, even without the solvent molecules, the degrees of freedom associated with the ligand and the receptor still represent a computational burden. As such, different approximations to allow a more effective sampling of the searching space can be employed. A rigid body approach considers both the ligand as well as the protein as rigid bodies, only exhibiting 6 degrees of translational and rotational freedom. This approach completely disregards any kind of flexibility within the complex. The flexibility of the receptor is important as protein-ligand interactions involve conformation changes in the protein. However, the flexible receptor methodology represents a computational expensive protocol, for the number of degrees of freedom conferred to the protein.

Other approach applied to the pose search step, while ensuring minimal computational burden is to infer on the ligand flexibility while assuming a rigid protein receptor [57].

- Flexible Ligand - Search Docking

There are three general types of approaches to deal with the ligand flexibility - systematic methods, stochastic methods, and simulation methods.

- Systematic Methods

The first approach can be further divided in three methods - conformational search, that tries to incrementally rotate all the simple bounds until 360 degrees in order to explore all the conformational space, until all the possible conformations have been generated and evaluated. This method is quite expensive as it can lead to a "combinatorial explosion", since the number of generated conformations grows immensely with the number of rotatable bonds, making its applicability very limited. The fragmentation method tries to tackle the flexibility issue by incrementally growing the ligand, either by docking several fragments into the active site, followed by a linkage to recreate the initial ligand ("place and join" approach) or either by dividing the ligand rigid core fragments that are docked and flexible regions, subsequently added. Some examples of software using this technique are LUDI [58], FlexX [59], DOCK[60], ADAM [61], and Hammerhead [62].

Database methods, on the other hand, try to address the combinatorial explosion issue by using pre-existing libraries of generated conformations (ensembles), in order to deal with ligand flexibility. FLOG [63] is a software that follows this approach, by generating a set of 25 database conformations per molecule, based distance geometry, subsequently subject to a rigid docking protocol.

- Stochastic Methods

Stochastic search algorithms sample the conformational space by performing random changes the ligand(s). Then the resulting changes are either kept or discarded according to a given criteria. There are three types of approaches to this algorithm - the Monte Carlo (MC) method, Genetic Algorithm (GA) method and the Tabu Search Method. In the MC method the alteration in the ligands evaluated according to the Boltzmann Equation criteria. This poses an advantage in relation to Molecular Dynamics (MD) methods as a simpler energy function is used[64] . Examples of software following these methods are Prodock [65], ICM[66], MCDOCK [67], DockVision [68] and QXP [69].

Genetic Algorithms are based on concepts from the theory of biological evolution for docking. Contrary to MC or MD methods, this approach starts from an initial population of different conformations of the ligand with respect to the protein. Each of the conformations is defined by a several parameters, like orientation or translation (constituting a gene). The full set of ligand's variables are called a genotype whereas

their spatial location represents a phenotype. The given gene is then subject to biological events - mutations, crossings, and migrations - until a final population that optimizes a predefined scoring function is reached. Example of software following this method include AutoDock [70], DIVALDI [71], DARWIN [72], and GOLD [73].

Tabu search methods rely on the existence of generated and then stored restrictions in order to prevent the algorithm the revisit already explored regions, thereby promoting the analysis of new conformational space. PRO_LEADS [74, 75] is the most popular algorithm that employs this approach.

▫ Simulation Methods

Simulation Methods are other type of flexible ligand search methods that are based on the calculations of the solutions to Newton's equation of motion. The two major approaches that exist are the molecular dynamics (MD) and pure energy minimization methods. Energy minimization methods include direct searches (e.g., simplex), gradient methods (steepest descent), conjugate-gradient descent methods and least squared methods, otherwise not very used as only the local minima can be reached. Nevertheless, several discussed docking algorithms discussed above use energy minimizations as an intermediate step. Examples of software are ProDock [65], ICM [66], QXP [69], DARWIN [72], DOCK 4.0 [60], ADAM [61] and Hammerhead [62].

The evaluation and ranking of the computed poses are a key aspect of any Molecular Docking protocol. A scoring function should therefore allow the distinction of the true binding sites among all predicted poses or between active and random compounds.

▫ Scoring Functions

The scoring of the poses, and in particular the dominant pose, reflects the likelihood of a structure to be observed in an experiment. The abundance of each pose depends on the energy of the complex, in relation to that same pose. However, and in accordance with “molecular recognition”, there is a pose that is more frequent than all the others. In this situation only minor rotational oscillations are verified around the dominant pose (with RMSD values inferior to 2 Å).
The scoring reflects the intensity of attractive interactions - i.e., the energy required so that the ligand (and the protein, if a flexible body protocol is employed) adopts the docked

conformation, plus the desolation cost of both the protein and the substrate. As a rigorous scoring function would be too expensive, assumptions can be made in order to decrease the complexity of the scoring functions, with cost in terms of accuracy. This, on the other hand, represents a major drawback on scoring functions, the relation between speed and accuracy [76].

There are 3 main types of scoring functions – based on force fields, empirical and knowledge based. The force field-based scoring functions are based on the principle that the interaction energy is comprised of intra and intermolecular terms. The intramolecular terms arise from the adopted conformation, whereas intermolecular terms are given to the energy arising from electrostatic dispersions and/or repulsions. They are translated through a combination of Van der Walls energy term (using a Lennard-Jones potential) with an electrostatic energy term (using a Coulomb formulation, with a distance-dependent dielectric function that lessens the contribution from charge-charge interactions [76].

Empirical Scoring functions are designed to reproduce experimental data, based on the idea that binding energies can be approximated by the sum of several uncorrelated terms [76, 77]. The determination of the terms' coefficients used to calculate the binding energies is resolved by regression analysis, using experimentally determined binding energies or a training set of previously resolved receptor-ligand complexes.

Knowledge-based scoring functions are designed to follow rules and general principles statistically derived to reproduce experimentally determined structures. Typically, these methods use simple atomic intersections pair potentials thus allowing large compound databases to be efficiently screened. These potentials are based on the frequency of occurrence of contacts between different pair of atoms and other typical intersections in large datasets of protein ligand complexes of known structures. This causes that its application is restricted to the available information, in a limited set of structures.

Consensus Scoring tries to combine information from different scoring functions to try to compensate possible errors arising from individual scoring procedures. Correlation between several functions can negatively influence the outcome, since it can amplify an error, instead on balancing it. Bellow there is a table summarizing several existing scoring functions, as well as information related to them.

TYPE	EXAMPLES
FORCE-FIELD BASED	D-Score [78], G-Score [78], GoldScore [73]
EMPIRICAL	Böhm scoring function[79, 80] , F-Score [59], ChemScore [81], ChemPLP [82]
KNOWLEDGE BASED	Muegge´s Potential of Mean Force (PMF) [83-85] Drug-Score [86], Astex Statistical Potential (ASP) [66]
CONSENSUS SCORING	X-CScore [87], MultiScore [88], GFScore [71]

Table 2 - - Types of scoring functions for the docking algorithm (Left) with examples (Right).

1.2.2 Construction of the Initial Complex

Upon the construction the model, for the refinement of our initial structure, a molecular docking protocol was employed in order to find the optimal geometry for which the substrate can bind to the enzyme. For that matter, L-Phenylalanine was used as a substrate, using GOLD software, version 5.2.2 as the docking algorithm [73]. As for the algorithm parameters, the binding site center was defined using a methane fragment as an anchor. A cavity of 6 Å around this initial point was defined, with the respective cavity structure being extracted to a file. As stated earlier, L-Phenylalanine was used as the docking ligand, with the Genetic algorithm being set for 100 runs. We then applied 3 scoring functions as implemented in the software: GoldScore, ChemScore and ASP. The scoring function that revealed to infer the best resulting geometry for this substrate was ASP.

Having confirmed for the correct protonation state all the important residues in the reaction, two water molecules, present in the outskirts of the catalytic center, were added. Here, we allowed full flexibility to the waters, that could rotate or translate in a radius of 2 Å of its initial position, and the “toggle” option was selected. The last only take into account the water molecules that are favorable for the binding. Consequently, the algorithm was initiated. All this process was computed at its slowest rate, therefore producing the most accurate poses.

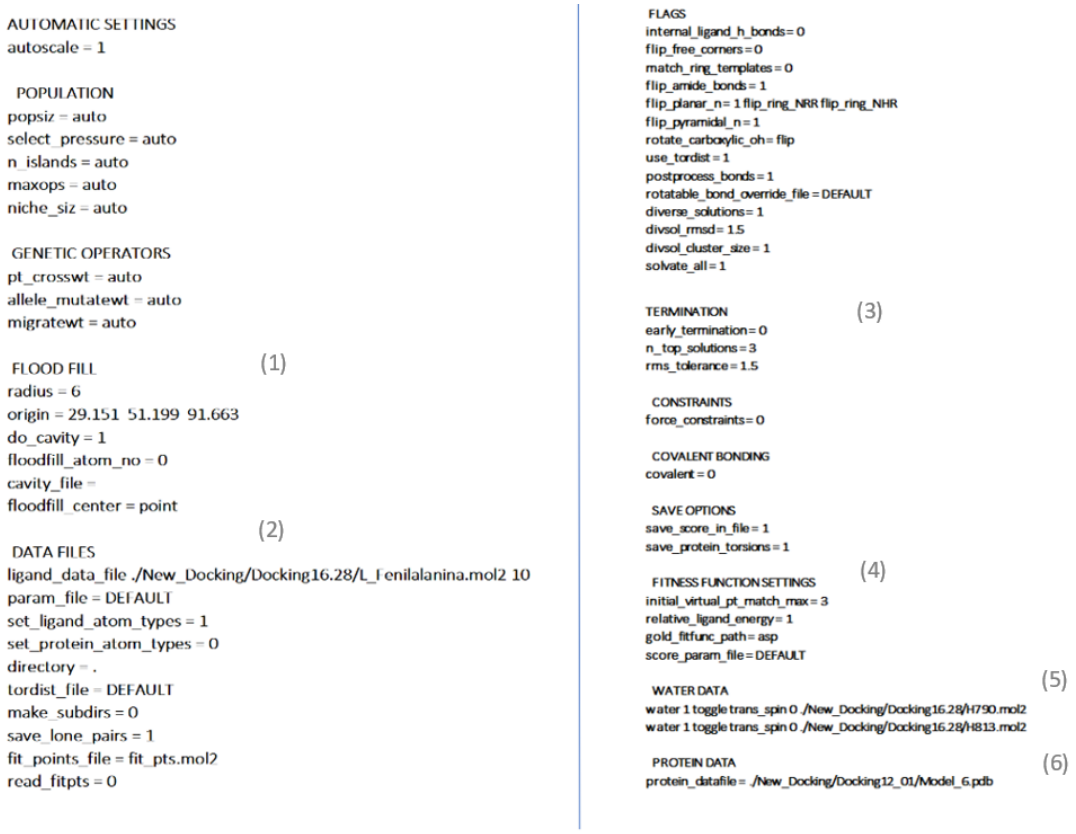


Figure 15 - GOLD script used for the generation of 10 poses. (1) – Flag for the definition of the anchor for the ligand. (2) – Flag for the definition of the data file to be used as ligand. (3) – Flag to prevent early termination of the docking, allowing for more conformational diversity in the solutions. (4) – Flag for the definition of the scoring function. (5) – Flag for the definition of the water molecules to be used. (6) – Flag for the definition of the receptor.

2. Parametrization of two systems

The energy of any molecular conformation can be calculated through its force according to its atom's positions and a set of parameters. A force field can therefore be used to calculate a system's energy – or rather the total potential energy – as the sum of mechanical contributions, as well as non-binding interactions, as depicted in the following equation:

$$E_{total\,potential} = \underbrace{E_{bond} + E_{\theta} + E_{\omega}}_{E_{intramol}} + \underbrace{E_{vdw} + E_{el}}_{E_{intermol}}$$

Equation 1 - Energy calculation of a force field.

Where the total potential energy of the system ($E_{total\ potential}$) is given by the Energy associated with the deformation arising from the bonds between two atoms (E_{bond}), the energy associated with the angle variation in a bond, and the energy associated with the torsion of dihedral angles. These terms are called intramolecular energy terms. The last two terms arise from energy contributions from non-binding interactions, both dependent (E_{el}) and independent of charge (E_{VdW}). These terms are meant to translate all the possible interactions between the different atoms in a system, according to the following schematics:

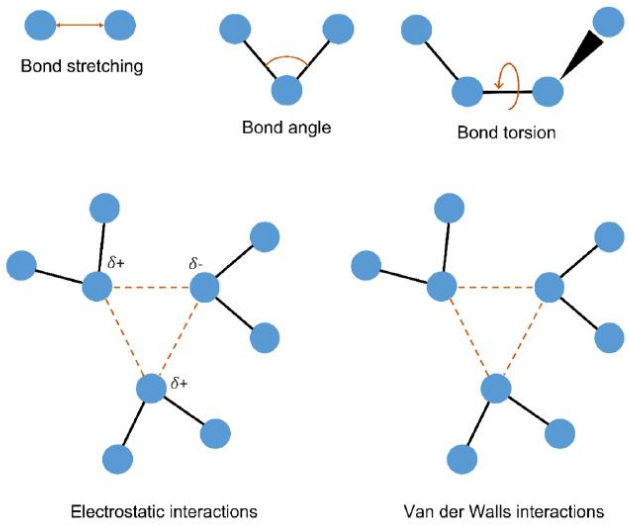


Figure 16 - Schematics of all the terms accounted in a force field energy calculation.

A parametrization of all the system's components is therefore required. Nowadays, there are many available biomolecular fields whereas, despite punctual differences, most of them, if not all, account all the terms depicted in the equation above. Examples of such force fields are AMBER [89], CHARMM [90, 91], GROMACS [92], OPLS [93]. For this work however, and in line with this team line of work, AMBER force field was use for all the simulation calculations regarding this enzyme.

- Assisted Model Building with Energy Refinement (AMBER)

AMBER is a suite of packages that, among other tasks, allows the molecular simulation of a complex, and consequent analysis [89]. It comprehends several libraries of

parameters designed to fully characterize every residue, as well as hydrogens in physiological conditions. These libraries describe the atoms' type, their force and equilibrium constants, for every bond, angle, and dihedral, as well as the Van der Walls forces and partial charges for every atom, and hydrogen, according to the equation:

$$U(\overrightarrow{R}) = \sum_{bonds} K_b\ (b - b_0)^2 + \sum_{angles} K_\theta\ (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}\ [1 + \cos(n\emptyset - \gamma)] + \sum_{nonbonded} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_1 R_{ij}}]$$

Equation 2 - AMBER force field energy calculation.

Where K_b, k_θ, and V_n represent force constants related to bond connections, angles, and dihedrals, respectively; b, θ and ∅ represent the connection length, angle and dihedral, with the index 0 concerning the equilibrium values for each individual term; γ corresponds to the phase angle (for values between 0 and 180 degrees);A_{ij} and B_{ij} are the terms of repulsion and attraction of dispersive interactions; q_i and q_j are the partial atomic charges of atom i and j; ε_1 is the dielectric constant of the surroundings.

In order to comprehensively study the dynamics of the protein in physiological conditions, as well as to confirm the proposed mechanism proposed for this enzyme, two separate systems were constructed, starting from two initial complex. As stated in the introduction, this reaction starts with a protonation of the substrate, in its zwitterionic form, by a His residue, resulting in an iminoacid. This is followed by a hydride transfer between the iminoacid and the FAD molecule, in its fully oxidized form. To mimic this first step, a complex was created, comprised of the substrate, in zwitterionic state (L-Phenylalanine), the protein, plus the corresponding FAD molecule, in its fully oxidized state (FADox).
The second part of the mechanism comprises the semi-oxidation step, that sees the oxidation of the FAD molecule, with the concomitant formation of the α-ketoacid, and the release of H_20_2 and NH_4^+. To study this part of the mechanism, a different system was created, comprising the substrate (in its iminoacid form), the protein, plus the FAD molecule, in its semi-reduced state (FADH).
To estimate the dynamics of such complexes, we started by creating libraries of parameters (defining bonds length, angles and dihedrals) for the different constituents. The parameter library is called topology of the system. Additionally, we defined the initial coordinates of systems, i.e., the spatial location occupied by each atom in the initial structure. Then, both structures were refined, by relaxing each of the complexes,

conducting them to an energy minimum, ultimately allowing the correction of the geometry of the system to its ideal state.

2.1 Setup of the Initial Structures

This step was first achieved with the employment of a script designed to clean *.pdb* files - *pdb4amber*. This script strips the file of irrelevant information (for the calculations), such as headers, connectivity, and structural information ("CONNECT", "HELIX", "SHEET" or "SSBOND" tabs, for example) as well as renaming the residues, followed by a respective renumbering, in order for the file to be understood by AMBER.
The usage of the script was followed by the confirmation of the protonation state of specific residues.

```
pdb4amber -i [input_pdb].pdb -o [output_pdb].pdb --reduce --dry
```

Figure 17 - pdb4amber script used for the initial preparation of .pdb files to be used in AMBER suite of programs

This is dependable on pk_a values and can influence the denomination of an aminoacid, depending the on the site where it is protonated. For example, His in physiological conditions is protonated, whereas its atom name could be defined as "HID", "HIE" or "HIP", depending on the location and the number of the nitrogen atoms protonated in the imidazole ring. This can either be performed by specialized software, such as H++ webserver [94], propKa [95] or manually.

2.1.1 FADox

AMBER suite has already implemented in the code the parameters for standard amino acids or nucleic acids but requires extra libraries for non-natural amino acids and organic molecules. There are, however, some databases with available parameters for extra amino acids, cofactors, ions. R.E.DD. B. is a database of RESP and ESP charges of high quality and reproducibility, and force field libraries for small molecules and molecular fragments [96]. Since the standard force fields of AMBER (ff14SB and GAFF/GAFF2 [97]) does not have parameters for FADH nor fully oxidized FAD, this database allowed the creation of parameters for the fully oxidized FAD (FADox for shortage) molecule, by assembly 4 molecular fragments constituting the molecule [98].

This step started with a manual deprotonation of the N5 of the FADH molecule present in the complex created in the previous step, saving the resulting *.pdb* file afterwards. *Pdb4amber* was then applied, resulting in the generation of 2 *pdb* files – one for the protein, and one for the FAD molecule, for which standard AMBER force fields does not have parameters for.

The script adopted from the R.E.DD.B database, makes use of the *tLeap* module, and it allowed the complete definition of all the atom types belonging to the parts composing FADox, followed by the assembly those parts. At this step, a new library, with all the FAD atom types, bonds, angles and dihedrals was saved (*.off* file). At the same time, we saved the new structure (*.mol2* file) for this assembled molecule.

To simulate the initial system, i.e, the system that corresponds to the first step of the chemical reaction, the substrate has to be in its zwitterionic state. However, AMBER suite only has libraries of parameters for amino acids in physiologic conditions, meaning that L-Phenylalanine parameters only exist in its positive form. This setback was resolved by the download of the parameters for this species (95 that were previously developed by the Bryce Group (at the University of Manchester) and are deposited on Amber Parameters Database [99].

```
loadoff FAD FAD.off
source leaprc.protein.ff14SB
source leaprc.water.tip3p
FRCMOD = loadamberparams script3.ff
loadoff zaa.off
cofactor = loadpdb FADox2.pdb
savemol2 cofactor cofactor.mol2 0
saveamberparm cofactor cofactor.prmtop cofactor.inpcrd

protein = loadPDB 5TS5_Model_protein_test.pdb
saveamberparm protein protein.prmtop protein.inpcrd
savepdb protein protein.pdb

complex = combine {protein cofactor}
charge complex
addions complex Na+ 9
solvateoct complex TIP3PBOX 15.0 iso
saveamberparm complex complex.prmtop complex.inpcrd
savepdb complex complex.pdb
```

Figure 18 - Script used for the assembly of the complex with the fully oxidized form of FAD, the zwitterionic state of L-Phe, and the protein.

The complex was then assembled by the *combine* feature of *tLeap* (i.e., combining three different molecules: protein, FAD and substrate), followed by the neutralization of its negative charge – that saw the addition of 9 counterions of Na^+ — the solvation of the complex by a TIP3BOX octahedral water model with 15 Å of radius, meant to mimic the interactions between the complex and the solvent. Finally AMBER topology (*.prmtop*) and coordinates (*.inpcrd*) files were generated though *saveamberparam* feature. (Figure 18)

2.1.2 FADH

The parameters for the semi-reduced FAD molecule were also downloaded from the AMBER parameter database [100], whereas the parameters for the iminoacid form of L-phenylalanine were generated through a charge (RESP) model calculation. For that matter, the 3D structure (in Tripos MOL2 format) of the intermediate ((imi)-L-Phe) was created using *GaussView* [101].

Then, the geometry of the structure was optimized using Gaussian software (93), version 09, under the B3LYP/ 6-31G* level of theory. The output of this operation was then used as the input for the next step that saw the generation of electrostatic potentials, under a Hartree-Fork/6-31G* level of theory.

The previous steps culminated in the calculation of the atomic charges by the restricted Electrostatic Potential (RESP) method and assigned to the respective atoms, using the antechamber package of the AMBER suite [102]. The *.prep* file outputted by this package was then served as input for the *parmchck2* package, also included in AMBER suite, a software designed to validate the accuracy of the parameters. After the generation of the force field modification (*.frcmod*) file by the *parmchk2*, two separate *tleap* scripts were created – the first (Figure 19) intended to create the parameters' library (*.lib* file) for the intermediate, the topology and coordinates files and the new *pdb* structure.

(1)

```
source leaprc.protein.ff14SB
source leaprc.gaff
source leaprc.water.tip3p

loadamberprep int.prep
loadamberparams int.frcmod

intermediate = loadmol2 intermediate.mol2
saveoff intermediate int.lib
savepdb intermediate int.pdb
saveamberparm intermediate intermediate.prmtop intermediate.inpcrd
quit
```

(2)

```
source leaprc.gaff2
source leaprc.protein.ff14SB
source leaprc.water.tip3p

loadoff FAD.lib
loadamberprep int.prep
loadamberparams FAD.frcmod
loadamberparams int.frcmod

FAD= loadpdb FAD.pdb
Prot = loadpdb 5TS5_Model_pdb4amber.pdb
Int= loadMol2 intermediate.mol2
Complex = combine {Prot FAD Int}

bond Complex.331.SG Complex.412.SG (i)
bond Complex.10.SG Complex.173.SG
solvateoct Complex TIP3PBOX 15
addIONS Complex Na+ 7
savepdb Complex Complex.pdb
saveamberparm Complex Complex.prmtop Complex.inprcd
quit
```

Figure 19 - Script used for the generation of libraries pf parameters, topologies, coordinates, and 3D structures. (1) Script used for the iminoacid. (2) Script used for the complex.

The second script (2) resulted in the generation of topology and coordinates files for this complex, by combining the parameters of the protein, FADH molecule and the intermediate. The four disulfide bridges previously present in the enzyme was explicitly kept by the *bond* feature (i). Once again, in order to mimic the influence of the solvent in the dynamics of the complex, an octahedral TIP3PBOX water model system was added, with 15 Å of radius. The complex had an overall charge of -9, which was neutralized by the addition of 9 counterions of Na^+.

2.2 Geometry Optimization

Before the production of molecular dynamics trajectories, aimed to study the behavior and motion of the proteins in physiological conditions, it is a standard practice to optimize the geometry of complexes. This step is needed to avoid less favorable contacts and clashes, and attempts to drop the potential energy to a local minimum [89] and relax the system, therefore reducing any unfavorable bonds and clashes that could have arisen from the parametrization process.

The energy minimization process relaxes the initial atomic positions, avoiding undesirable steric repulsions and clashes resulting from misdirected or close atoms. Jumping this step could affect the final confirmation and reactivity [89].

The following scripts were created with different atoms constrains, aimed to slowly correct clashes resulting from the automatic placement of water molecules, counterions and hydrogen atoms by *tLeap.* A first script was generated, with the intention of minimizing only hydrogen atoms (1) in the system. A second one was then employed, relaxing the water molecules (2), while in the third and last script, all the restrains were lifted, minimizing all the system (3).

(1)

```
Minimzing of the waters with 25 kcal/mol restraints
500 steps of steepest
&cntrl
imin=1,
ntx=1,
irest=0,
ntpr=50,
ntf=1,
ntb=1,
cut=9.0,
nsnb=10,
ntr=1,
maxcyc=2000,
ncyc=500,
ntmin=1,
restraintmask=':!:WAT',
restraint_wt=25.0,
&end
&ewald
ew_type = 0
&end
```

(2)

```
Minimzing the hydrogen atoms with 25 kcal/mol restraints
500 steps of steepest
&cntrl
imin=1,
ntx=1,
irest=0,
ntpr=50,
ntf=1,
ntb=1,
cut=9.0,
nsnb=10,
ntr=1,
maxcyc=2000,
ncyc=500,
ntmin=1,
restraintmask=':*&!(:WAT|@H=)',
restraint_wt=25.0,
&end
&ewald
ew_type = 0
&end
```

(3)

```
Minimzing the system
500 steps of steepest
&cntrl
imin=1,
ntx=1,
irest=0,
ntpr=500,
ntf=1,
ntb=1,
cut=9.0,
nsnb=10,
maxcyc=20000,
ncyc=500,
ntmin=1,
&end
```

Figure 20 - Scripts used for the energy minimization of the system. (1)- Water minimizations. (2) - Hydrogens minimizations. (3)- All system minimization.

All the minimizations (imin=1) saw a cutoff for intermolecular interactions of 10 Å. 500 cycles of optimization (ncyc=500) were employed, 250 of which used the Steepest Descent algorithm, while the other 250 cycles followed the Conjugate Gradient algorithm. Each minimization ended with the generation of a *.rst7* file (with atomic coordinates and velocities of the system), which was then used as an input for the subsequent steps.

The SANDER (Simulated Aneling with NMR-Derived Energy Restrains) module, from the AMBER suite was used to perform the calculations, having as reference the previous generated structure (4).

(4)

```
sander -O -i min_wat.in -p Complex.prmtop -c Complex.inprcd -r Complex_min_wat.rst7 -o Complex_min_wat.out -ref
Complex.inprcd

sander -O -i min_H.in -p Complex.prmtop -c Complex_min_wat.rst7 -r Complex_min_Hs.rst7 -o Complex_min_Hs.out -
ref Complex.inprcd

sander -O -i min_all.in -p Complex.prmtop -c Complex_min_Hs.rst7 -r Complex_min_All.rst7 -o Complex_min_All.out -ref
Complex_min_Hs.rst7
```

Figure 21 - Script instructing SANDER module to conduct all the minimizations.

3.Strutural Analysis of *Bothrops Atrox's* LAAO

3.1 Molecular Dynamics (MD)

- Background

A static structure can only be representative of a single moment in time, analogous to picture capture, which is only a frame in a complete movie. Biological systems cannot be described as single frames but need to be seem as a movie, where the full complex is dynamics. The full movie can give atomic details, as well as the inter-interactions, and conformational changes to better understand biological functions. Molecular dynamics allows the prediction of the behavior of all atoms in a system over time, giving information about the system's motion, as well as the outcome of induced perturbations in the system (e.g., mutations, protonation's, ligand binding, phosphorylation, etc.) [103]

To predict the dynamics of a system over time, MD calculations describes the position and velocity of every atom though the application of Newton's second law of motion - $\vec{F} = m\vec{a}$ [104].

For this calculation, an initial set of positions is deterministic, that could be either retrieved from an X-ray structure or modelled; and to each atom is given initial velocities, randomly attributed from a Maxwell-Boltzmann distribution over a selected temperature. Boundary conditions are also required, with the goal to mimic the interactions with the mean, as well as a description of how the molecules and atoms in the system interact with each other, namely though the potential energy given by a force field equation [105].

A well-known drawback of Molecular dynamics, however, is that it cannot be used solely for the study of catalytic mechanisms, as it cannot predict the formation or break of atomic bonds [103].

- Motion Studies

The MD trajectories were carried out using the AMBER18 software, within the pmemd.MPI module. The first step of this protocol was to refine the complexes, in order to remove undesirable contacts points, as stated earlier. To achieve this, the scripts referred to the geometry optimizations (section 2.2), discussed above, were employed, using the SANDER module.
The second stage of the molecular dynamics simulations saw the gradual heating of the system (1), raising the temperature gradually from 0K (tempi=0) to 300K (temp0=300), using the Langevin thermostat (ntt=3) over the course of 5 ns.

This was followed by 10ns of equilibration (2), in which the system was subjected to constant pressure of 1 bar (ntb=1), recording the resulting trajectories every 10000 steps (ntwx=10000). The frequency of collisions was established at 1 *p*s (gamma_ln=1.0) and 5000000 simulation steps (ntslim=5000000) with an integration time of 1 *f*s (dt= 0.001) considered. A cut-off of 10 Å for non-ligand interactions was set, stating that only interactions falling within that radius were considered. Not having this parameter set would off course increase the needed computational time, according to the number of atoms and/or interactions in the system.

The final step of this protocol, called production step (3), saw the application of an isothermal-isobaric ensemble, subjecting the system to constant temperature and pressure (ntb=2), thereby allowing a more accurate representation of the system density and consequently a more coherent simulations of biological systems. In this particular step, only the coordinates were read (ntx=1), and not the velocities, under 1000000 simulation steps (ntslim = 1000000), with 1*f*s (dt=0.001). The reliability of the results was assured by producing 5 replicas for each system. Several files were generated as output, corresponding to energy (*.mdout* and *mdinfo*), coordinates, and velocity (*.nc* and *.rst7*) information for every 5000 steps (ntpr=5000).

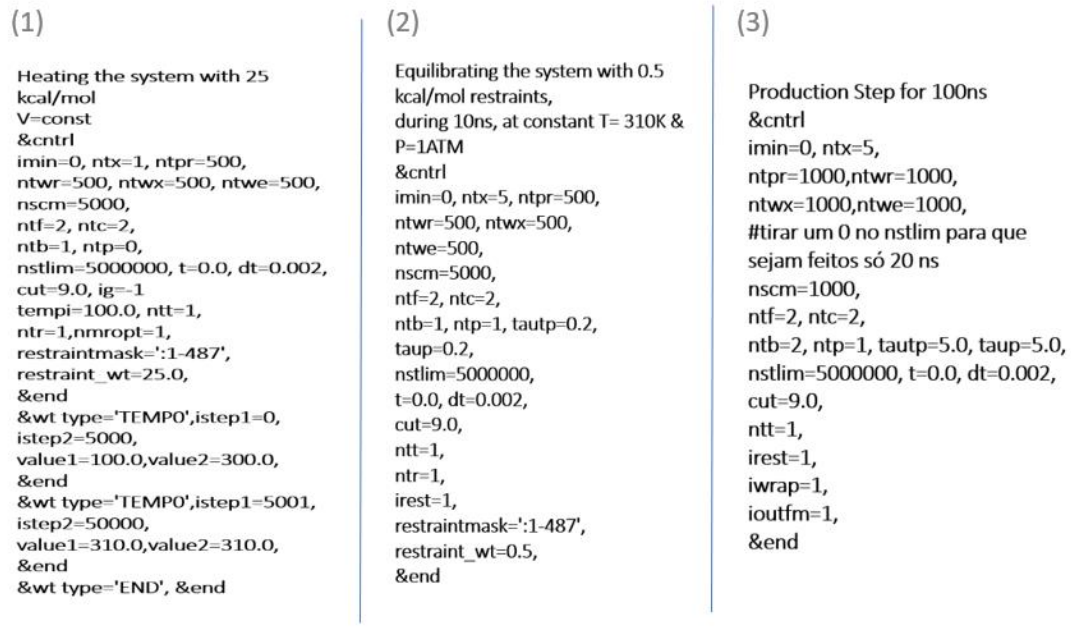


Figure 22 - Scripts used for the energy minimization of the system. (1) – Heating phase. (2) – Equilibration phase. (3) – Production phase.

3.2 MD Analysis

3.2.1 Root-Mean-Squared deviation Calculations

One of the primary parameters that can be used in order to infer the stability of a proteins throughout a MD simulation is the analysis of the RMSD. It measures the displacement of atomic positions, by comparing the distance between two sets of atomic coordinates, therefore allowing to analyze how much the geometry of a structure has fluctuated in a tridimensional space. Considering two structures (*v* and *w*), with corresponding two sets of atom coordinates, the RMSD can be calculated considering the expression:

$$\mathrm{RMSd}_{(\mathrm{v,w})} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(v_x^i - w_x^i\right)^2 + \left(v_y^i - w_y^i\right)^2 + \left(v_z^i - w_z^i\right)^2}$$

Equation 3 - Calculation of the Root Mean Squared deviation for a pair of atoms (v and w) in the positions x, y and z.

This metrics indicates the impact of the complex re-adjustments along the MD trajectories calculated and when the system reaches the most stable conformation. The

calculations allowed to inspect the stability of the complex, as well as its flexibility [106]. In general, a structure is be considered stable if RMSD values in relation to the average structure are below 2 Å, even though there are reports stating that these values can be significantly bigger for larger complexes. [106, 107].

In the context of the present work, we have evaluated the RMSD values comparing the system evolution with the first frame of the trajectory and with an average structure that resulted from a stable part of the trajectory (last 10ns-see below).

3.2.3 Root Mean Squared Fluctuation Calculation

- Average Structure

An average structure resulting from the most stable part of trajectories, and necessary for to estimate the more flexible regions of the complex (see below the RMSF section), was computed over the last 10 ns of the simulation and as stated before, was generated using CPPTRAJ and the average command. To carry on this calculation, we stripped the water molecules and sodium counterions from the initial systems. Two different scripts were drafted (1 & 2), for the two systems. Once generated, this structure was then minimized to relax the amino acid side chains.

(1)

```
trajin md10.nc
autoimage
center origin :*
strip :WAT,Na+
average Avg_Struct.pdb pdb
average Avg_Struct rst7

rms :1-484 out Protein.data
rms  :485 out Substrate_ZF.data
rms :486-489 out FADox.data

run
```

(2)

```
trajin md10.nc
autoimage
center origin :*
strip :WAT,Na+
average Avg_Struct.pdb pdb
average Avg_Struct rst7

rms :1-484 out Protein.data
rms  :486 out Substrate_Imi.data
Rms :485 out FADH.data

run
```

Figure 23 - Script for the retrieval of the average structure over the last 10ns of simulation, and RMSD values for the protein, the substrate, and the cofactor. (1) - Complex with the fully oxidized FAD. (2) - Complex with semi-reduced FAD molecule.

- Root Mean Squared Fluctuation Calculation

After the retrieval and the consequent minimization of the average structure, the positional fluctuation per residue in the complex was analyzed using the calculation of the Root Mean Squared Fluctuation (RMSF). This parameter measures the fluctuation of an atom or coordinate over time, in relation to a reference position. This measurement allowed for the inspection of the flexibility of every constituent of each of the complexes, and whether if there were any major changes resulting from the conformational changes.

The RMSF calculation can be made over the following equation:

$$\mathrm{RMSF}_{(\mathrm{i})} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\langle\left(r_i'(t) - r_i(t_{ref})^2\right)\rangle}$$

Equation 4 - Calculation of the Root Mean Squared Fluctuation for an atom (r), over a period of time (t) in relation to a reference (t_{ref}).

Where r' is the position of the atoms in the residue i after the superimposition on the average structure, and ri is the initial position of the residue. T is the time of the trajectory, whereas t_{ref} is the reference time and the angle brackets represent the squared distance of the atoms in the residue average. As stated, RMSF values were calculated for every part of each of the complexes – The protein, the substrate, and the FAD molecule – according to the following script:

```
trajin md1.nc
center origin :*byres
autoimage
strip :WAT
strip @Na+

parm Complex_m.prmtop
reference avg5TS5_Min.rst7 parm Complex_m.prmtop
center reference
atomicfluct out rmsf_Protein.data :1-484 byres
atomicfluct out rmsf_Fadox.data :486-489 byres
atomicfluct out rmsf_Substratelmi.data :485 byres
run
```

Figure 24 - Script for the retrieval of the RMSF values for the protein, the substrate, and the cofactor. Analysis conducted using CPPTRAJ software.

3.2.4 Distance and Angles Analysis

The distance between key residues during the MD were analyzed. The data here generated could determine a pattern of permanent and transient interactions in the system, allows to determine the stability of the active site and the substrate/substrate, and also to select the frames to use for the subsequent studies.

Here, we paid particular attention to the residues that helped to stabilize important molecules for the reaction, whether being the FAD molecule, or the substrate. For the FAD molecule, we have evaluated its interaction with the Arg90 (where a hydrogen bond was observed between the N3 of the isoalloxazine ring and the O of the Arg backbone). For the substrate, those residues were Arg90 (Putative hydrogen bonds between the terminal carboxyl group of the substrate and the NH1/NH2/NE of Arg90) and Gly464 (to inspect a possible hydrogen bond between the backbone of Gly464 and the N group of the substrate).

This analysis was also conducted in order to inspect the variation of the reactive distances between the catalytic His223 and the FAD molecule, specifically the distance between the N atom of the substrate and the N2 of the His223; and the distance between the Cα of the substrate and the N5 of the isoalloxazine ring, participating in the hydride transfer.

Lastly, and in order to select frames of the MD corresponding to initial structures of the complex to be used in the QM/MM studies, the angle between His223 (NE2) and the substrate (N atom and adjacent Hydrogen), and the angle between the substrate (Cα - Hα) and FAD (N5) were also analyzed. The following script depicts the described protocol.

```
trajin md1.nc
.
.
.
trajin md10.nc

center origin :*
autoimage
rms first 1-484@CA,C,N,O out rms_protein_bb.dat
distance d1 :223@NE2 :485@N1 out distance_his223_NE2_ZF_N_Calc1_FAD.data
distance d2 :223@ND1 :485@N1 out distance_his223_ND1_ZF_N_Calc1_FAD.data
distance d3 :372@OH :485@OXT out distance_tyr372_OH_ZF_O1_Calc1_FAD.data
distance d4 :372@OH :485@O out distance_tyr372_OH_ZF_O2_Calc1_FAD.data
distance d5 :223@CA :409@OE2 out distance_His223_CA_Glu209_OE2_Calc1_FAD.data
distance d5 :223@CA :409@OE1 out distance_His223_CA_Glu209_OE1_Calc1_FAD.data
distance d2 :223@NE2 :485@H1 out distance_his223_ZF_H1_Calc1_FAD.data
distance d3 :223@NE2 :485@H2 out distance_his223_ZF_H2_Calc1_FAD.data
distance d4 :223@NE2 :485@H3 out distance_his223_ZF_H3_Calc1_FAD.data
distance d5 :485@OXT :90@NE out distance_arg90_NE_ZF_O_Calc1_FAD.data
distance d6 :485@OXT :90@NH1 out distance_arg90_NH1_ZF_O_Calc1_FAD.data
distance d7 :485@OXT :90@NH2 out distance_arg90_NH2_ZF_O_Calc1_FAD.data
distance d8 :485@O :90@NE out distance_arg90_NE_ZF_OXT_Calc1_FAD.data
distance d9 :485@O :90@NH1 out distance_arg90_NH1_ZF_OXT_Calc1_FAD.data
distance d10 :485@O :90@NH2 out distance_arg90_NH2_ZF_OXT_Calc1_FAD.data

distance d11 :90@O :485@N2 out distance_arg90_O_FAD_N3_Calc1_FAD.data
distance d12 :464@O :485@N1 out distance_gly464_O_ZF_N_Calc1_FAD.data
distance d13 :485@C3 :485@N2 out distance_ZF_CA_FAD_N5_Calc1_FAD.data
distance d13 :485@C3 :485@N2 out distance_ZF_CA_FAD_N5_Calc1_FAD.data
distance d20 :485@O :322@NH1 out distance_ZF_O_Arg322_NH1_Calc1_FAD.data
distance d21 :485@O :322@NH2 out distance_ZF_O_Arg322_NH2_Calc1_FAD.data
distance d22 :485@O :322@NE out distance_ZF_O_Arg322_NE_Calc1_FAD.data
distance d23 :485@OXT :322@NH1 out distance_ZF_OXT_Arg322_NH1_Calc1_FAD.data
distance d24 :485@OXT :322@NH2 out distance_ZF_OXT_Arg322_NH2_Calc1_FAD.data
distance d25 :485@OXT :322@NE out distance_ZF_OXT_Arg322_NE_Calc1_FAD.data
distance d1 :223@ND1 :224@OD1 out distance_His223_ND1_Asp224_OD1_Calc1_FAD.data
distance d2 :223@ND1 :224@OD2 out distance_His223_ND1_Asp224_OD2_Calc1_FAD.data

angle angl1 :485@C3 :485@HA :485@N5 out angle_ZF_HA_FAD_N5_Calc1_FAD.data
angle angl2 :485@C3 :485@N :223@NE2 out angle_ZF_CA_his_223_Calc1_FAD.data
```

Figure 25 - Script used in the calculation of the distances and angles throughout the simulation. Analysis conducted using the CPPTRAJ software.

3.2.4 Dihedral Angle Analysis

One of proposed structural variations that occurs following the coupling of the substrate is the change in the conformation of His 223 [35]. This change is thought to block the substrate in the active site, with the aid of Arg 322.

To infer the validity of this affirmation, the dihedral angles in the His223 residue were analyzed, corresponding to the orientation adopted by the imidazole ring and therefore its adopted conformation. Additionally, to infer the structural stability of the residues in the active site, the dihedral angles of Arg90 and *Tyr*372 This protocol was achieved by the following script:

```
trajin md1.nc
  .
  .
  .
trajin md10.nc

center origin :*
autoimage
rms first 1-484@CA,C,N,O out rms_protein_bb.dat

dihedral dih1 :322@CG :322@CD :322@NE :322@CZ out dihedral_Arg322_Calc1_4_FAD.data
dihedral dih2 :322@CB :322@CG :322@CD :322@NE out dihedral_Arg322_Calc1_3_FAD.data
dihedral dih3 :322@CA :322@CB :322@CG :322@CD out dihedral_Arg322_Calc1_2_FAD.data
dihedral dih4 :322@C :322@CA :322@CB :322@CG out dihedral_Arg322_Calc1_1_FAD.data
 dihedral dih5 :90@CG :90@CD :90@NE :90@CZ out dihedral_Arg90_Calc1_4_FAD.data
dihedral dih6 :90@CB :90@CG :90@CD :90@NE out dihedral_Arg90_Calc1_3_FAD.data
dihedral dih7 :90@CA :90@CB :90@CG :90@CD out dihedral_Arg90_Calc1_2_FAD.data
dihedral dih8 :90@C :90@CA :90@CB :90@CG out dihedral_Arg90_Calc1_1_FAD.data
 dihedral dih9 :372@N :372@CA :372@CB :372@CG out dihedral_Ty372_N_CA_Calc1_FAD.data
dihedral dih10 :372@CA :372@CB :372@CG :372@CD2 out dihedral_Tyr372CA_CB_Calc1_FAD.data
```

Figure 26 - Script for the retrieval of the dihedral angles throughout the simulation. Analysis conducted using CPPTRAJ software.

Finally, the presence of water molecules around the substrate, which could aid the hydride transfer between the iminoacid and the N5 of the oxidized FAD molecule was also inspected, using the *VolMap* plugin of VMD. This software ran over aligned trajectory files and computed occupancy maps, that are no more than 3D grids, were each individual space of the grid is stepped to be 0 or 1, whether it finds one or more atoms in it, or not. The occupancy map will therefore correspond to a fractional occupancy of every grid point in the trajectory.

4. Study of the catalytic mechanism

4.1 Quantum Mechanics

Contrary to Molecular Mechanics, quantum mechanics treats electrons explicitly. This allows more comprehensive and accurate study of phenomena happening in lapses of

times, such as chemical reactions, where there are formation and cleavage of bonds, electrons transfers, etc.

The characterization of all the physical properties of a system is made possible through the fundamental postulate of quantum mechanics [46, 104], which describes microscopic systems as wavefunctions.

However, and despite how complete and informative this type of methods can be, they are very complex and can only be used for systems up to hundreds, making their usage limited to small biological complexes, much smaller of those involving an enzymatic system.

4.2 Schrödinger Equation

One of the postulates of quantum mechanics was presented by Erwin Schrödinger in 1925. The equation was able to describe atomic, subatomic, and macroscopic systems, by the prediction of both nuclei and electronic behavior. It enabled the calculation of the system's energy (E).

However, and despite all the information that it could provide, and the simplifications assumed, the equation is still quite complex, as it can only be solved analytically for one particle systems, requiring further simplifications to be used for multi-body systems [46, 108]. The Born-Oppenheimer is one of the available approximations. Essentially, this simplification enables the breaking of the Hamiltonian into two terms – the nuclei and the electronic term. This is assumed, as it is considered that the nucleus has much greater mass than the electrons, resulting that the electrons adjust instantly to nuclear movements.

The nuclei term is handled though classical mechanics equations, while the electronic term is treated over quantum mechanics. As the position of the nuclei is considered fixed, for every arrangement, the Hamiltonian is solved considering only the electrostatic Hamiltonian. However, when there's is a change in nuclear positions, the nuclear Hamiltonian term needs to be added to the final term to account it to the system total energy, depicted in the Schrodinger equation [104].

The Born-Oppenheimer approximation is still limited, however, as it can only be solved for species comprising of two nuclei and one electron, such as H_2^+ [104]. In this sense, alternative approaches to predict nuclear and electronic behavior, ensuing its

applicability for bigger systems, can be divided in ab initio and semi-empirical approaches [46, 104, 108, 109].

4.3 *Ab initio* approaches

4.3.1 Hartree-Fock

Hartree-Fock (HF) is the most common *ab initio* method, and also the simplest. Is allows to obtain the energy and the orbital configuration for polyelectronic system [108]. It gives the probability of finding an electron around an atom, as dependent on its distance to the nucleus, rather than its distance to other electrons [46, 108] .Therefore, HF discards the effects of electronic correlation, i.e., the stabilizing effect resulting from alternative electronic distributions that decrease electronic repulsion. This allows de decomposition of the initial Hamiltonian in many one-electron equations, where an orbital and an orbital energy is given for each one-electron equation [108, 110]. However, despite this assumption, the average effect of electronic repulsion is still accounted for, though the integration of a repulsion term [108]. Accounting for the electronic repulsion only as an average often leads to huge errors in energy calculations, especially when applied to reactive chemical events, where electrostatic correlation plays a big role. This translates in energy values often greater than the exact energy.

4.3.2 Density Functional Theory (DFT)

One method, capable of achieving good levels of accuracy and applicable for systems up to hundreds of atoms, while maintaining manageable levels of computational power, is the density functional theory (DFT) [111]. This reduction in the computational burden is achieved by the replacement of the concept of individual electrons with electronic density (ρ). It states that the electrons' density distribution is the only requirement to calculate the ground state energy of a molecule.

Moreover, the reduction of computational power when in comparison with wavefunction methods, such as HF, DFT methods is due to the sole usage of 3 dimensions – x, y and z - being independent on the number of electrons of the system, unlike the wavefunction methods that are dependent on 3N dimensions, where N is the number of electrons. The energy of a system, according to DFT can be depicted as following:

$$E_{DFT} = E_k[\rho] + E_{ne}[\rho] + E_{ee}[\rho]$$

Equation 5 - Energy calculation through the density functional theory.

where E_k [ρ], describes the kinetic energy, E_{ne} [ρ] accounts for the interactions between the electrons and the nuclei, and E_{ee} [ρ], represents the electrostatic interactions. The terms for kinetic energy and for the interactions between electrons cannot be explicitly solved, however. Kohn-Sham orbitals is a solution to this problem, as it separates the kinetic energy term into kinetic energy for non-interaction electrons plus residual kinetic energy. The energy term for the interaction between electrons also suffers a division, originating two terms - a term for Coulomb interactions and an exchange term - k[ρ]. The sum of this newly created term is also known as exchange correlation functional, which can only be solved by attribution an approximated functional. Hence, the DFT functional is given by:

$$E_{DFT} = E_{k_{ni}}[\rho] + E_{ne}[\rho] + J[\rho] + E_{XC}[\rho]$$

Equation 6 - Energy calculation of a system through the density function theory and by applying an exchange correlation functional.

The electronic density of the system is assigned to the previous equation by allocating a guess density from a set of initial electronic orbitals. Then, a new set of orbitals is applied, and a new electrons' density is generated. Sets of orbitals are generated until the calculated ground-state energy equals the one from the previous guess or bellow a given threshold. When this happens, it is deemed to be self-consistent. The ground state energy of a system is therefore given by a self-consistent field (SCF) calculation.

However, DFT still has some drawbacks, as medium and long-range VDW interactions are not accounted for energy calculations, and its accuracy is very dependent on the selected exchange correlation functional [112].

▫ Exchange Correlation Functionals

Exchange correlation functionals tend to be simple and to provide promising results, representing an appealing tool in computational chemistry. The simplest one is the local density approximation (LDA) in which the electron density is considered homogeneous

across all space. An approximation that considers the heterogeneity of the electrons density is the generalized gradient approximation (GGA). It introduces the dependency on the gradient of the electrons density, achieving more reliable results for biomolecules. In a hybrid GGA – which includes B3LYP - a percentage of Hartree-Fock exact exchange is blended with exchange and correlation potentials.

These hybrid potentials still present some limitations like the ineffective treatment of charge transfer processes, overestimating their interaction, and over polarization of molecular orbitals and they are also not able to account for dispersion interactions. Additionally, they also give wrongful, energies to describe the dissociation of simple two-center systems [113].

▫ B3LYP

B3LYP, standing for Becke-3-parameter Lee-Yang-Parr is the most used exchange correlation potential, given its accuracy for a wide variety of compounds, like organic molecules [108].

▫ Basis Sets

A basis set can be defined as a set of functions that can be combined to describe molecular orbitals. It has great impact on the accuracy of the results. They can be created by merging linear combinations of basic functions and angular functions. Already existent basis set can be employed, such as Slater type orbitals (STO) and gaussian type orbitals (GTO). STO are only applicable for mono or diatomic systems, and they are too computationally demanding, making their usage advisable only when higher levels of accuracy are required. As some groups prefer to use STO as basis sets, GTO are the preferred choice for QM calculations, as they represent a simpler ensemble of mathematical functions, while being applicable to systems with more than two atoms. GTO presents a diminished computational effort, as their accuracy and description of valence atom is reduced [114]. This issue can be resolved by the application of, more complex basis sets, like the addition of Gaussian mathematical functions, that in term enables a better description of valence electrons' orbitals, while the addition of polarization and diffuse functions can improve the flexibility of the orbitals. A widely used GTO basis set, and the one used for the conduction of QM/MM studies for the present

work, that has a good compromise between computational cost and accuracy is the 6-31G(d) basis set.

4.4. Quantum Mechanics/Molecular Mechanics (QM/MM) studies

The knowledge of catalytic reaction mechanism represents an important point for the development of inhibitors, based on the TS theory. As referenced before, Molecular Mechanics alone is not able to describe some phenomena to occur in a biological system, like the breaking and formation of bonds or electronic arrangements.

In this sense, in 1976, Whersell and Levitt [115] developed the hybrid method QM/MM, a tool that combines QM and MM methods, allowing the division of a biological system in two (or more) parts, each treated with different levels of theory, thus enabling the study of systems up to thousands of atoms, while ensuring lower levels of computational effort.

The active region of an enzymatic system, involving electronic rearrangements, the breaking and forming of chemical bonds, which represent events that happen in infinitely short periods of time is treated under a quantum mechanics formalism, whereas the outer part of the system(remaining protein and as well as the solvent, which may contain up to thousands of atoms), not actively involved in the catalytic mechanism and therefore not requiring to be described with greater levels of theory is handled under a classical mechanics formalism. This part is commonly treated with classical force fields, such as AMBER or GROMACS.

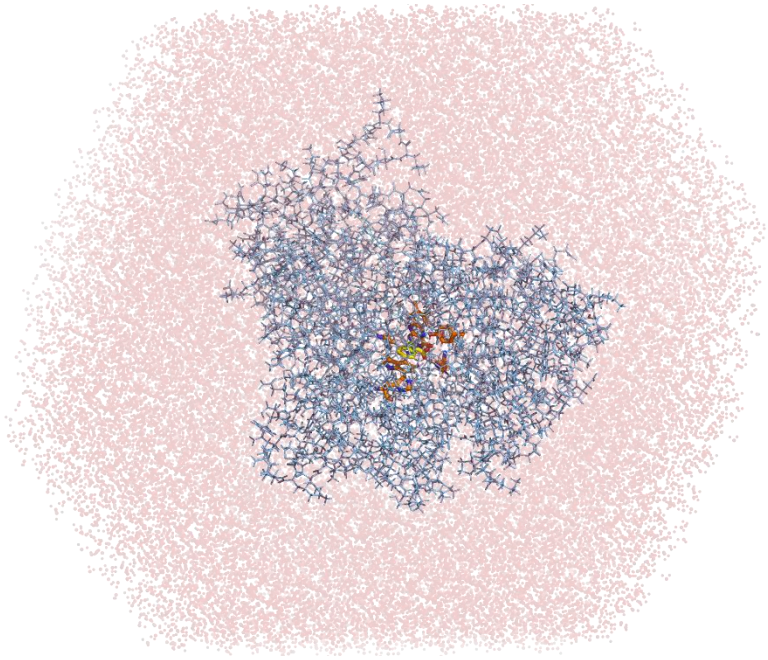


Figure 27 - Protein system treated over a QM/MM scheme, where the part of the system playing an active part on the reaction mechanism (in orange sticks) is treated under a Quantum Mechanics formalism, with the rest of the system (in blue lines) is treated with a Molecular Mechanics formalism.

▫ QM/MM schemes

There are two strategies in QM/MM methods, additive and substrative schemes. In addictive schemes, the energy of the system is described as the sum of the energy of the QM layer (inner layer) and the MM layer (outer layer), plus a term accounting for the coupling between these two layers.

$$E_{QM/MM\ fullsystem} = E_{QM\ innerlayer} + E_{MM\ outerlayer} + E_{QM\ coupling}$$

Equation 7 - Calculation of the energy of a system over an additive scheme.

In a subtractive scheme, the system's energy is obtained by firstly determining the energy of all system, through a lower theoretical level (MM), followed by the calculation at bot QM and MM level of the inner layer. The MM contribution of the inner layer is subtracted from the MM contribution of the whole system, and the QM energy added, resulting in the final total energy for the system (Equation 8)

$$E_{QM/MM\ fullsystem} = E_{QM\ innerlayer} + E_{MM\ fullsystem} - E_{MM\ innerlayer}$$

Equation 8- Calculation of the energy of a system over a substrative scheme.

Subtractive prove to be better in relation to additive schemes, given that they do not need a parametrized expression to calculate the energy for coupling region's term.

ONIOM (69) (Our Own N-layer Integrated molecular Orbital molecular Mechanics) developed by Morokuma and co-workers in 1986, is a subtractive QM/MM method that attempted to solve this problem by incorporating an electrostatic embedding scheme [116].

In the context of this work, the ONIOM framework was employed. Among other things, this method enables geometry optimizations, energetic and frequencies calculations.

▫ Boundary Schemes

The electrostatic interactions between the QM and the MM regions can be delt by a mechanical embedding approach (ME), or tough electrostatic embedding (EE) approach. In ME, the electrostatic interaction in those regions are only handled at a MM level. Contrary to this, in EE there is an incorporation of electrostatic interactions of the MM region into the QM equation [116].

▫ Mechanical Embedding

ME embedding scheme was first implemented in the ONIOM methods [117]. This is the simplest and the least demanding method to treat the QM/MM electrostatic interactions.

Its simplicity is attributed due to neglecting of the electrostatic effects by the surrounding environment on the QM calculations, which mean the surroundings (treated with MM methods) do not polarize the QM region. In this scheme, the QM calculation only considers the structural constrains imposed by the MM layer. For MM calculations, a fix set of point charges is attributed to every atom in the QM system. Therefore, the accuracy of these interactions, depend strongly and solely on atomic charges defined by the user.

The boundary interactions are delt by molecular mechanics, making that one of its main limitations is the need for constant reparameterization of the atomic point charges of the reactive region, due to the fact that along the reaction, the charge distribution in the reactive system suffers alterations [116].

To include polarizations of the QM wave function by the environment (improving QMMM electrostatic interactions), a term that describes the atomic charges of the MM environment system can be added.

- Electrostatic Embedding

The Electrostatic embedding scheme includes a polarization of the inner layer, by adding a set of point charges of the MM region. This results in the addition of sets of parameterized point charges into the Hamiltonian equation. This provides a better accuracy, as its mimics more faithfully what happens in a real biological system.

The main advantages in relation to the ME scheme are not requiring a set of derived charges for the high layer, plus not needing a constant charge update, given that the point charges can be derived from the wavefunction.

The high accuracy related to this embedding translates in a significant computational cost. Moreover, the polarization of the MM region by the QM system is not accounted, which can still affect the results, despite know being very relevant [116].

- The link atom approach

When the QM and the MM layers are connected, special considerations are going to be required. The breaking of covalent bonds generates unpaired electrons at the boundary region, that can greatly influence the electronic structure of the QM layer. A strategy used to address these issues is by the application of a link atom approach, so that atomic open valences that are generated by the bond cleavage between the layers can be completed.

In this approach, an atom, usually a hydrogen is added at an appropriate position along the boundary between both layers. Using hydrogen is the simplest and the most used boundary approximation even though other monovalent atom can be used as link atom These link atoms are only going to be accounted for by the QM layer, so that they are only required for the substrative scheme calculation [116].

4.2 Initial Structures Preparation

After the Molecular Dynamics protocol, the resulting frames were analyzed and the one with the most favorable distances and angles between key amino acids for the reaction to take place were selected. With respect to the complex involving the fully oxidized form of the FAD molecule and the zwitterionic form of the substrate, an initial snapshot was selected so that the distance between the NE2 of the His223 and the N terminal of the substrate was around 2.5 Å, while the distance between the C_α of the substrate and the N5 of the FAD molecule was around 4 Å, with additional detail on the angle between these residues to be around 90 degrees, so that the H_α could be point directly to the N5 atom.

Regarding the system involving the semi-reduced FAD molecule and the iminoacid, only the distances between the previously referred residues were taken into attention for the selection of an initial frame.

Lastly, and in order to minimize the computational effort required for the calculations, the size of the solvent shell was reduced, by defining a cap of water molecules distancing 5 Å from the protein's edge and the two frames were saved in *.xbgf* format.

4.3 Calculations Setup

The calculation scripts were prepared using *molUP* extension, under the VMD software. This plugin offers a personalized graphical user interface (GUI) to the Gaussian software. It was developed in order to aid the user to create Gaussian input files, and to analyze the resulting output files. The creation of the input script is facilitated by the existence of several options regarding the initial structure (definition of layers, frozen atoms, atomic charges) as well as predefined keywords offered by the plugin, depending on the job at hand.

4.4 Geometry Optimization

The first step of this protocol comprised the geometry optimization, under a mechanical embedding. This started with the definition of the layers – the low layer (LL) and the high layer (HL). The HL involved the residues (or parts of them) directly participating in the reaction, such as the isoalloxazine ring of the FAD molecule, the imidazole ring of His223 or the substrate. Other residues (completed or not) included in the layer were Phe 227,

Tyr 372, and a water molecule (residue 501). Overall, this layer consisted in 115 atoms, with a net charge of +1.

The QM layer was treated under a higher theoretical level, the DFT level with a hybrid functional – B3LYP, in a combination with the 6-31G(d) basis set. There remaining part of the system was described with MM methods, whereas a lower level of theory was employed, the AMBER ff14SB force filed.

ONIOM automatically places hydrogen atoms in order to complete the open valences, resulting from the breaking of bounds made at the definition of the layers. In this sense, atoms from the QM region covalently linked to the MM region were capped with hydrogen atom to saturate their valences.

Apart from the initial optimization of the structures, made with both ME followed by the employment of EE, all the other optimizations – for all the intermediates of the reaction – where optimized under an electrostatic formalism (EE), made with the intent to account the polarization of the QM subsystem by the MM layer. This approach is commonly used in an effort to reduce to computational effort during a QM/MM optimization.

4.5 Determination of Stationary Points

The energetic profile of a catalytic mechanism is enabled by the performance of several linear transit scans. They allow single point energy calculations over an incremented set of coordinates. For example, to study the initial protonation of the substrate, corresponding to the first step of the reaction mechanism, the distance between the NE2 of His223 and the H2 atom of the substrate was incrementally shortened, until a reactive distance was reached, and a proton transfer could be inferred. The putative transition structures were the ones corresponding to the maximum value found in the Potential Energy Surface (PES) calculated scheme of the reaction coordinates. The structures were then fully optimized (under an EE formalism), and their vibrational frequencies calculated, to confirm that the obtained structure was indeed an intermediary.

Once a transition state was found, intrinsic reaction coordinate calculation (IRC) was made to confirm the correct reactants and products of the reaction. The IRC method is a predictor-corrector reaction path integration method combined with ONIOM method to obtain steepest-descent reaction path for large systems [118] In this case, the core system walks along the IRC path and the remaining system is minimized after each IRC step [119].

This step was repeated for every proposed step of the reaction, until and energetic profile was set, with every intermediary of the reaction, as well as the final product.

5. Virtual Screening (VS)

VS is a technique used for the identification of HITS (i.e., molecule that is predicted to bind to the receptor with good affinity) from large chemical libraries through computational means, either though structural knowledge about the protein target (structure-based VS) or though known bioactive ligands (ligand-based VS).

The ligand-based approaches utilize structural data from a set of known experimental active molecules (actives) in order to identify novel drug candidates for a potential experimental evaluation. The structural data can derive form methods like similarity and substructure searching, QSAR, pharmacophore-based search, and 3D shape matching [44]

Alternatively, structure-based VS mainly employs a docking approach, often exhaustive, where the candidate molecules are iteratively docked against a biological target and rank them based on their predicted binding affinity (though the docking score).

Overall, the VS approach is highly dependent on the quantity and quality of available data and the predictability of the underlying algorithm. As a consequence, there is no universal guideline or workflow for the VS approaches, and the researcher has to apply his computational knowledge and experience to find the active drug candidate from the sea of drug databases and chemical libraries applying the best possible source of tools as per his requirements. However, despite being the method of choice for drug discovery in the pharma industry, various drawbacks arise from it, namely the high cost, the time-demanding character of the process as well as the uncertainty of the mechanism of action of the biological target.

5.1 Structure-Based Virtual Screening (SBVS)

In SBVS, large libraries of drug-like compounds that are commercially available are computationally screened against targets of known structure, by the employment of a molecular docking protocol, where the discrepancy often lies in the considered threshold for which the computed pose is considered for the results [44].

The obtained docking poses are then ranked and then post-processed by examining the calculated binding scores, validity of the generated pose, undesirable chemical moieties, metabolic liabilities, desired physicochemical properties. After this post-processing a number of compounds are then selected and experimentally tested.

5.1.1 Repurposing of small-molecule therapeutics (SMTs)

The specific target of a venom component is nearly impossible because the several non-enzymatic, low-molecular weight toxins that lack binding cavity. However, inhibiting some key venom toxins might succeed in delaying the venomous toxicity, as some toxins "trigger" major toxic events, while others reinforce them. Inhibiting the triggering toxins is therefore a very valid strategy against snake bite envenoming.

Since snake venom toxins evolved from gene duplication in the venom gland, this implies that most of venom enzymes and proteins have close analogs in humans that may have been the target of drug discovery studies. This statement constitutes the basic outline of drug repurposing studies.

- Chemical Database

With this idea in mind, and in the scope of this work, a drug repurposing protocol was employed in order to search for SMTs that could potentially be used to inhibit the effects of LAAO, by blocking its active site, therefore preventing the beginning of its catalytic cycle. The SMTs were searched and promptly put together on a database of chemical compounds, made of 9400 molecules that have successfully passed the stage 2 of drug discovery studies, entering the clinical stage. This information represents the main advantage of drug repurposing protocols, since it makes them less expensive to conduct, while also ensuring a more secure and safe approach [120] .

- Virtual Screening

A molecular docking protocol was them conducted though the database. This was achieved by, first dividing the database in 17 smaller input files that were the given to GOLD software, followed by a molecular docking run, manually setting a reference site, and using the ASP scoring function to rank the poses. 10 solutions were asked for each

ligand, allowing diverse solutions to be generated, consequently analyzing their interactions within the active site.

▫ Generation of Representative Clusters

Upon the generation of the poses by the virtual screening protocol, the top 250 achieved results were selected, according to the obtained ASP scoring function. This was followed by the clustering of the results, culminating in a final selection of 12 compounds to be tested *in vivo* against this enzyme. The generation of the clusters was made in two distinctive approaches – by hierarchical and non-hierarchical clustering. The hierarchical clustering method was achieved through the application of Ward's algorithm, whereas the non-hierarchical methods comprehended the Jarvis-Patrick algorithm, and the Diversity Selection tool.

The hierarchical clustering works by joining two successive clusters, merging the ones with the least distance between them. This distance is based on molecular fingerprints or other descriptors. The fast computation of this clustering method is ensured though the application of Murtagh's reciprocal nearest neighbor (RNN) algorithm.

Specifically, in the case of Ward´s clustering, the criteria to join two clustering is the Error Sum of Squares, more precisely, the objective is to minimize the sum of squares upon the merging (Ward's minimum variance method [121]), i.e., the square of the intra-clustering distance.

Similar to hierarchical clustering, non-hierarchical clustering creates and merges clusters in order to maximize or minimize some evaluation criteria. The main difference between them is the initial ordering, or lack of it (in the case of non-hierarchical) of the clusters.

Jarvis-Patrick is based on variable-length lists of nearest neighbors, where for each structure, it collects the set of nearest neighbors that has a dissimilarity (distance) less than a T threshold value. A cluster is then created if there are overlapping of neighbors and if R – number of nearest neighbors in common – is above a minimum [122].

The other non-hierarchical clustering method applied in this work – the Diversity Set Selection – represents an important point for pre-clinical drug discovery research and in design of diverse chemical libraries for combinatorial chemistry and compound selection for High Throughput Screening (HTS) [123].

This technique attempts to generate the most diverse solutions from a chemical dataset according to their molecular descriptors and the dissimilarity between compounds, where firstly a molecule is used as a query, and then the most dissimilar all is picked. This process repeats itself until the database is exhausted or the sorting aborted by the user. [123].

All these functions and tools were conducted using *Chemaxon* [124]– A software with many chemical tools aimed at treating and/or analyzing chemical compounds. The most suitable technique would be the one that generated the greatest number of clusters, while assuring chemical diversity between them.

Results and Discussion

- Creating the 3D-LAAO model

The correct and precise form of the tertiary structure of a protein is crucial for the success of subsequent steps of this study. Therefore, to complete the 3D structure of *Bothrops atrox's* LAAO (PBD:5TS5 [1]) we applied a Homology Modeling protocol, using the MODELLER software. The selected templates were the structures of our protein of interest (PDB: 5TS5) and the structure of L-amino acid oxidase from *Calloselasma rhodostoma* in complex with L-Phe (PDB: 2IID [125]). As similarities in the primary sequence of a set of proteins implies a similar folding, the sequence alignment becomes an important tool in the construction of a tridimensional structure. The sequence similarity between the target sequence and the *Calloselasma rhodostoma* template was analyzed through a BLASTp algorithm, by accessing the NCBI software suite [126]. The target and the analyzed template share very high sequence similarity — 86.6% — thus ensuring the usage of this template.

Another parameter that influences greatly the success of Homology Modeling is the resolution of the used template structures. An X-ray structure solved at high resolution displays a precise position of the atoms in the protein. An X-ray structure is considered to have good resolution when it is resolved with resolutions 2.5 Å, with lower values of atomic resolution implying greater structural detail.

Effectively, the structures used as templates (PDBs: 2IID; 5TS5) display atomic resolution values of 1.8 Å and 2.3 Å, respectively, denoting good starting points for the construction of a model.

To complete the structure of the PDB: 5TS5, we have produced 6 models. Subsequently, we evaluated the structural similarities between the produced models and the target protein using VMD's *RMSD calculator tool*. Results showed that, overall, the models mimicked quite well the folding of the initial target structure, displaying RMSD values between 0.862 and 0.885. RMSD values bellow 1 Å, are indeed within the experimental error associated with the crystallographic technique. Finally, the superimposition between the initial target protein (5TS5) and the selected model showed an RMSD value of 0.317 Å.



STRUCT	RMSD	DOPE_SCORE
Target	0.000	0.000e+00
Model 1	0.883	-5.854e+04
Model 2	0.884	-5.860e+04
Model 3	0.862	-5.848e+04
Model 4	0.875	-5.862e+04
Model 5	0.880	-5.861e+04
Model 6	0.874	-5.869e+04

Figure 28 - (Left) - Superimposition of the initial target system (in green) and the selected generated model (in gray). (Right) - Evaluation metrics of the 6 generated models with the chosen one highlighted in blue.

- How does the substrate “sits” in the catalytic cavity?

Once the 3D-model of the enzyme has been completed. We then asked the question — How does the substrate interacts with the active site of the enzyme?

The answer to this question was given by Molecular Docking. This technique was used to infer the best orientation of the substrate for the catalysis to occur.

The used substrate — L-Phe — is an L-amino acid that is a known substrate for the LAAO from *Bothrops atrox* [35]. We have then proceeded with a visual inspection of the top 30 of the 100 generated solutions.

The selected pose corresponds to the orientation that we found more suitable to the catalysis occur and, at the same time, formed key interactions with the residues Gly 464, Tyr 372, and Arg 90, as hinted in the study conducted by Anwar Ullah.

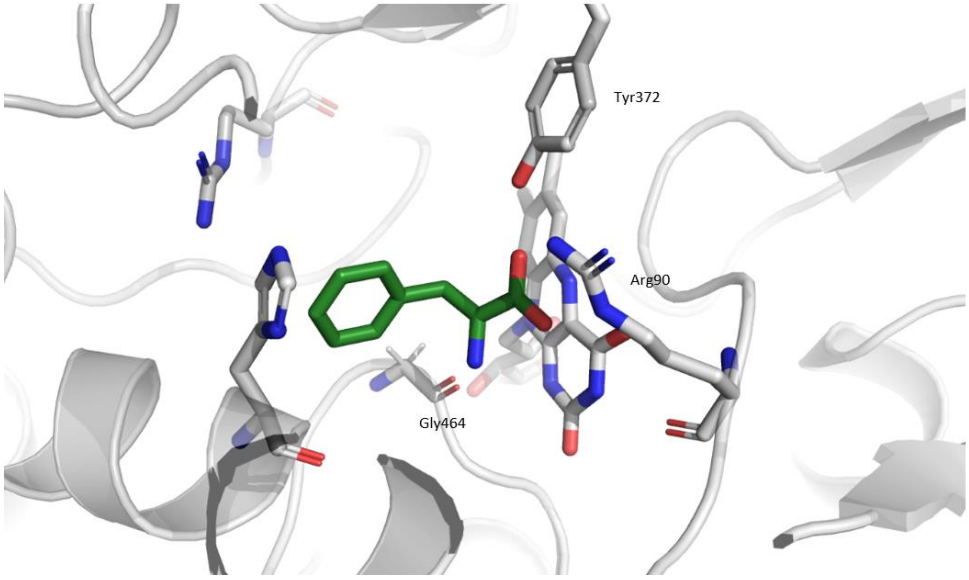


Figure 29 - Representation of the protein active site with the docked substrate (in green). Important residues for the stabilization of L-Phe a labelled in black, as noted by Anwar Ullah.

- The System Parameterization and Geometry Optimization

After the assembly of the initial complex of *Bothrops atrox* LAAO and the L-Phe, we proceeded with the setup of two separate systems that mimic the first two half-reactions of LAAO mechanism, at physiological conditions:

(a) The complex LAAO-L-PHE in presence of oxidated FAD.

The substrate in the zwitterionic state ((zwi)-L-Phe), plus the FAD molecule, in its fully oxidized state was parametrized and subsequently optimized, over energy minimizations.

(b) The complex LAAO-L-PHE in presence of the FADH.

The substrate acquires its iminoacidic form ((imi)- L-Phe) and the FAD molecule take its semi-reduced stage, as FADH.

The ionization state of the charged amino acids was calculated using the propka3.1 at physiological pH. Finally, the full system parametrization was conducted using the *tleap* module of AMBER 18.

- The Complex's Motions

Molecular dynamics simulations were carried out in three steps. In summary, the first one, comprised the minimization of the complex (at 0K) in order to avoid bad contacts and clashes. Then, we proceeded with the equilibration of the complex (10ns). At this point, we started by slowly rising the system temperature, until the system achieved 310K (the optimal temperature for the snake metabolism). Simultaneously, we equilibrated the system volume. Finally, we started the trajectory production and at this third step we changed from the NVT grand canonical to the NPT ensemble. The last step was carried over the course of 100ns, and 5 replicates were produced for each complex.

The goal of these calculations was fourfold:

(1) Determine the complex stability.

(2) Understand the complex dynamics and motions.

(3) Determine if any specific residue of the catalytic site (or even the protein) undergoes any conformational changes when the FAD is semi-reduced to FADH.

(4) Selection of different frames over the MD trajectories that could be used as input for the QM/MM calculations.

The system's stability was analyzed by the means of RMSD values over the course of the dynamics. This allowed for the inspection of positional deviations over a reference position. Additionally, the analysis of RMSF values allowed the observation of flexible regions in the protein and the possibility of any conformational changes. Distance analysis of key residues, over the course of the dynamics, helped to understand key interactions within the complex, namely in the active site and the residues that participate, directly or indirectly, in the catalysis.

These structural analyses were conducted using *cpptraj* tool of AMBER18, and the data was represented using the Python's *matplotlib* package.

- Stability of the Complexes: RMSD Analysis

For the oxidized systems, the RMSD values for the protein denotes good stability through the trajectories analyzed, with values consistently lower than 2 Å. This stability is also present in the fully oxidized cofactor — the FAD molecule — with average RMSD values of 0.5 Å. Finally, (zwi)-L-Phe experiences small fluctuations in RMSD values. It showed

good stability overall in the simulation, with RMSD values lower than 1.5 Å before *t=40ns,* and around 1.5 Å after *t=40ns*.

Regarding the FADH-complex system, the protein proves to be stable over the course of the dynamics, exhibiting RMSD values ranging from 1.0 to 2.5 Å. For most replicates, there is a leap in RMSD values around t=30ns, from 1.5 to 2.5 Å. In line with FAD in the oxidized system, the FADH molecule experiences very good stability, denoted by RMSD values consistently lower than 1 Å. In these systems, the (imi)-L-Phe experiences very good stability, with average RMSD's values of 0.5 Å, up until *t=30ns*. After, there is a leap in the average value, reaching 2 Å, for most replicates. The only exception is replicate 1, where RMSD values are, for the most part of the simulation, lower than 1 Å.

In both systems, the (zwi)-L-Phe, and later (imi)-L-Phe are the least stable molecules at the begin of the MDs trajectories, exhibiting recurrent fluctuations in RMSD values. Because the fluctuations occurred early in simulation time, and thereafter the RMDs remained constant, we suggest that the L-Phe was undergoing small structural readjustments in the binding pocket to achieve the most stable binding position.

Overall, the results denote good stability throughout the course of the simulations. The observed fluctuations in RMSD values are expected, as the systems undergo normal structural reorientations. The last are a result of the simulations dynamic character, aimed to mimic physiological conditions. Moreover, the structures arise from a molecular docking protocol, where the achieved pose is obtained through a rigid receptor approach. Therefore, during the protocol of molecular dynamics, the protein experiences minor structural fluctuations arising from the protein no longer being physically restraint, and thus adjusting the new mean conditions.

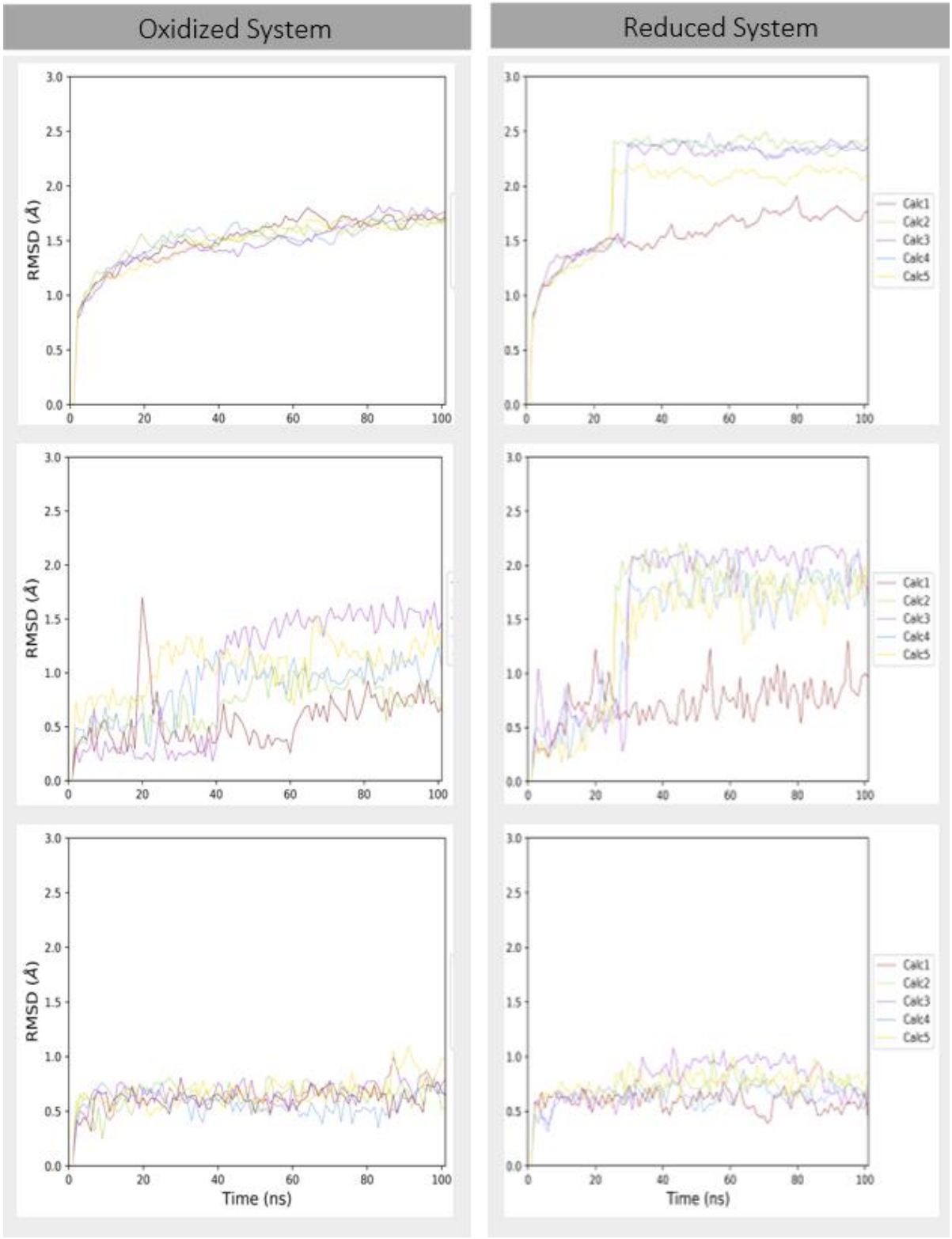


Figure 30 - RMSD values displayed by the Protein (up), Substrate (low), and FAD (low), for the 5 replicates over the course of 100ns, for the oxidized and semi-reduced system.

- RMSF Analysis

To inspect the flexibility of the protein, we calculated the atomic fluctuations of the system residues with respect to an average structure that was generated from a stable part of the trajectory, in each replicate (see methods section for details). Overall, the fluctuation profile for both models (Figure 30) hinted for a low flexibility of the protein, with RMSF values consistently lower than 2Å, denoting minor fluctuations.

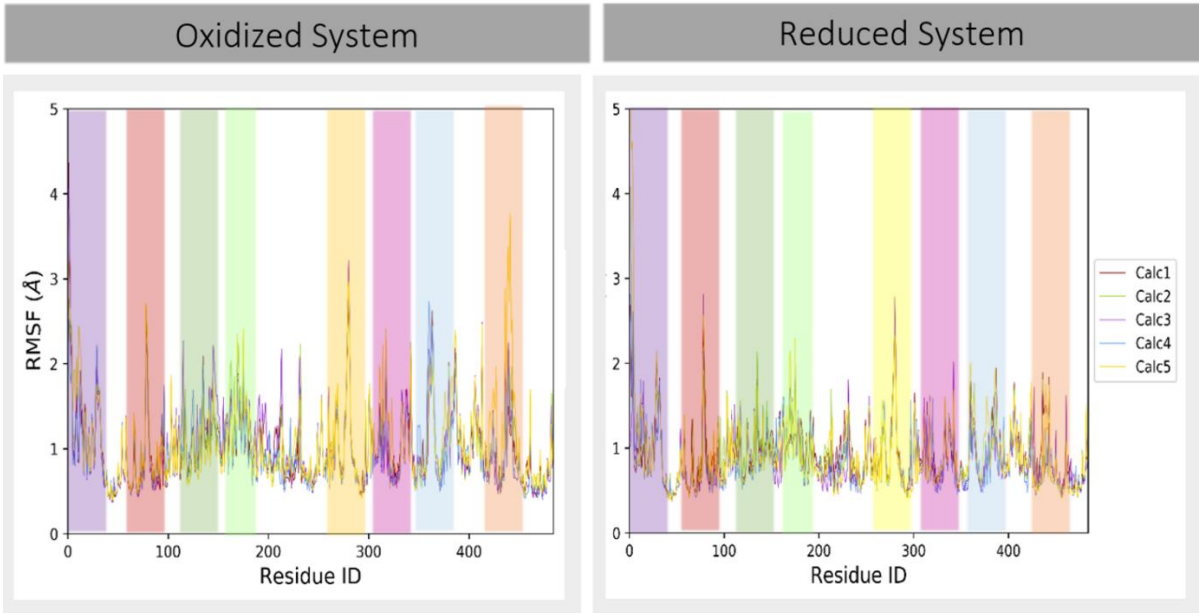


Figure 31 - RMSFs along the MD measured for the 5 replicates, for the oxidized and semi-reduced system. High RMSF zones highlighted in colors.

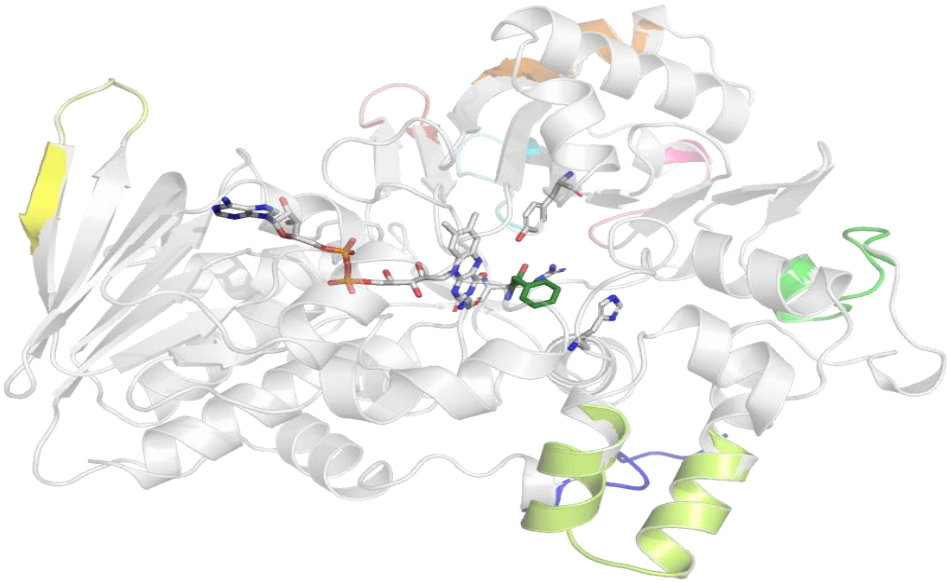


Figure 32 - Tridimensional representation of the protein, with high RMSF zones highlighted in colors.

Exceptions were observed, as expected, at the N-terminal (residues 1-10) that displayed higher fluctuations (until a maximum of 5 Å) and in the loop's regions (residues 130-142, 275-283, 336-344, 358-368, and 405-414). For the last, the RMSFs profiles showed several peaks where the values were equal or higher than 2.5 Å. The loops are the most flexible regions in proteins and in the LAAO are located at the protein surface, more exposed to the solvent, and therefore are subject to bigger structural oscillations.

- Distances

The integrity of the active site of both systems was analyzed by the visual inspection of several frames generated along the MD trajectories and through the analysis of distances of key residues in the catalytic cycle. We have considered as key residues for the reaction the amino acids that interact and stabilize L-Phe— Tyr 372, Arg 90, and Gly464 — and the residues that could play a direct role in the mechanism — FAD (at N5), His 223 or L-Phe itself.

This integrity was exploited for both semi-reactions present in the catalytic cycle, thus allowing to establish the importance of each residue in each reaction step.

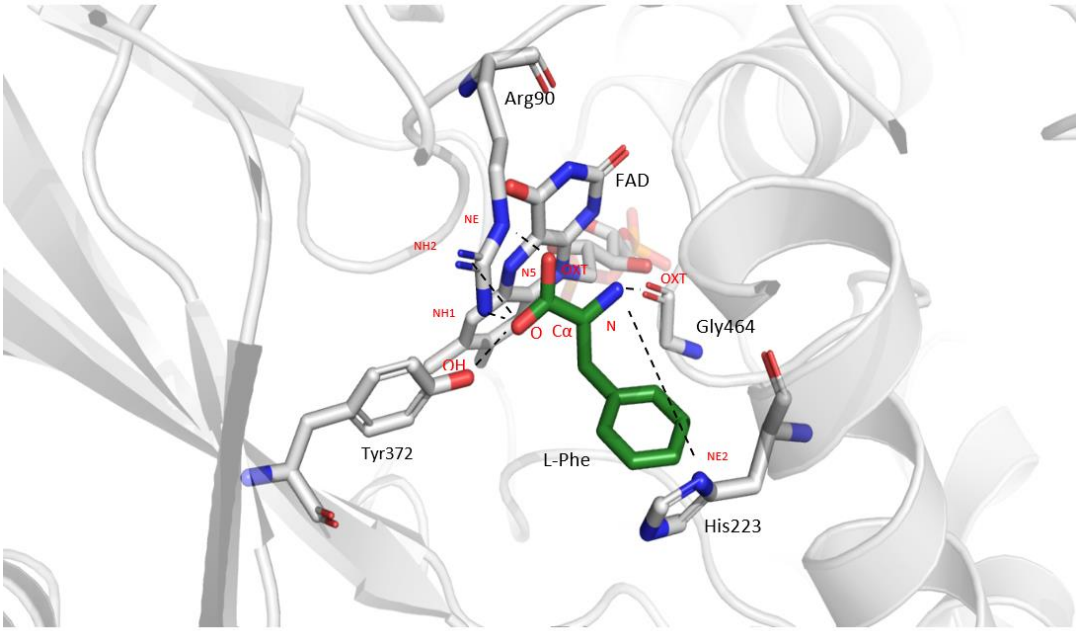


Figure 33 - Close-up of the active center of the protein, highlighting the residues important for the reaction, either by stabilization (Arg90, Tyr372, Gly464) or by a direct participation (FAD and His223). Residues labelled in black Distances measured are represented in dashed. Atoms considered for the measurements labelled in red.

- What is His 223 behavior in presence of the FAD and FADH cofactors?

We started by measuring the distance between HID 223 and the NH_3^+ group of (zwi)-L-Phe presence of FAD; and in the semi-reduced system, the distance between HIP 223 and the NH_2 group of (imi)-L-Phe. As can be observed in Fig. 33, it is clear that the HID223 remains close to the amine group of (zwi)-L-Phe, when in presence of FAD. In contrast, the HIP223 pulled itself away from the NH_2 of (imi)-L-Phe very early in the simulations, during the equilibration phase.

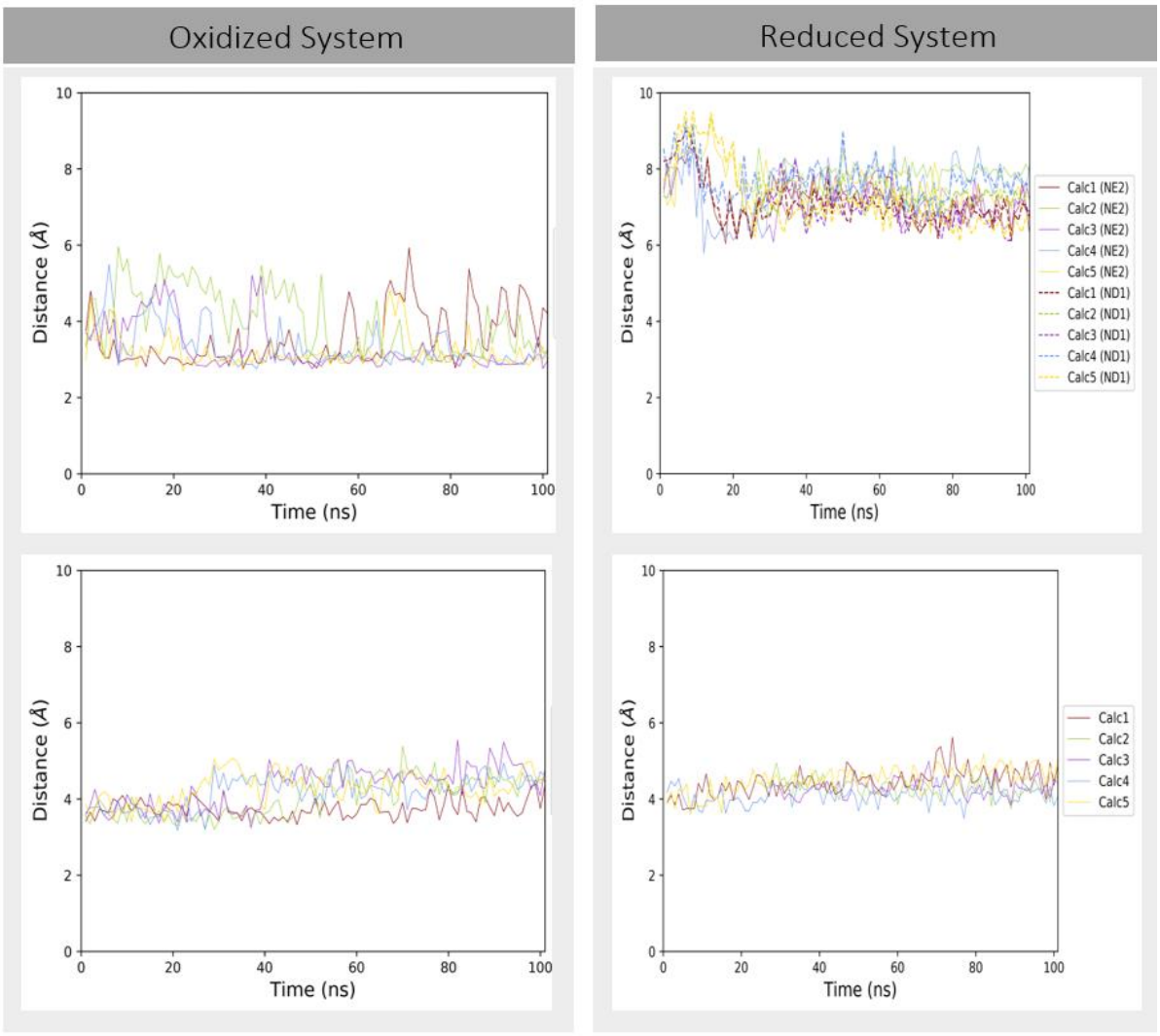


Figure 34 - (Up) Distances between the His223 and the N of L-Phe, calculated for the 5 replicates, for oxidized and the semi-reduced System. Distance measured for the NE2 atom of His223 represented in lines. Distance measured for the ND1 atom of His223 represented in dashed. (Down) Distances between the Cα of the L-Phe and N5 of the FAD molecule, calculated for the 5 replicates, for oxidized and the semi-reduced System.

However, the profile for the distance between the alpha carbon of (zwi)-L-Phe and the N5 of FAD, and the distance between the (imi)-L-Phe and FADH (N5) denoted that the L-Phe occupies the same position in the catalytic cavity in both systems (<5 Å).

Overall, the distance observed between the HIP223, and (imi)-L-Phe is far greater than the observed between the HID223 and (zwi)-L-Phe, denoting a withdrawal between this residue and the L-Phe happening between systems. The results indicate that the L-Phe itself does not re-orient in the catalytic pocket but rather the HIS223 suffers a conformational change.

Moreover, the analysis of the distances between the Arg 90 (atoms NE, NH1 and NH2) and the carbonyl group of the L-Phe for the oxidized system showed a strong H-bond between the Arg90 (NE and NH1) and the O of the L-Phe (~3 Å). The distance between Arg 90 and (zwi)-L-Phe remains constant throughout the dynamics. H-bonds were established between NE and the NH2 atoms of Arg90 and the O and OXT atoms of the zwitterionic residue, respectively. The distance between the two residues increases during some trajectory frames due to the entrance a water molecule, the last mediates the interaction between the two amino acids.

In contrast, the H-bond formed by the carbonyl group of Gly 464 and the NH_3^+ at the L-Phe was broken after 20ns of simulation in replicates 1 and 2; or even before, in replicates 3 to 5. Here, the distance values reached 7-8 Å. These data suggests that Gly464 is not responsible for the substrate stability.

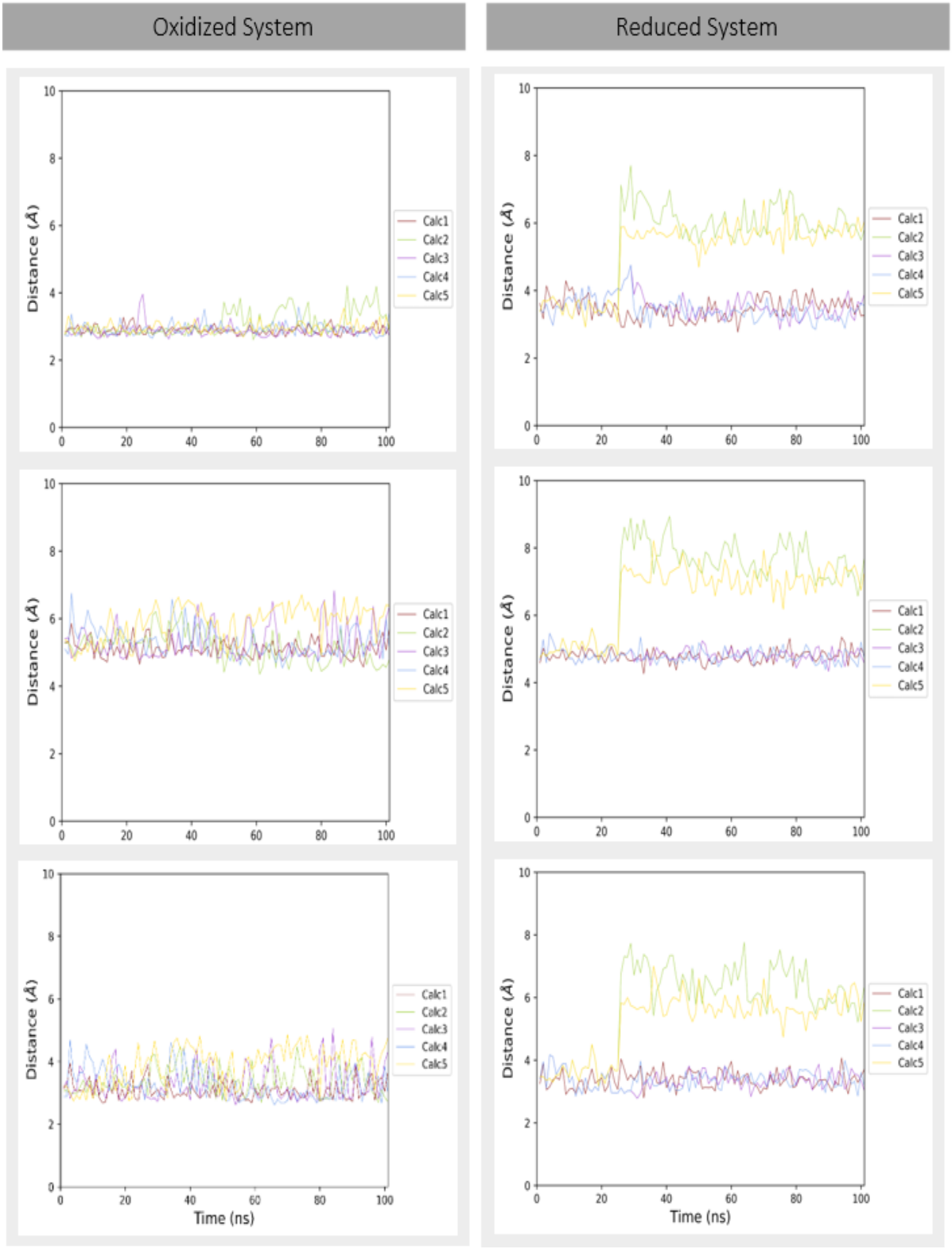


Figure 35 - Distances between the NE, NH1 and NH2 of Arg90 and O and OXT of L-Phe, over the course of the simulations, for the oxidized and the semi-reduced System. (Up) – NE – O. (Mid) – NH1 – OXT. (Down) – NH2 – OXT.

For the semi-reduced system, there is an H-bond between Arg90 and the (imi)-L-Phe for replicates 1, 3, 4 (~3.5 Å). Exceptions are observed for replicates 2 and 5, where there is a huge distancing between these residues, when *t=30ns*.

Regarding Gly464, the distance between this residue and the N-terminal of the substrate remained stable during the five trajectories calculated (3 Å). Here, in opposite to the oxidized system, Gly464 seems to be important to stabilize the (imi)-L-Phe in the active site.

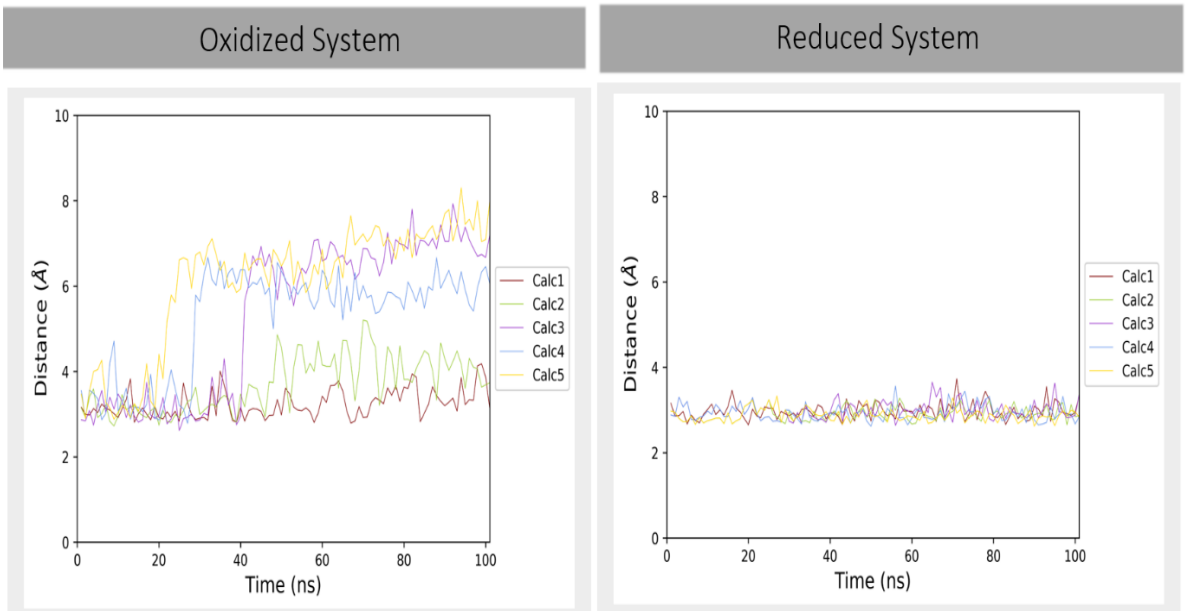


Figure 36 - Distance between Gly 464 (OXT) and the L-Phe (N), for the Oxidized and the Semi-reduced system.

Additionally, Tyr372 also played a key role in the substrate stabilization. The OH atom of Tyr372 and O atom at L-Phe formed an H-bond, exception is the replicate 2. This interaction in lost when in presence of FADH and (imi)-L-Phe.

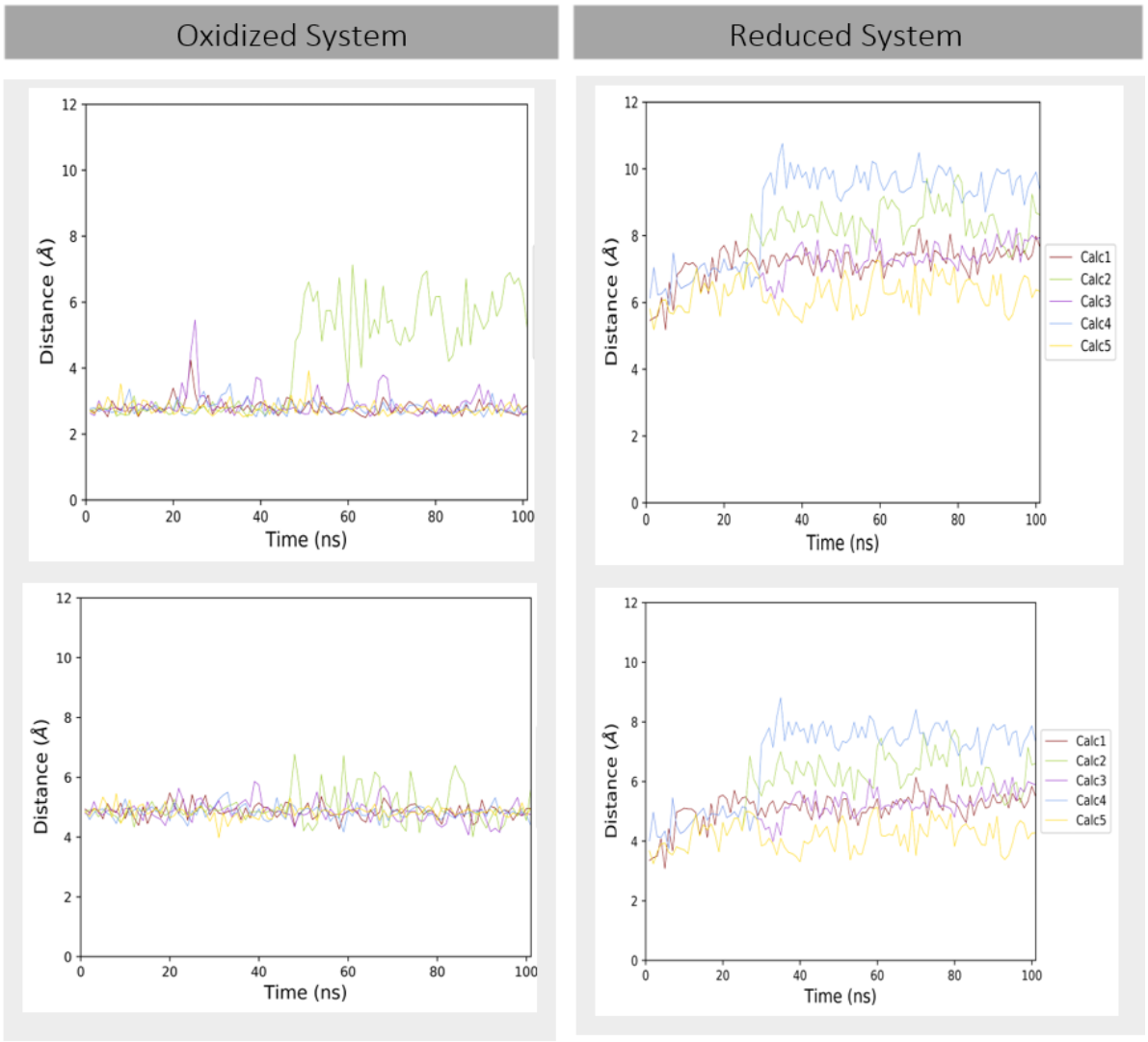


Figure 37 - Distance between Tyr372 (OH) and L-Phe (O and OXT) for the Oxidized and the Semi-reduced system. (Up) – OH-O. (Down) – OH – OXT.

For Arg322 the results suggested that this residue is not responsible for keeping the substrate neither the iminoacid (distances are as far as 10 Å) but could play another role in their accessibility.

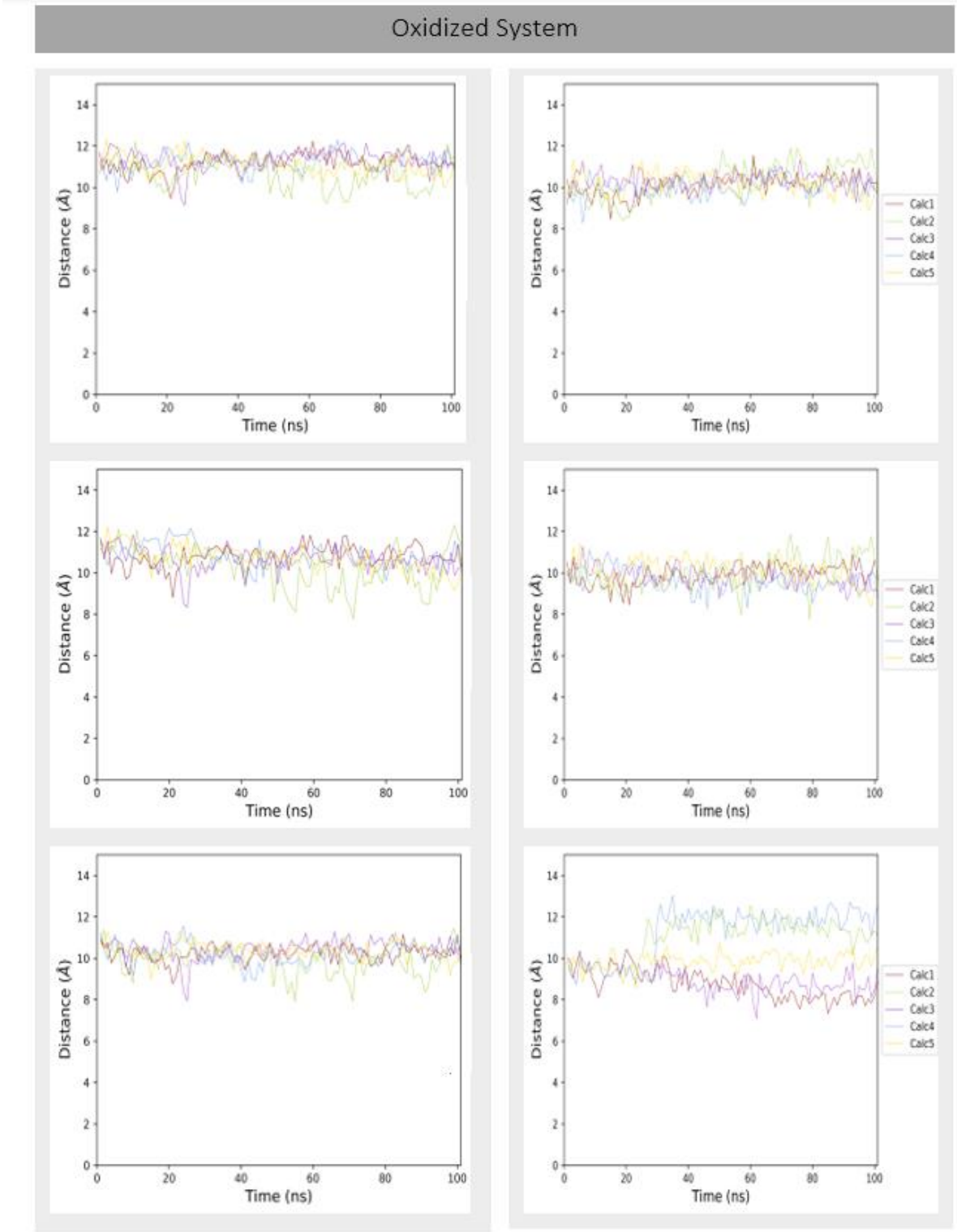


Figure 38 - Distances between Arg322 (NH1, NH2, and NE) and L-Phe (OXT and O), along the simulations, for the oxidized system. (Up) – NH1 – OXT and NH1 – O, respectively. (Mid) – NH2 -OXT and NH2 – O, respectively. (Down) – NE – OXT and NE – O, respectively.

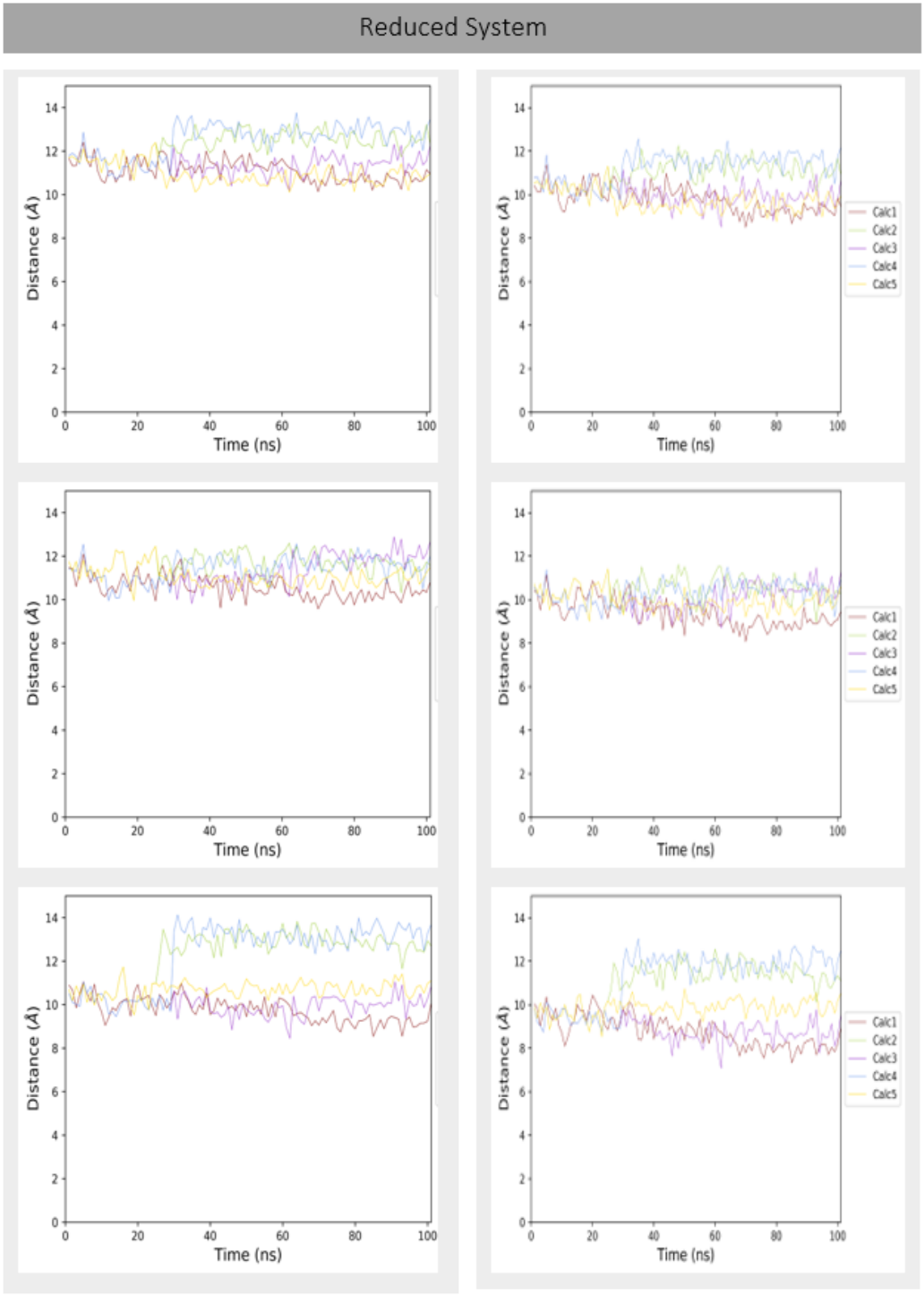


Figure 39 - Distances between Arg322 (NH1, NH2, and NE) and L-Phe (OXT and O), along the simulations, for the semi-reduced system. (Up) – NH1 – OXT and NH1 – O, respectively. (Mid) – NH2 -OXT and NH2 – O, respectively. (Down) – NE – OXT and NE – O, respectively.

At this point, we can conclude that the two forms of the L-Phe induce a great rearrangement in the residues that compose the catalytic cavity in order to form interactions that stabilize 1st the substrate, and then the intermediate. This is indeed what we observe, in the concomitant movement of Tyr372 and the Gly64.

- Is there a stabilization of His 223 throughout the MD?

The infer a possible stabilization of the His223 throughout the simulations, the distance between the ND1 atom of His223 and the OD1 and OD2 atoms of Asp224. For the oxidized system there is a hydrogen bond between the ND1 and OD1, depicted by a stabilized distance value of 3 Å for most replicates, when *t=40ns.* Exceptions are observed for replicate 5, where there is a great fluctuation in distance values until *t=80ns*, when the value converge to around 3 Å. For ND1-OD2 there is a confluence in distance values, for most replicates, around 5 Å. Exceptions are observed for replicates 2, 4 and 5, where the convergence in distance values is only achieved at *t=80ns* for replicate 2 and 5 and *t=90ns* for replicate 4.

For the semi-reduced system, huge fluctuation in value is observed for the distances between the OD1 atom and both the ND1 and NE2 atoms of HIP223, with values ranging from 4 to 6 Å. An outlier is observed for replicate 4, where the distance between NE2 and OD1 is 5 Å at the beginning of the simulation but, when *t=30ns* until the end of the dynamics presents distances values of around 3 Å. For the distance between the OD2 atom of Asp224 and HIP 223, the distances values considering the ND1 atom of His are on the average of 5 Å, for most replicates, an exception is observed for replicate 4, where when *t=30ns* the distance values are around 6 Å. The distance considering the NE2 of His and the OD2 atom of Aspartate stabilizes in values around 2.5 Å, when *t=30ns*. An exception is once again notes for replicate 4, where the distance value commences at 2.5 Å but rises in value, to 4 Å, at *30ns* of simulation.

Overall, the stabilization of His223 is ensured by Aspartate224, for the oxidized system, by a hydrogen bond established between the ND1 of the imidazole ring of His223 and OD1 of the carbonile group of Asp224. For the semi-reduced system, the stabilization of HIP223 is ensured by a hydrogen bound established between the NE2 atom and OD2 atom of aspartate.

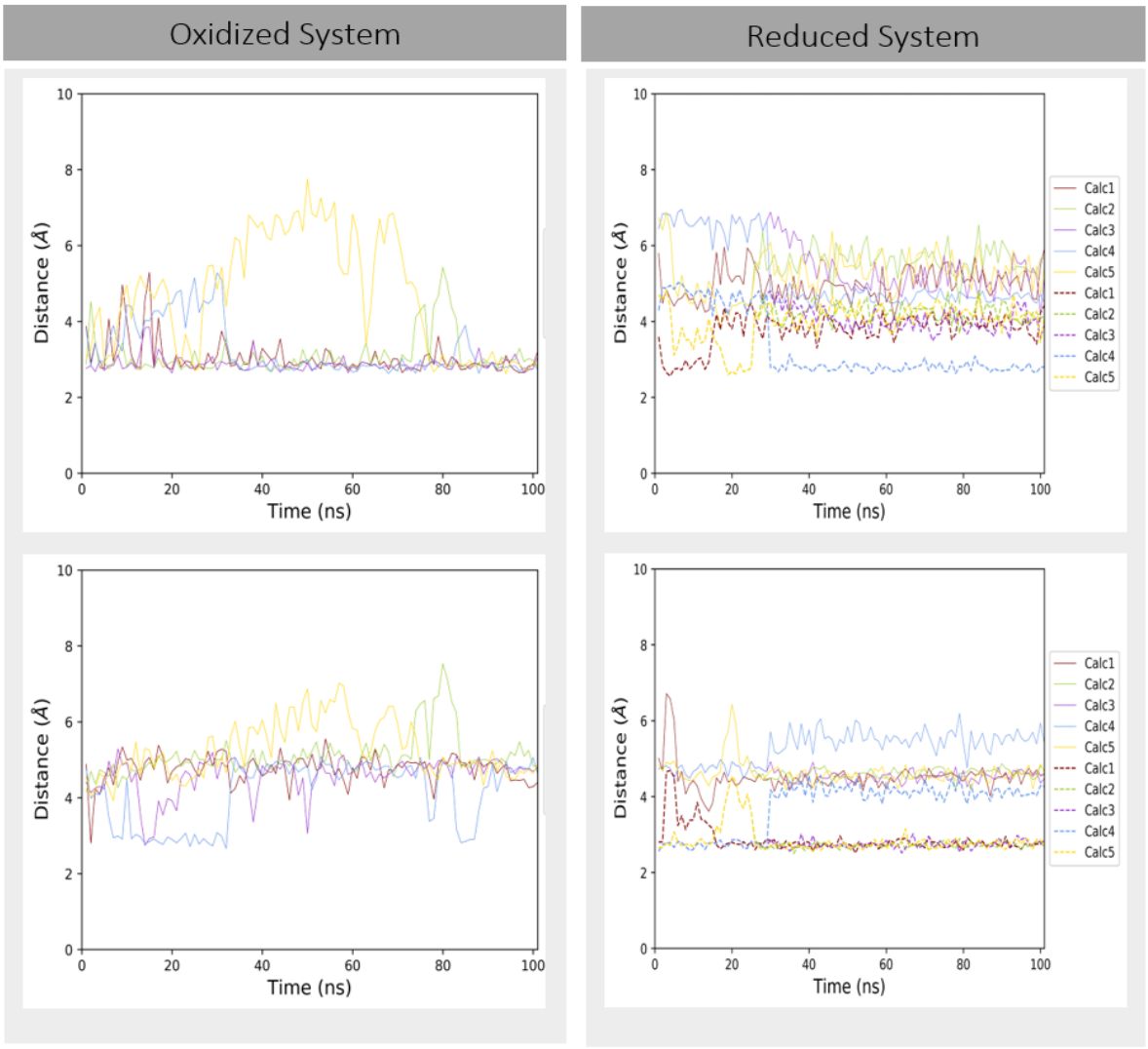


Figure 40 - Distance between Asp224 (OD1 and OD2) and His223 (ND1 and NE2 for the semi-reduced system) along the simulations and for the oxidized and semi-reduced system. (UP – OD1 – ND1/OD1 – NE2). (Down) – OD2 – ND1/OE2 – NE2.

- Dihedrals Analysis

The orientation of the His223 and Arg322 towards L-Phe and the possibility for a conformational change was inspected through the analysis of dihedral angles adopted by these residues along the MD trajectories. Additionally, possible conformational changes in the residues important for the stabilization of L-Phe in the active site were inspected.

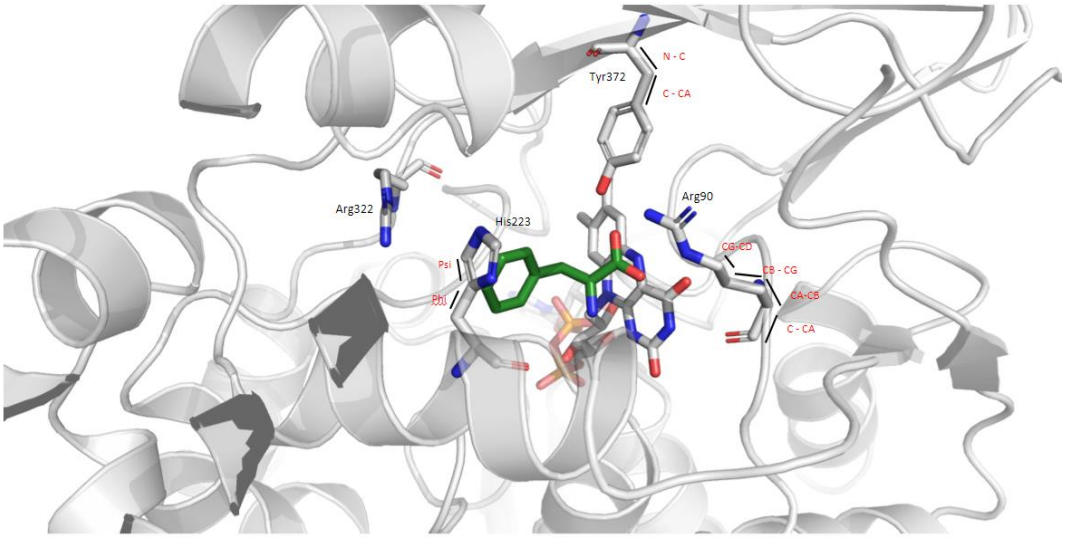


Figure 41 - Close-up of the active center of Bothrops atrox's LAAO, highlighting dihedrals that were measured (in lines). Dihedral angles labeled in red. Residues considered labeled in black. PS: Same labels for the dihedrals between Arg90 and Arg322.

In the case of His223, this analysis was conducted by inspecting structural alterations in the imidazole ring (Phi and Psi, respectively), while in Arg322 this analysis was used to analyzed structural changes in the aliphatic (between C-CA-CB-CG, CA-CB-CG-CD, and CB-CG-CD-NE, CG-CD-NE-CZ respectively) side chain. Additionally, the dihedral angles for the residues Tyr372 (N-CA-CB-CG and CA-CB-CG-CD2) and Arg90 (C-CA-CB-CG, CA-CB-CG-CD, and CB-CG-CD-NE, CG-CD-NE-CZ), residues important for the stabilization of the L-Phe were inspected.

For the oxidized system, the results points for a specific conformation for the His. Aditionally, the stability of this conformation is highlighted by the coeherent angle vlaues troughout the MD, with Psi and Phi angles displaying and average of -100 degrees and 75 degrees, respectively. The only outlier for this observation was the fifth replicate, where there are aberrant fluctuactions in dihedral values between 30 and 70 ns of simulations

For the semi-reduced system, structural changes in the imidazole ring of His223 were evident, as displayed by huge fluctuations in the angle's values. They happened when the simulation reached 20ns, for most of the replicates. The only discrepancy was in replicate 3, where a huge fluctuation in the dihedrals was observed latter in the simulation, when *t=40ns*.

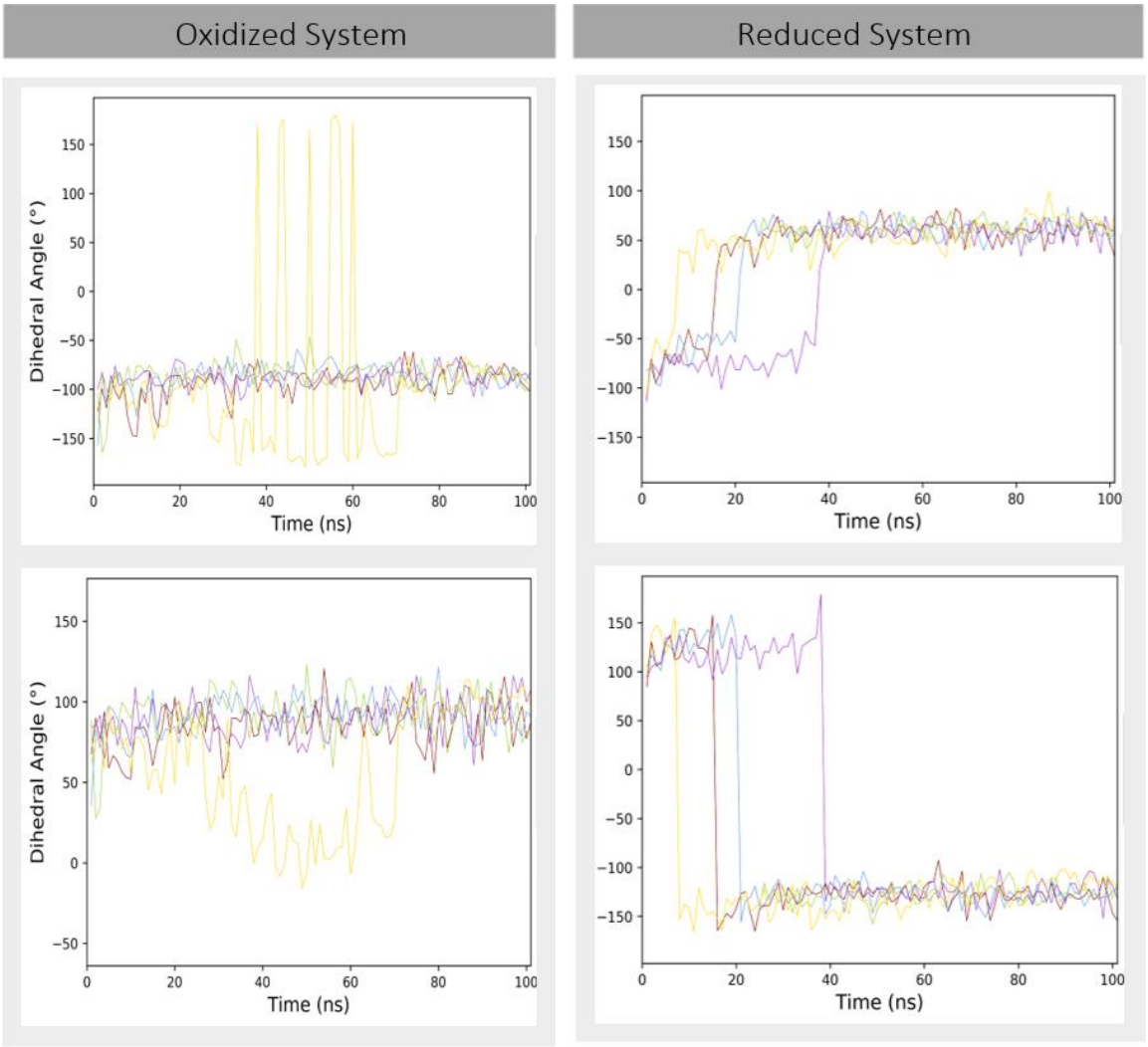


Figure 42 - Dihedral angles displayed by the imidazole ring of His223, highlighting possible conformation changes, for the oxidized system and the semi-reduced System. (Up) – Psi dihedral. (Down) – Phi dihedral.

The two sets of data measured opposite dihedrals in the imidazole ring, where these results pointed to a change in the adopted conformation for His, between the two systems present in the two semi-reactions comprising this mechanism.

This observation is in line with the proposal of *Moustafa et al.* that suggested that the His223 acts as gate, occupying a conformation that encloses the L-Phe in the catalytic cavity, and then experiences a conformational change in order to allow the exit of the product — the keto acid — unblocking its passage.

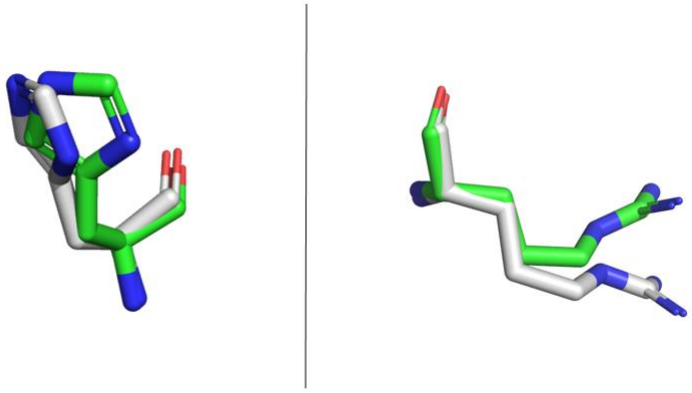


Figure 43 - Observed changes in dihedral angles for His223 and Arg 322, pointing for a change in conformation. (White) – Residues belonging to the first frame for the simulation of the oxidized system. (Green) - Last frame of the simulation of the semi-reduced system. Superimposition

Finally, the dihedral angles measured for the aliphatic chain of Arg322 suggested a very stable residue since the first 30ns of the MD simulations, in both systems. For the oxidized system, in the very begging of the simulations, there is a specific conformation in the aliphatic side chain, highlighted by coherent dihedral angle values namely for C-CA-CB-CG and CB-CG-CD-NE with a mean of 150 and -50 degrees, respectively. However, this was a conformational transition until the residue adopts its final position, as noted by the fluctuation graphs. For the semi-reduced system, similar conclusions can be taken. Indeed, for replicates 1,3, and 4, the value for the C-CA-CB-CG dihedral angles begins at 150 degrees, but at *t=30ns*, until the end of the simulations, stabilizes around 50 degrees. However, there are outliers for replicates 2 and 5. For CB-CG-CD-NE dihedral angle, the simulations begin with values of 50 degrees but at *t=30ns* these values stabilize at -50 degrees.

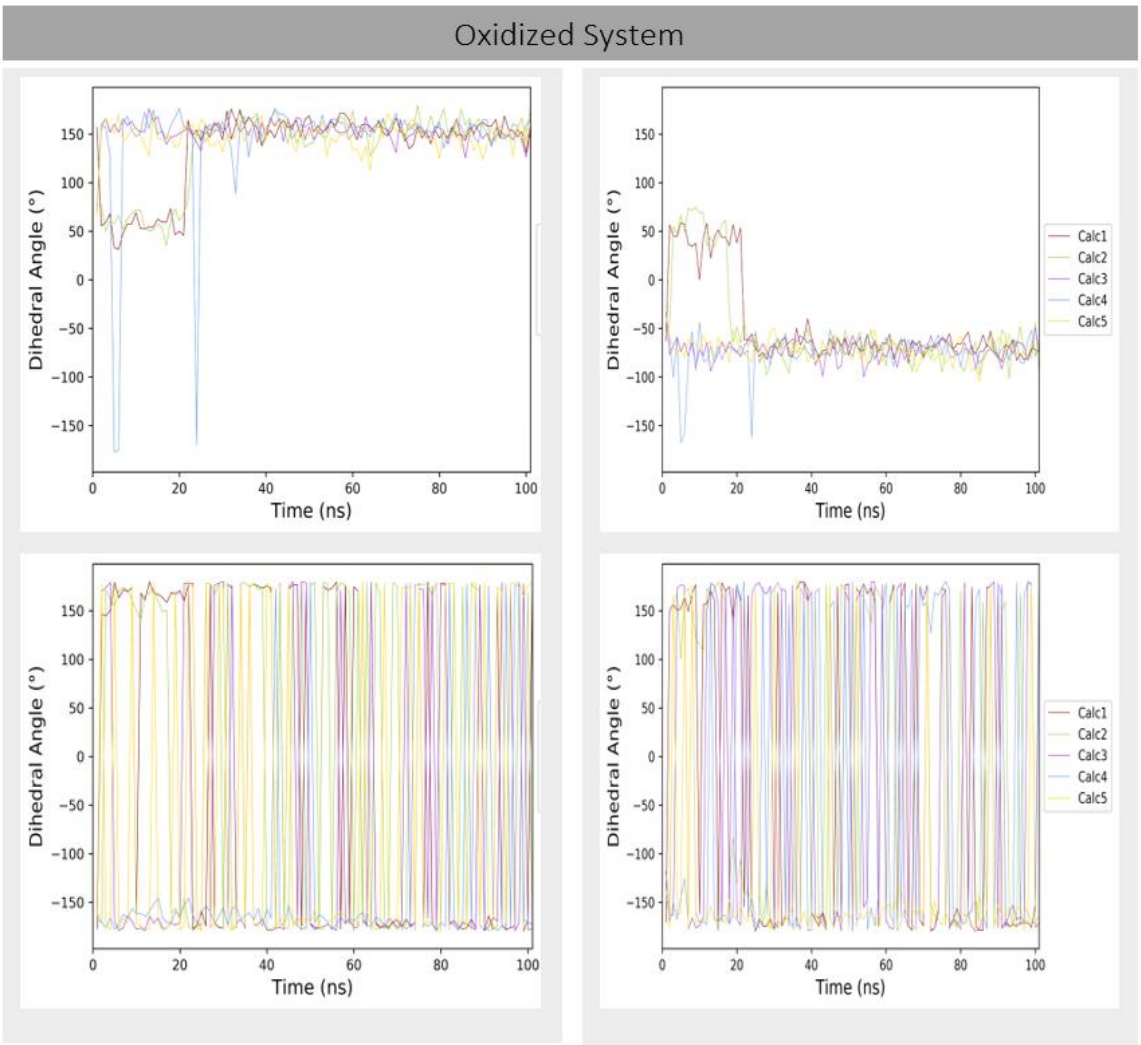


Figure 44 - Dihedral angles displayed by the aliphatic chain of Arg322, highlighting possible conformation changes, for the oxidized system.

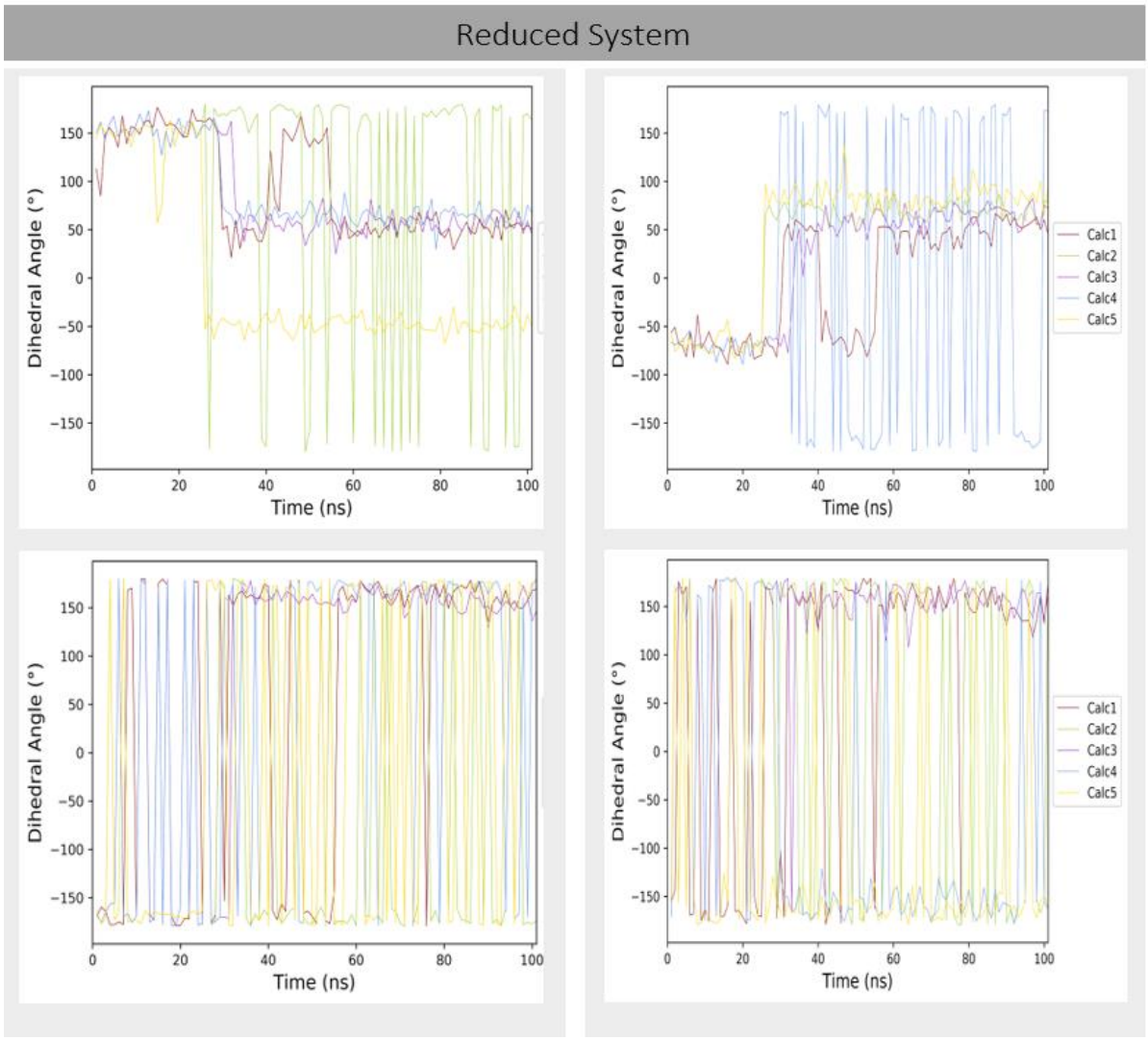


Figure 45 - Dihedral angles displayed by the aliphatic chain of Arg322, highlighting possible conformation changes, for the semi-reduced system.

In case of Arg90, both systems suggest great stability for the aliphatic chain of this residue throughout the MD, as depicted by light fluctuations in value for the C-CA-CB-CG, CA-CB-CG-CD, CB-CG-CD-NE, and CG-CD-NE-CZ dihedral angles. The dihedral angles of CA-CB-CG-CD estimated for the oxidized was around -75 degrees while in the semi-reduced system it was -80 degrees. The data indicate that in both cases the side chain of Arg90 points towards the same direction.

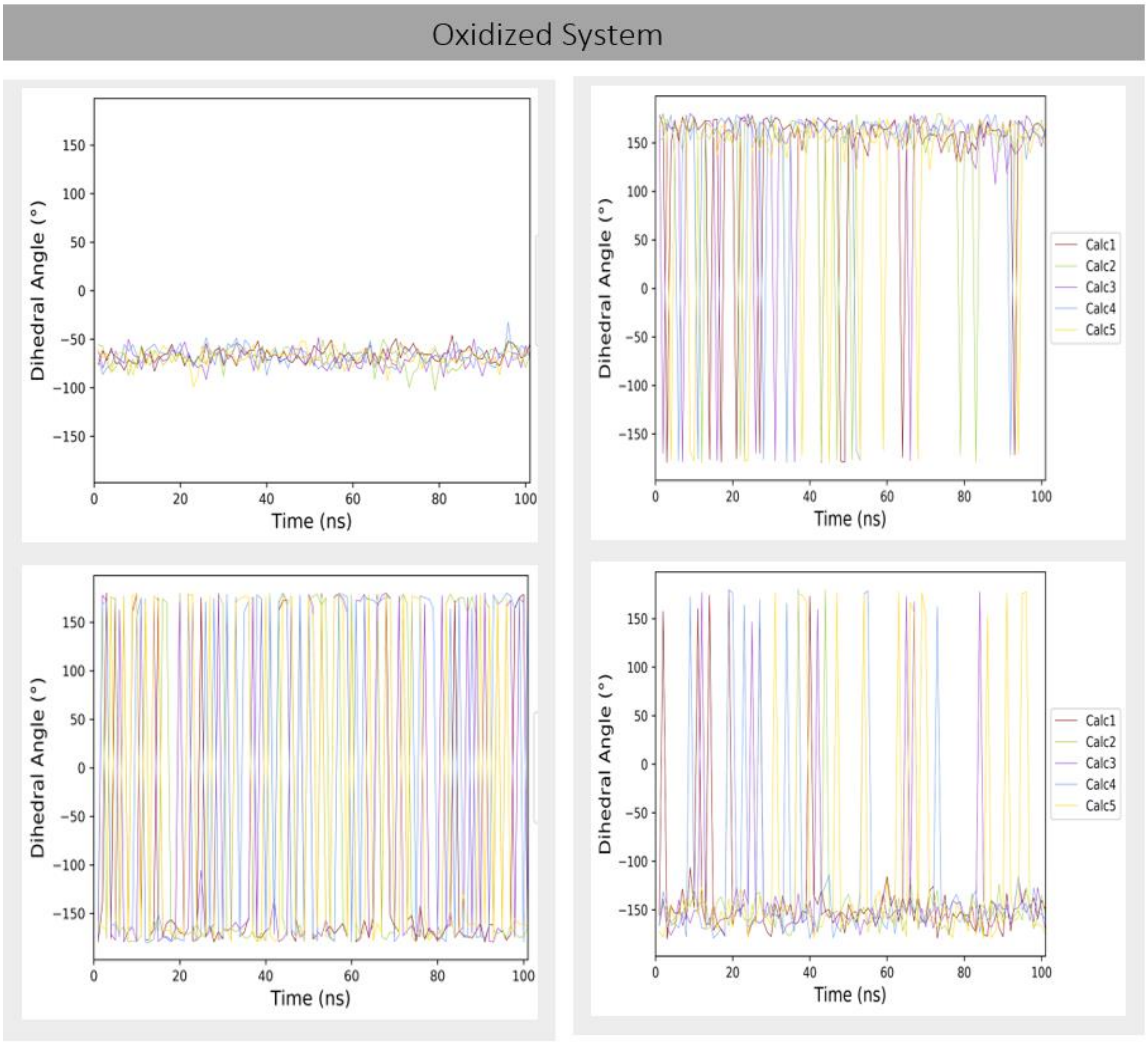


Figure 46 - Dihedral angles exhibited by Arg 90 along the simulations, for the oxidized system. (Up) – C-CA and CB-CG dihedrals. (Down) – CB-CG and CG-CD dihedrals.

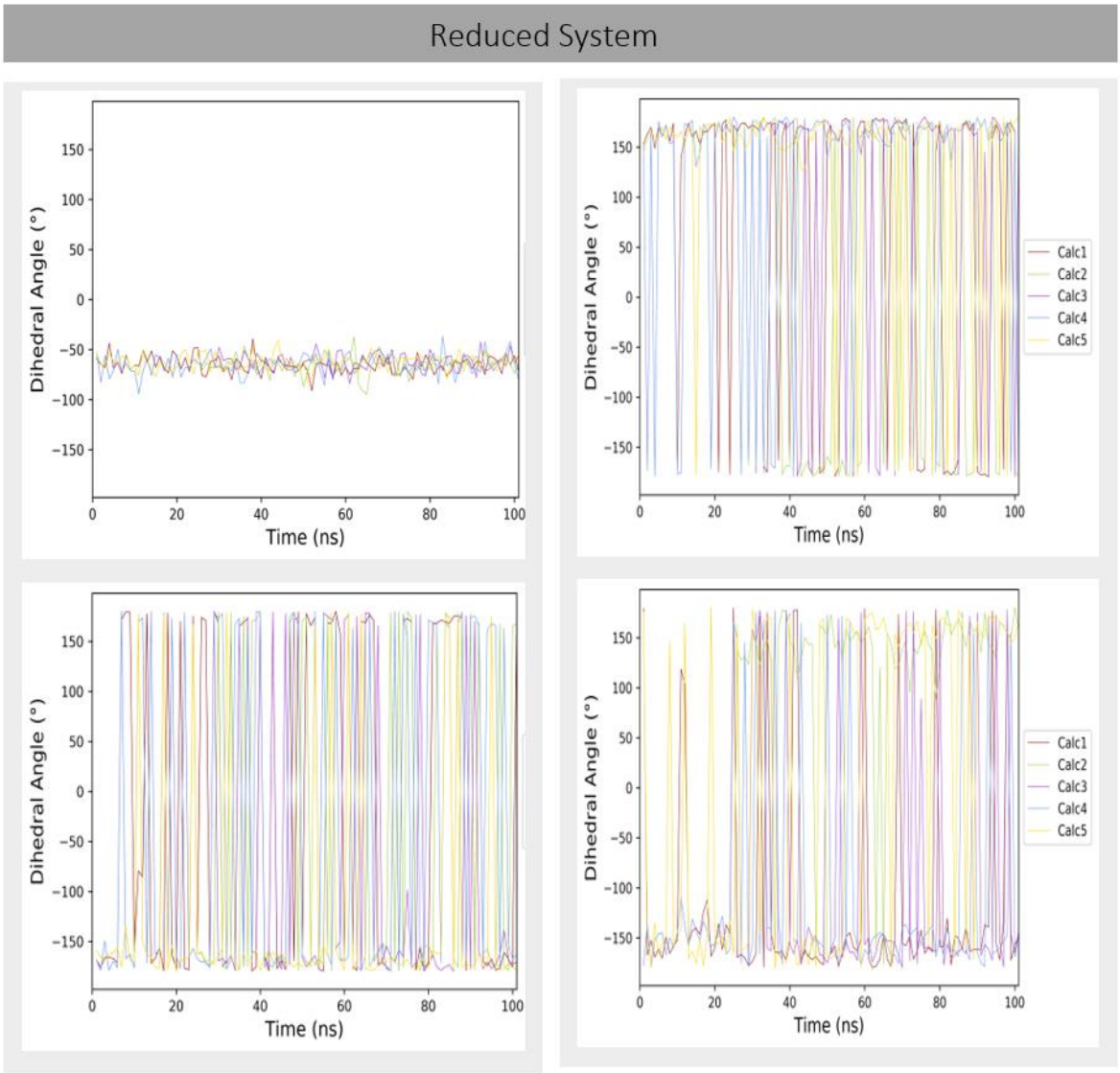


Figure 47 - Dihedral angles exhibited by Arg 90 along the simulations, for the semi-reduced system. (Up) – C-CA and CB-CG dihedrals. (Down) – CB-CG and CG-CD dihedrals.

Finally, the analysis of the dihedral angles for Tyr372 denoted similar behaviors for the systems along the simulations, where values for the CA-CB-CG-CD2 dihedral oscillated between +180 and -180 degrees. In contrast, the dihedral angle defined by the atoms N-CA-CB-CG, also oscillates between +180 and -180, exceptions were the replicates 4 and 5, in the semi-reduced system (~100 degrees).

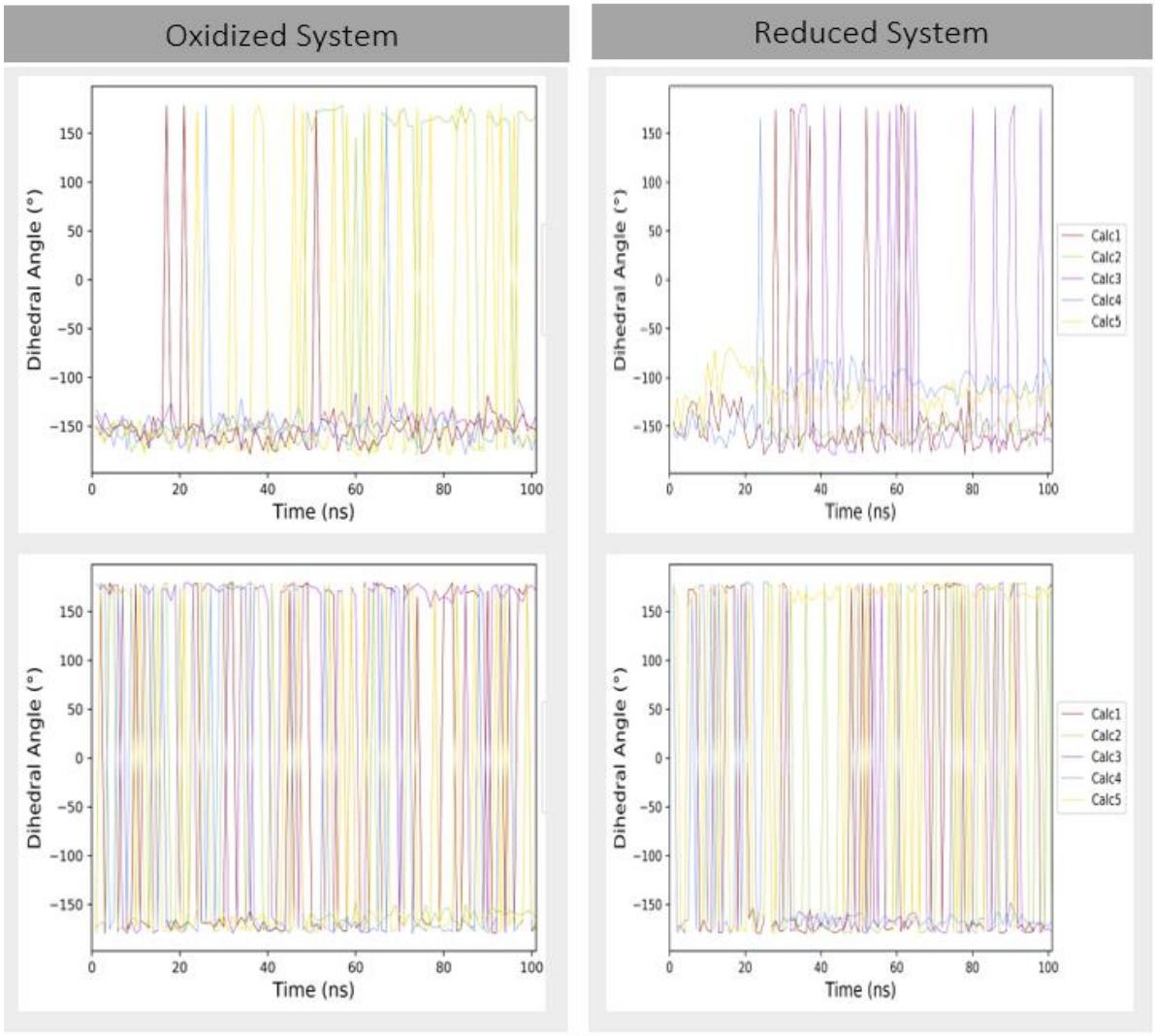


Figure 48 - Dihedral angles exhibited by Tyr372 along the simulations, for the oxidized and the semi-reduced systems. (Up) - N-CA-CB-CG dihedral. (Down) – CA – CB – CG – CD2 dihedral.

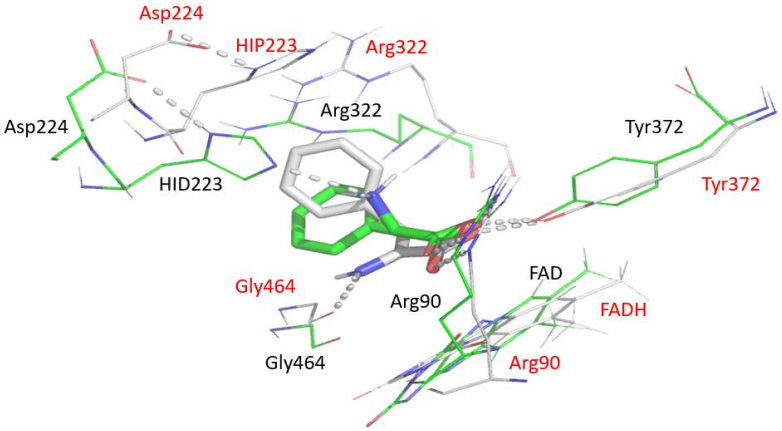


Figure 49 – Superposition of the minimized average structure of the oxidized (in green) and the semi-reduced (in white) systems, highlighting the possible conformational changes. Residues belonging to the oxidized system labelled in black. Residues belonging to the semi-reduced system labelled in red.

Quantum Mechanics/Molecular Mechanics Studies

- Selection of the Initial Structure

Several structures were inspected to be used as a starting point for further steps under this protocol. Frames resulting from the production step of Molecular Dynamics, plus the minimized crystallographic structure were analyzed.

The selection of the frames containing the structure to be used as a starting point for the study of the first semi-reaction of the mechanism, started from the analysis of the angles established between the Nε2 atom of His 223 and the guanidine group of the (zwi)-L-Phe, plus the one established between the Hα of (zwi)-L-Phe and the N5 of the FAD molecule. Additionally, the distances between the catalytic residues (His223 and FAD) and the L-Phe were also used as a criterion. The angle between His223 and (zwi)-L-Phe played a minor role as a criterion given the reduced distance between those two residues. The angle between the Hα of (zwi)-L-Phe and N5 of the oxidized FAD molecule played a bigger role in the selection of frames to be used. At the end, a favorable orientation and distance for the Hα of L-Phe in relation to the N5 atom of the isoalloxazine group of FAD was inspected. Two initial frames were selected from this step. (Figure 50)

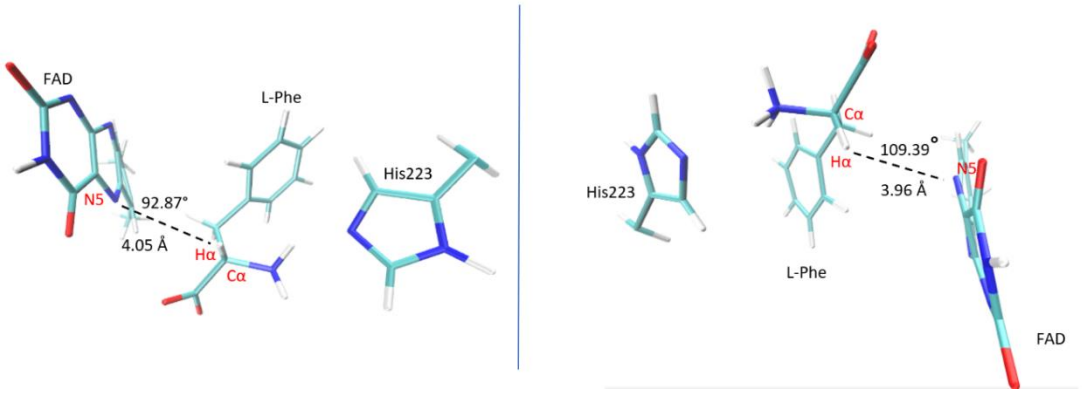


Figure 50 - Analysis of distances and torsional angles between the FAD (N5) and (zwi)-L-Phe (Hα), comprised of the Cα and Hα atoms of the L-Phe, plus the N5 atom for the isoalloxazine group of the FAD molecule, from structures arising from the production phase of MD. Considered atoms for the analysis are labelled in red. Considered residues are labelled in black. Measured distances and torsional angles are labelled in black.

The minimized crystallographic structure arising from the production step was also selected. Ultimately, this structure was selected was the initial point for the subsequent steps.

- Preparation of the System – Definition of Layers and Freezing

The division of a chemical system to be treated with different degrees of theory, as well as the freezing of atoms revolving the high layer (HL), represents a pivotal step in any reaction mechanism study. Two different strategies were adopted in the selection of the layers to be used. The fist approach saw the HL being comprised of 151 atoms, made of residues of Leu88, Met89, Arg90, His95, His223, Tyr372, Trp465, Wat501, (zwi)-L-Phe, plus the isoalloxazine group of FAD.

As for the second approach, the HL as comprised of 115 atoms, made of residues of Arg 90, His223, Tyr372, Trp465, Wat501, (zwi)-L-Phe, and the isoalloxazine part of FAD (Figure 51).

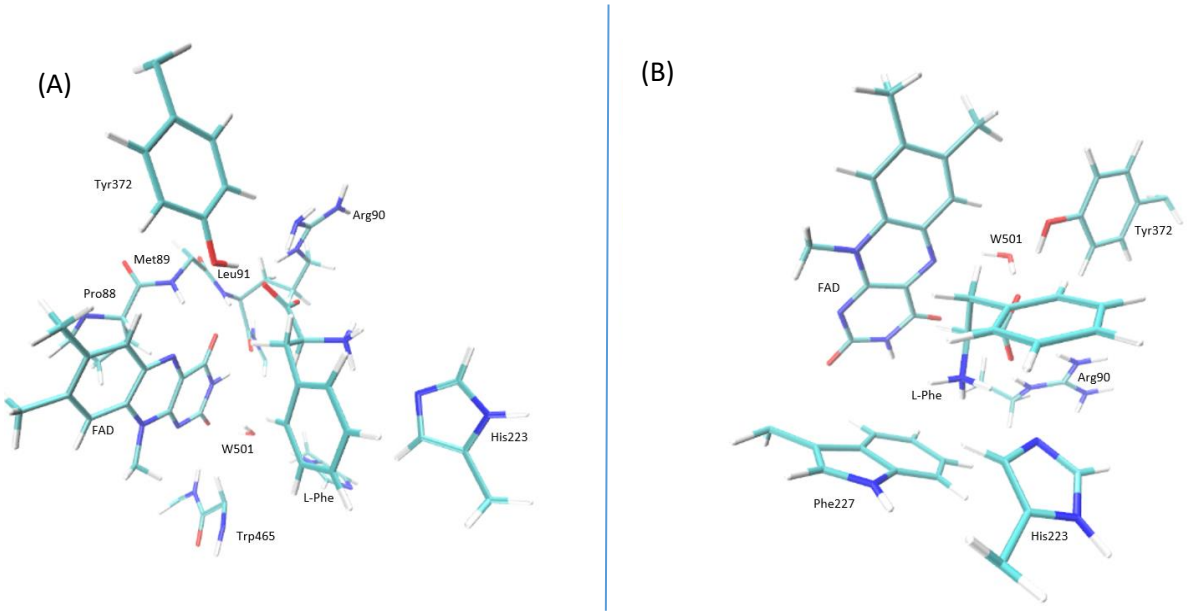


Figure 51 - The two analyzed strategies for the composition of the HL, with 151 atoms (A) and 115 atoms (B).

Moreover, for the freezing of atoms, again different approaches were conducted, differing in the size of the free layer, with a radius of 15 Å and 5 Å from the HL being adopted, respectively. Overall, the system to be optimized was made of the minimized structure arising from the production step of MD, where the HL for the system was made of 115 atoms and the unfrozen layer had a 5 Å radius around the HL.

- Geometry Optimization

Geometry optimization was conducted over the minimized crystallographic structure. The structural integrity of the system after this step was maintained. Moreover, the catalytic distances between His 223 and the N-terminal group of (zwi)-L-Phe have decreased, from 2.08 to 1.65 Å. Also, the distance between the α Carbon of (zwi)-L-Phe and the N5 of the isoalloxazine group of FAD has experienced changes, from 3.67 to 3.05 Å. The high layer of the initial structure before and after the optimization are depicted on figure 52, highlighting the catalytic distances.

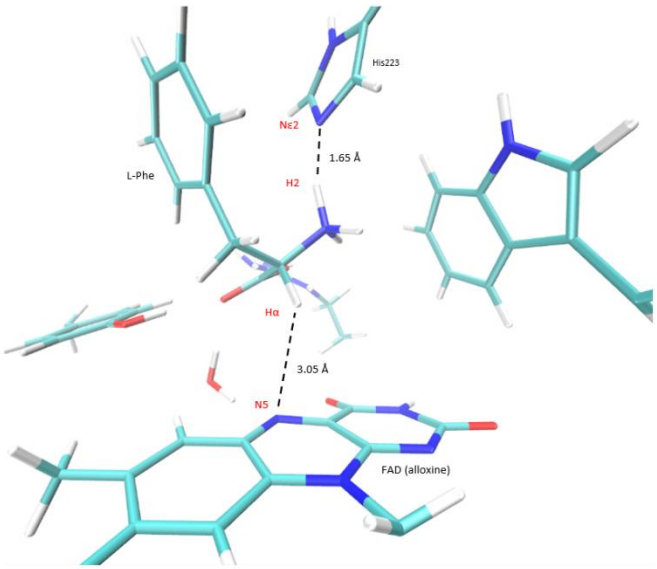


Figure 52 - Highlight of the system's HL after ME+EE. Catalytic important residues and distances labeled in black. catalytic important atoms labeled in red.

- Step 1 – Deprotonation of L-Phe

The first step of the LAAO's catalytic mechanism, as proposed by [35, 39, 40] sees the deprotonation of the amino acid in its zwitterionic form, by a residue of His. In order to infer the liability of this step on *Bothrops atrox's* LAAO, a Potential Energy Surface (PES) scan was conducted, to calculate the enthalpy and later the Gibbs free energy associated with this step.

To achieve this, the H2 atom of (zwi)-L-Phe was incrementally closed to the Nε2 of His223, so that a favorable distance for the deprotonation has achieved. A scan of 14 steps with a -0.1 Å incrementation was implemented. This culminated with a proton transfer from the substrate to the His. The graph with the associated PES is represented on figure 53.

The energy barrier calculated for the first transition state (TS1) was 0.34 kcal/mol. According to these results, it can be denoted that the first reaction is favorable, as the energy barrier is quite small. It is lower than kBT which means that the reactant will not resist a simple bond vibration. Moreover, this reaction happens spontaneously, as the Gibbs free energy for the reaction is negative, given that System's total energy for the 1 intermediary (INT1) is lower than the one calculated for the stating reactant. In

conclusion, the true ground state of this system has the His223 positive and the substrate deprotonated.

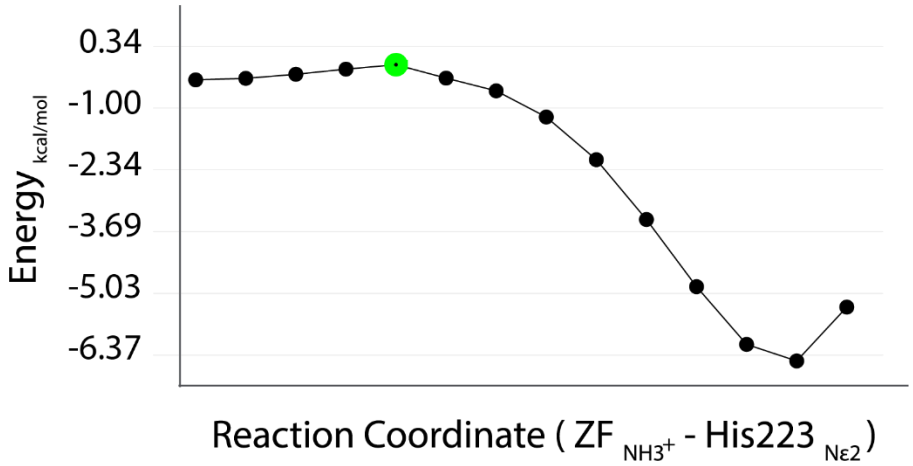


Figure 53 - Energetic profile of the Potential Energy Surface (PES) for the deprotonation step of the catalytic mechanism. The structure corresponding to the transition state is highlighted in green.

- Key Aspects and Future Perspectives

The selection of the initial structure as well as the division of the HL and the freezing atoms took some deliberation before making a final decision. The selection of the minimized structure arising from the MD as the initial structure, in opposition to the systems coming from frames generated in the production phase of the MD as due to the latter not converging trough EE.

Regarding the construction of the HL, the system composed of the minimized structure, where the HL was made of 151 did not converge with EE. Through visual inspection of the system, it was observed an exposure of the substrate to the solvent, posed a major drawback for the optimization of the system. Thus, a new division of the HL, plus the reduction of the radius for the freezing layer was adopted.

The analysis of the PES arising from step1 painted one conclusion – the high plausibility of the first step being the deamination of the substrate by His223. However, the obtained profile of the Total Energy of the System was quite distinct.

From the analysis of the minimized structure's PES, a hypothesis was proposed, that a residue in the LL was stabilizing the positive charge of the protonated His223, thus generating low values for the System's Total Energy. Through visual inspection the

residue was Aspartate224 that was stabilizing His 223 by the negative charge conferred by the COO^- group. This observation implies that the addition of Asp224 to the HL would be a more suitable approach.

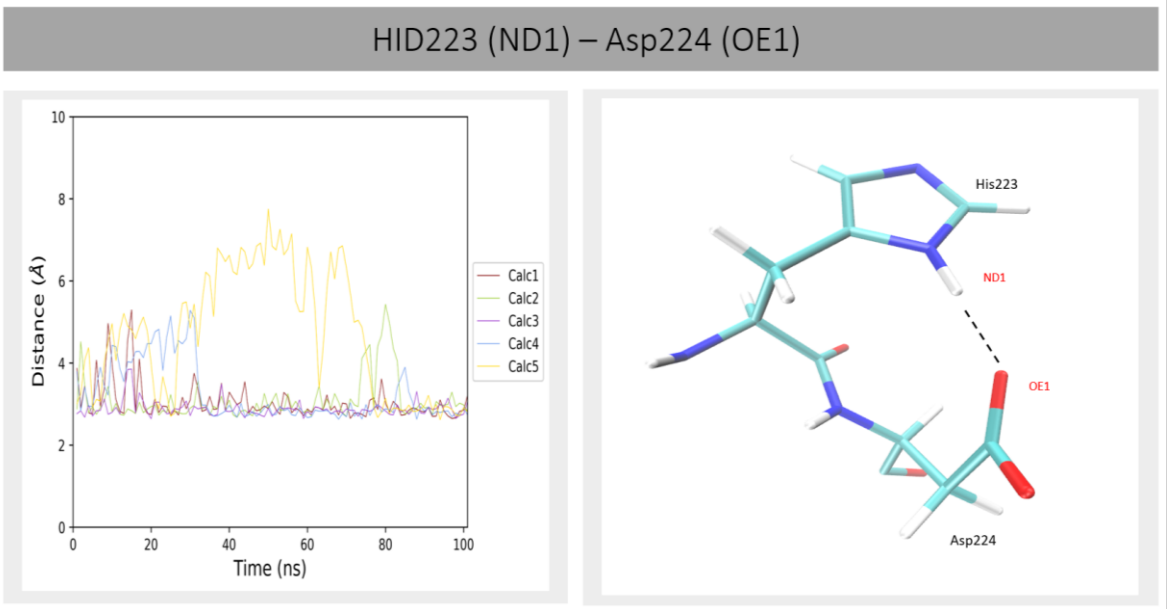


Figure 54 - (Left) - Distance between HID223 (ND1) and Asp224 (OD1) for all the simulations, for the oxidized system. (Right) – Tridimensional representation of the HID223 establishing a hydrogen bond with Asp224.

A conclusion was drawn that the hydrogen bond between Aspartate 209 and His 223 was frequent throughout the MD.

The analysis of the first PES allowed to confirm the first step for the semi-reduction reaction for LAAO, with a barrier of 0.34 kcal/mol.

The future perspectives for the study of this enzymatic mechanism would first pass through a better energetic characterization of the intermediates for the two initial steps. This characterization would be achieved first though a geometry optimization under EE of the intermediates, followed by the calculation of vibrational frequencies for the reaction coordinate. Furthermore, linear transit scans in the reverse direction should be performed.

Finally, for the study of the semi-oxidation step, a new system should be constructed where a reactive 0_2 close to the N5 of FADH and the side chain of Lys345 should be modelled, in order to mimic the proposed catalytic triad [35]. For the time being, only

parametrization of the system comprised of the protein plus the semi-reduced FAD molecule is assembled.

The searching for novel LAAO inhibitors

- Structure-Based Virtual Screening

SBVS is a computational technique that aids for the discovery of new drugs based on the target tridimensional structure. This structural information allows the VS software to evaluate the protein–ligand recognition. VS is not more that the application of massive molecular docking calculations, at a fast pace, which run simultaneously, in parallel, and only exploit the elementary possibilities for ligand binding.

In order to speed up the discovery of putative new drug candidates that block the pathological effects of *viperidae* LAAO, we have selected a diverse chemical library of molecules, *circa* 9700. These molecules approved at least for phase I of the drug discovery pipeline and were made available by the Drug Bank database. This strategy aimed for the finding of good LAAO inhibitors with known toxicity.

- Selection of the VS pipeline and scoring function

The VS calculations were carried out using the GOLD software. To date, there are no known effective inhibitors that block the LAAO catalysis. Thus, it was impossible to test the efficiency of several scoring functions implemented in the docking software's and its calibration using decoys. However, we have considered the ASP scoring function, implemented in GOLD, as the most suitable to perform the VS. This scoring function was considered preferential, over the GoldScore and ChemScore functions, because it was the only one that could reproduce the substrate in a favorable orientation for the catalysis.

After performing the VS using the GOLD software and the ASP function, we looked at the scoring values attributed to the compounds and selected the top 250 for posterior filtering, based on their chemical diversity, which were then visually inspected for validation.

- Clustering Analysis

In order to reduce the 250 compounds, equal likely to bind to the LAAO, to 50 (or 100) compounds to be tested *in vitro*, we have clustered the results according to their chemistry dissimilarity using the *Chemaxon* package [124].

To better understand the chemical similarity (or dissimilarity) of the compounds we generated two types of fingerprints: 2D pharmacophore fingerprints (PF) and Extended connectivity pharmacophore fingerprints (ECPF).

At this point, the treatment of the 250 compounds list was conducted using two clustering methods. In a first one, we used a hierarchical algorithm, and secondly a non-hierarchical algorithm to generate the desired clusters.

Non-hierarchical approach

Using the Jarvis-Patrick clustering method we exploited the dissimilarity of the compounds using the Tanimoto index (t) and their similarity to be considered neighbor molecules (c). We have tested different metrics for the combination of both parameters (t and c) in order to:

(1) Have the 50 molecules groups similar populated.
(2) Avoid clusters with a single molecule, which may be active against LAAO, but are more probable to fail.

Hierarchical approach

The next stage saw the employment of Ward's hierarchical clustering algorithm, with the metrics to tune for this method being the number of generated clusters (*c*), attending to their chemical similarity. Here, we asked for 50 and 100 clusters (number that could be reduced to 50 upon a visual validation). Indeed, the results pointed for more populated clusters, in opposition to non-hierarchical clustering, hinting for a more appropriate method.

When c=50 was given as an input parameter, most of the generated clusters had a population of 4 compounds, with an observed minimum and maximum cluster population of 3 and 10, respectively. Regarding c=100, the most observed cluster population was

2, with 1 and 6 being the maximum and minimum population values. Because the number of molecules in each cluster is reduced 2-fold when increasing from 50 to 100, the results showed that we were not gaining any information when folding the number of clusters. And therefore, 50 clusters were already adequate to explain the top 250 compounds diversity.

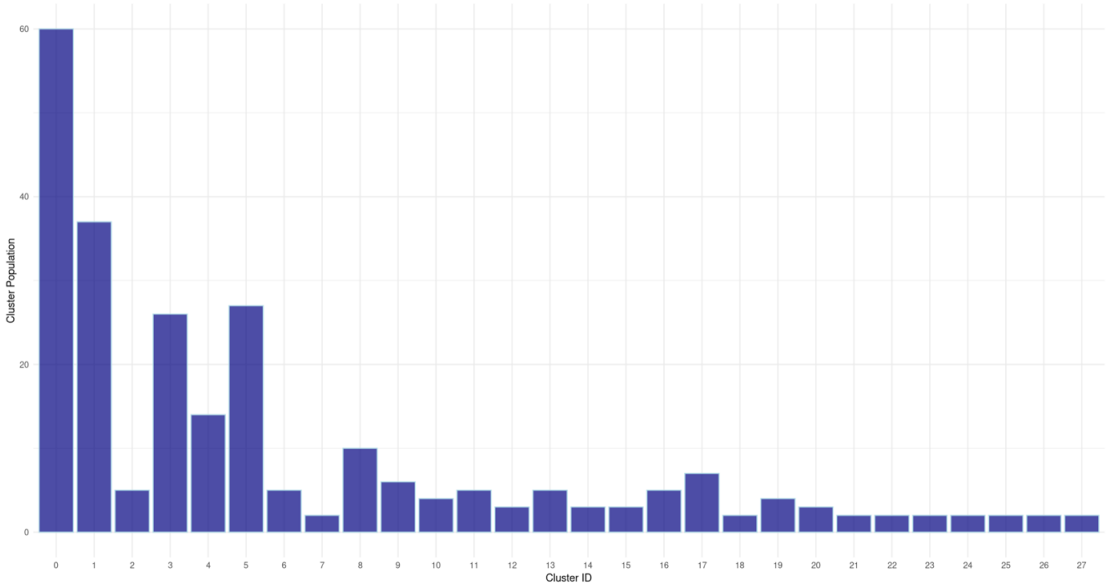


Figure 55 - Number of clusters generated for the top250 ranked compounds, when using a non-hierarchical clustering algorithm, with Tanimoto and Dissimilarity parameters of 0.6 and 0.8, respectively.

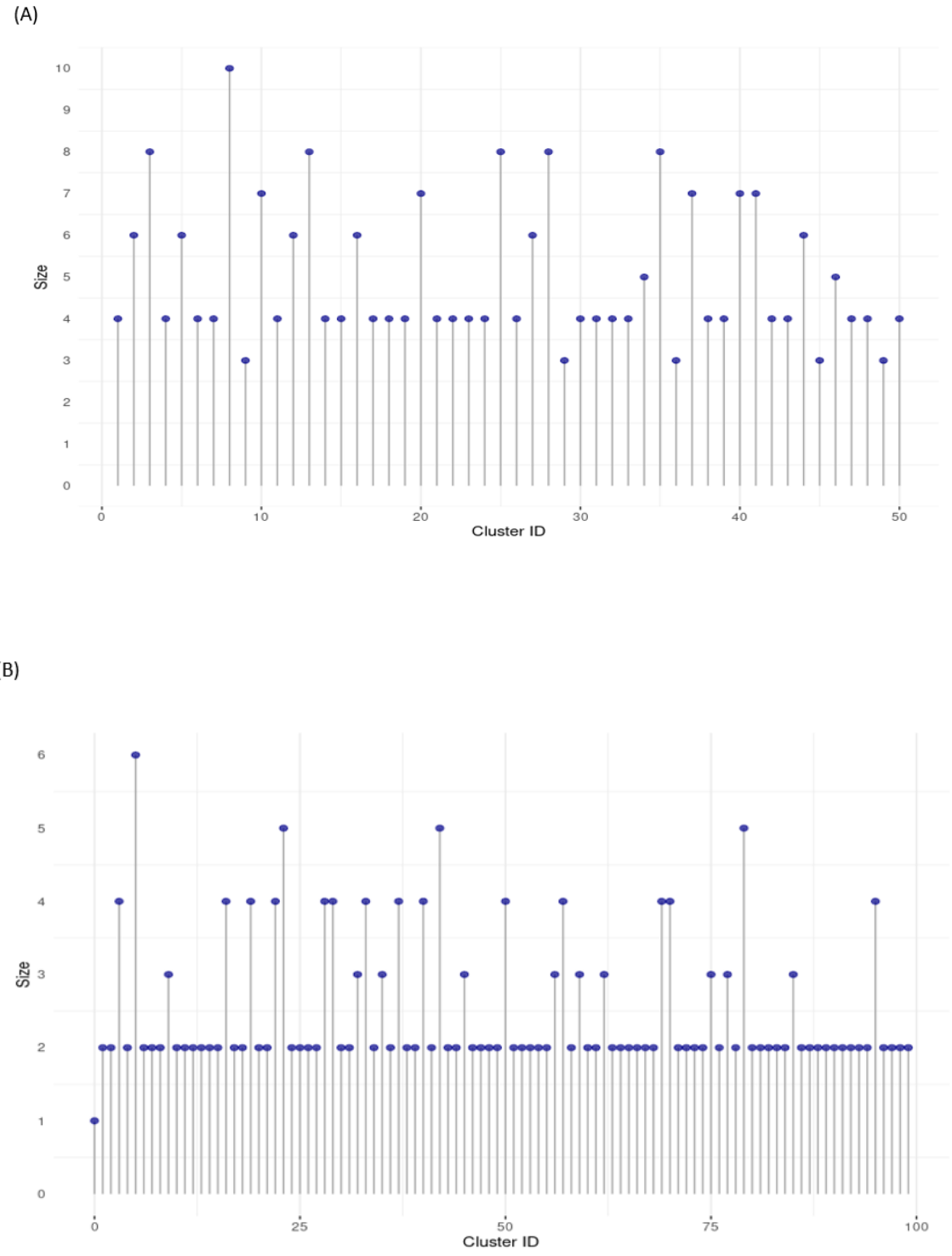


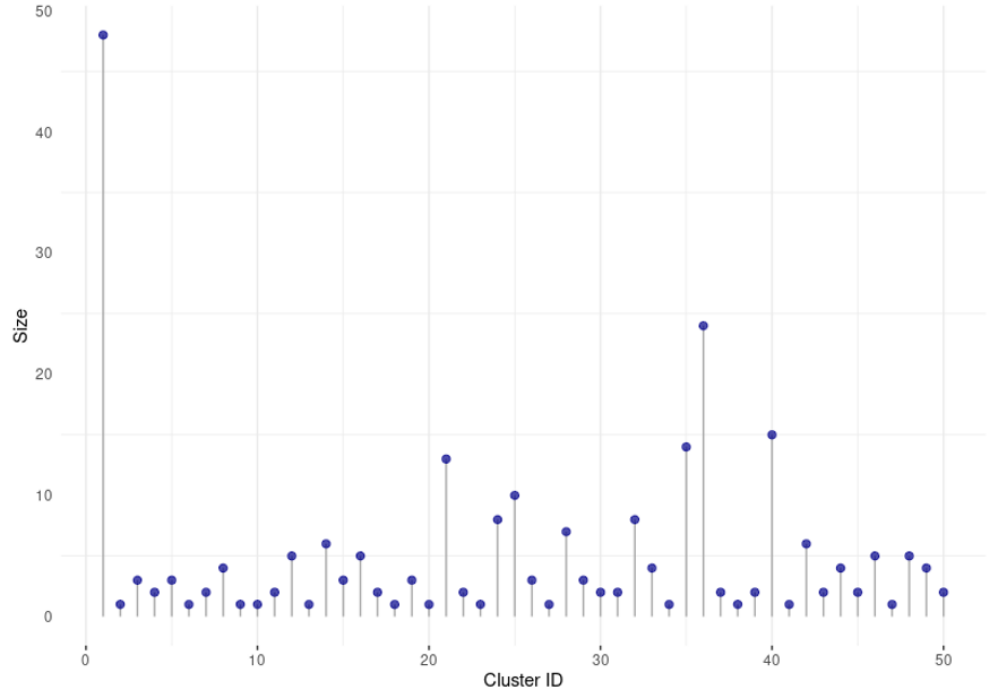
Figure 56 - Clusters' population, when using Ward's hierarchical clustering algorithm. (A) – 50 generated clusters. (B) – 100 generated clusters.

Results indicated that for non-hierarchical clustering algorithm, PF were to be used rather than ECFP, mainly given the poor number of clusters generated, regardless of the parameters used. Indeed, PF produced more satisfactory outcome, where the optimal parameters for the metrics were t=0.6 and c=0.8.

However, in both studied cases using the non-hierarchical algorithm (and PF and ECFP chemical descriptors), the results proven to be non-satisfactory, given the high proportion of singletons generated, i.e., the number of clusters composed solely of a compound.

Finally, we exploited the Diverse Set Selection Clustering Algorithm as implemented in *Chemaxon*. This non-hierarchical clustering splits the molecules into a number of clusters imposed by the user, without considering chemical fingerprints, but facilitating the sampling of the 250 compounds to obtain the most diverse compounds that span the entire chemical space. Diverse Set clustering generated better results when 50 clusters were generated, given that 100 clusters generated more singletons. Once again, the results pointed for similar outcomes, indicating a somewhat condensed chemical diversity.

(A)



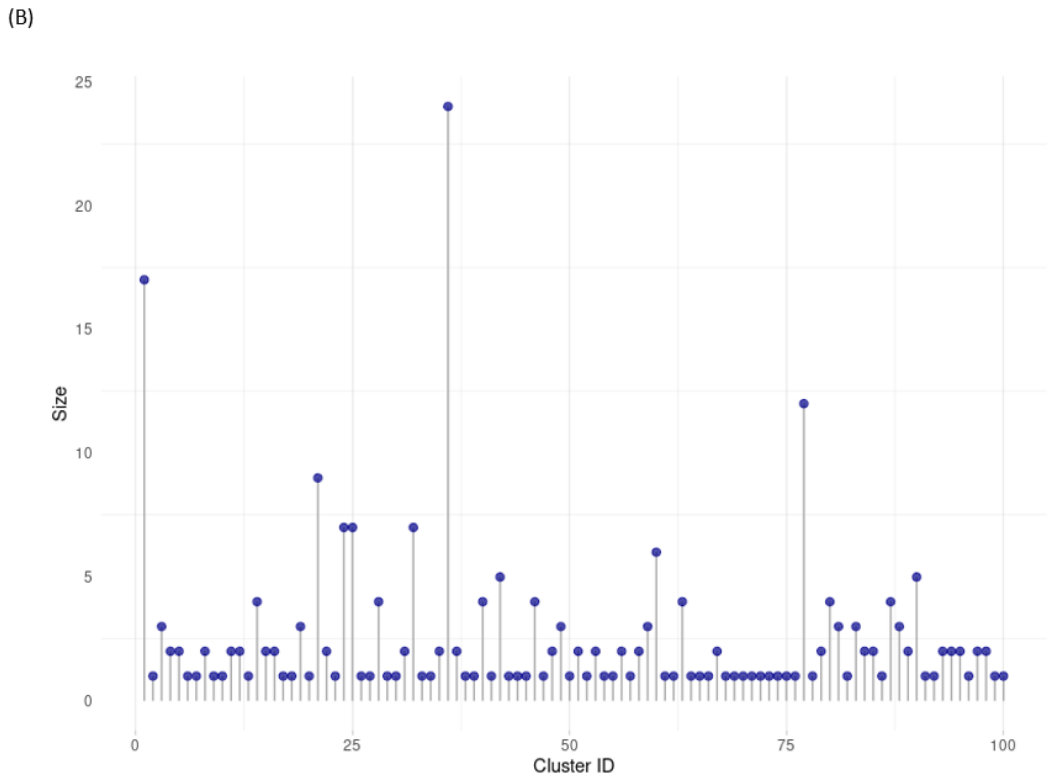


Figure 57 - Number of clusters generated with respective populations, when using a Diverse Set clustering algorithm. (A) – 50 generated clusters. (B) – 100 generated clusters.

Finally, we have compared the chemical diversity of two generated 50 clusters datasets: the hierarchical cluster — Ward — and the non-hierarchical cluster — Diverse set clustering. We have compared each cluster representative compound, using *Chemaxon's compr*. This tool calculates the average, maximum and minimum dissimilarity between molecules, which belong to different datasets (or over compounds of the same database), computing their Tanimoto index. In average, the chemical difference between the Ward's and Diverse Set representatives was 0.850 (min=0.017 and max=0.970). The results indicated that the two metrics lead to completely different compound representatives. This supports the claim that we must select compounds that were identified using the Ward's clustering and other compounds should be selected based on Diverse Set clustering. These results were also supported by means of Principal Component Analysis (see Fig. 57).

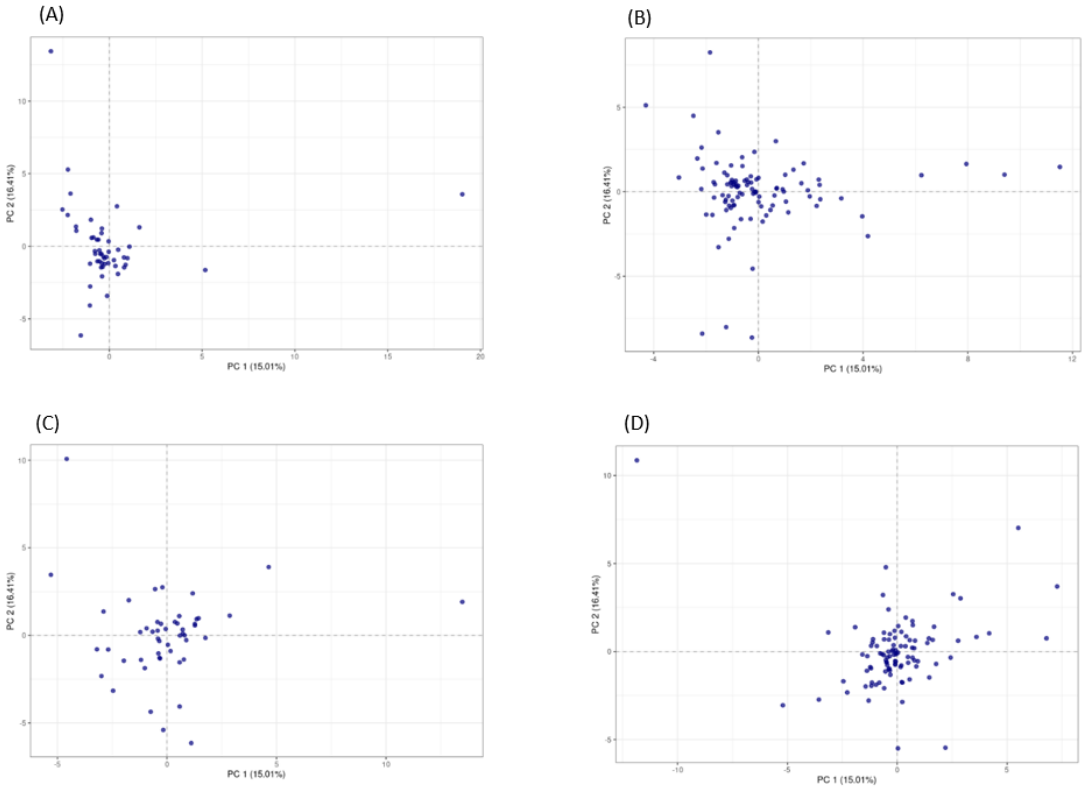


Figure 58 - PCAs for the different clustering algorithm applied - Ward (left) and Diverse Set (right). (A) & (C) - 50 generated clusters. (B) & (D) - 100 generated clusters.

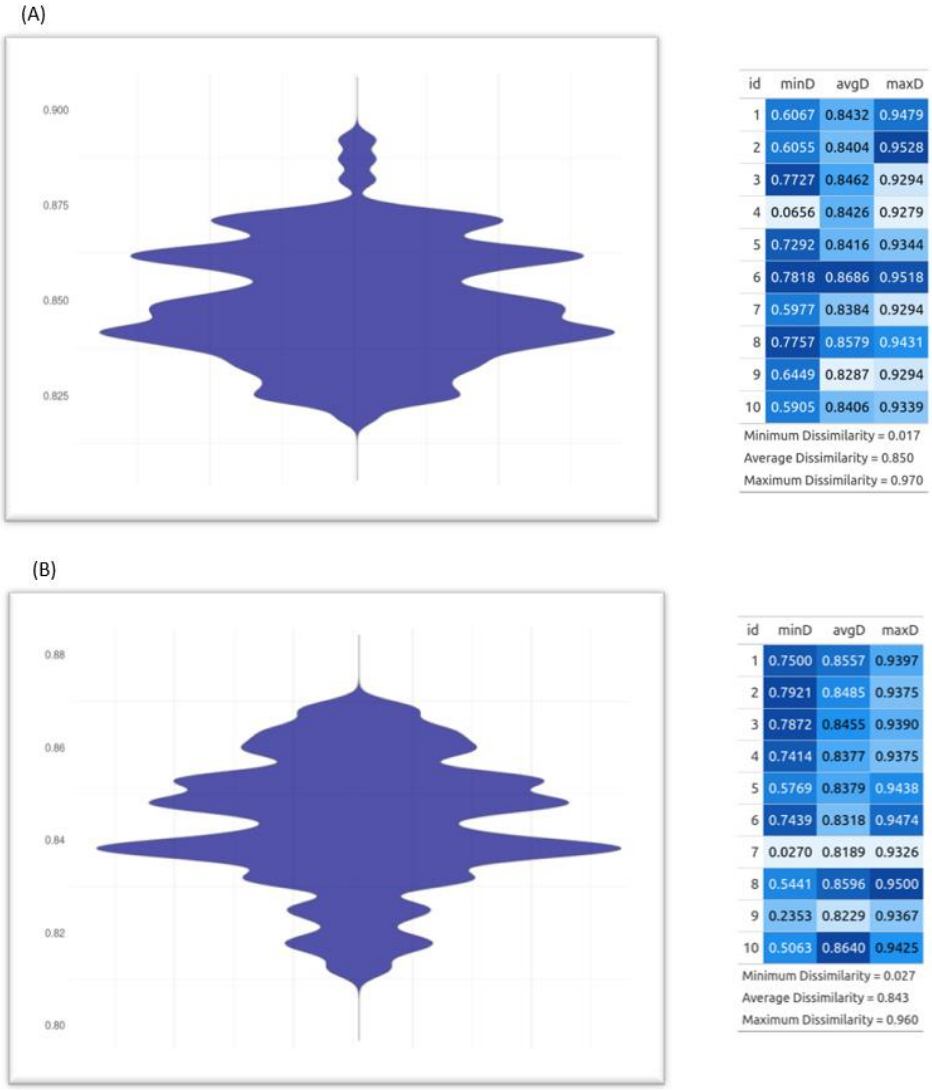


Figure 59 - Chemical diversity comparison between two sets of data, displayed over a violin plot and an example table highlighting minimum, average and maximum calculated dissimilarity values (A) - when 50 clusters were generated. (B) - when 100 clusters were generated.

- Favorable Interactions

The other employed strategy for selection of the compounds was a visual inspection. We selected only compounds which made interactions (H-bonds and pi-stacking interactions) that are considered key for the enzyme's catalytic activity. Indeed, literature points for interactions with key residues important for the catalytic cycle, either for the initial accommodation of the substrate, or for posterior stabilization. Examples of those residues are the His223, Arg90, Lys345 and/or Gly464 [35]. With that information in mind, it becomes reasonable to assume that compounds that establishes

bonds with these residues can mitigate or impede the catalytic cycle to occur, by acting as competitive inhibitors [127].

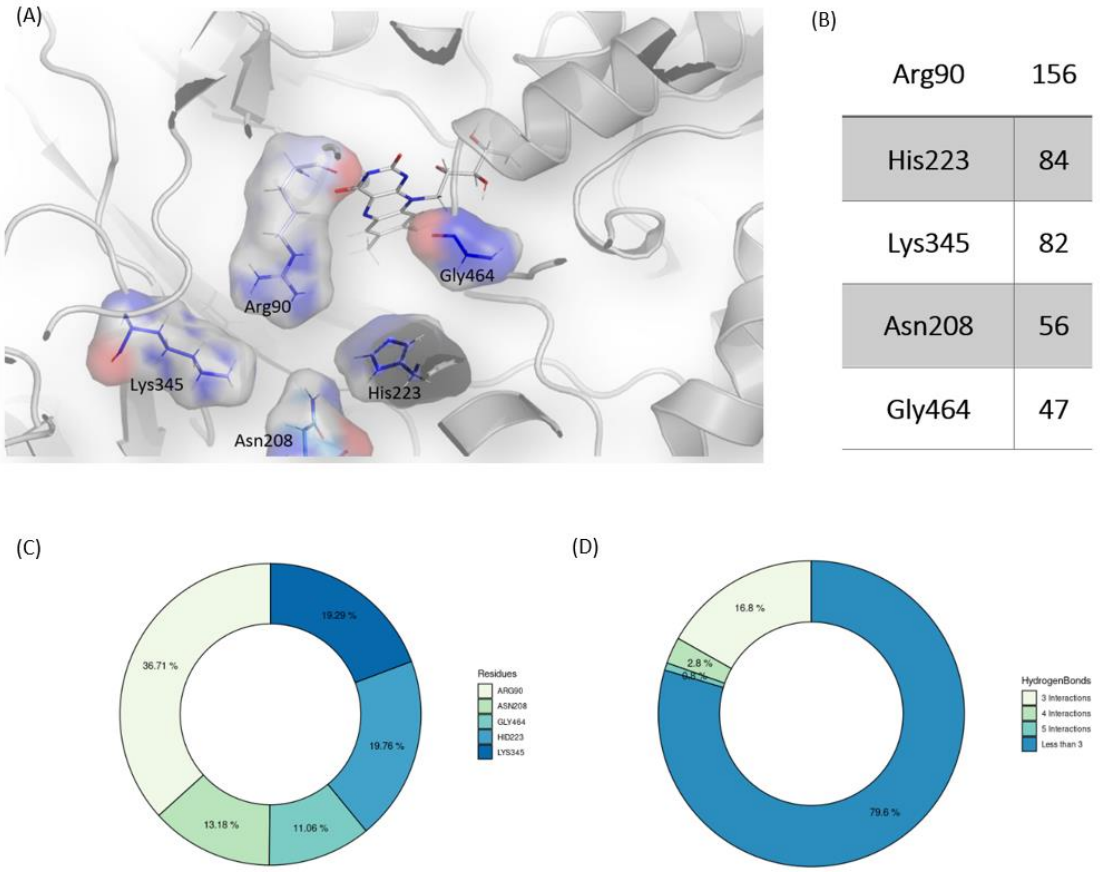


Figure 60 - (A) -Representation of the protein with highlight on the more observed hydrogen bonds, represented in surface. Image made with Pymol software. (B) - Table with the absolute frequency of the top 5 more observed residues. (C) - Relative frequency the top 5 most observed residues (D) – Relative frequency of the number of “relevant” interactions made by the docked ligands.

We started by quantifying the number of H-bonds, which are very strong interactions between the protein and a ligand, with high costly to break, and the residues evolved in this kind of electrostatic interactions. The hydrogen bonds made within the protein were inspected using *Ligplot+* [128].

As can be seen in Fig. 60, H-bonds between the top 250 and the Arg90 or/and Lys345 are present in more than 56% of the solutions analyzed. Other important H-bonds are

stabilized with the His223 (~20%), the backbone of Gly464 (11%) and Asn208(13%). Moreover, ~80% compounds formed at least 1 or 2 H-bonds with the protein.

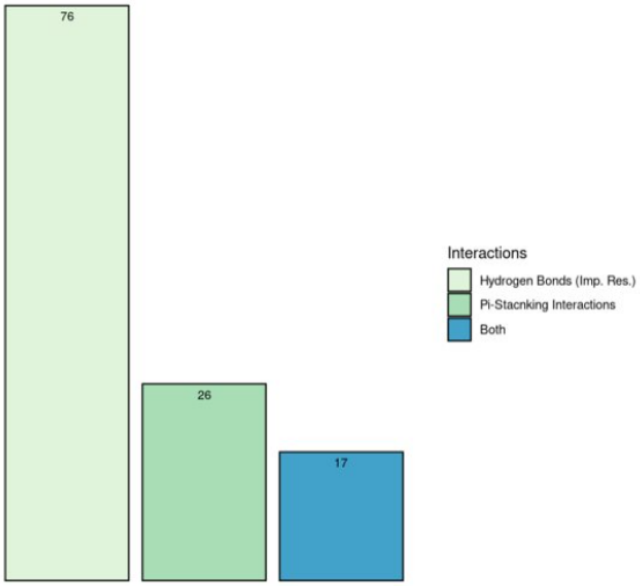


Figure 61 - Absolute frequency of the residues with 3+ frequent interactions (Hydrogen Bonds (Imp.Res)), displaying Pi-stacking interactions Pi-Stacking interactions), and observed Pi-Stacking interactions and residues with 3+ frequent hydrogen bond interactions (Both)

Thereafter, the pi-stacking interactions were analyzed using *PyViewer* plugin of *Pymol.* This information was cross-referenced with the previous one, therefore indicating the more promising residues to be tested *in vivo,* by establishing relevant interactions, while displaying increased chemical stability.

- Cross-Referencing

Finally, the results arising from this step were confronted with the ones coming from the previous steps. This allowed the selection of compounds to be further tested *in vivo*. Indeed, by the intersection of compounds arising from the clustering analysis step and the favorable interactions stage, a greater confidence in the selection of the drugs to be tested was achieved. This is due to the chemical diversity of the selected compounds and the pattern of interactions established with the protein, increasing the probability of retarding or even stopping the catalytic mechanism.

This final cross-referencing was made regarding the representative structures of each generated cluster. The representatives were used as the database, while the compounds

arising from the previous step (Important Residues) were used as a query. Regarding Ward´s clustering algorithm, when 50 clusters were generated, 7 representatives were selected. When 50 clusters were generated from the diverse set clustering with 50 clusters generated 4 representative compounds were deemed most promising to test, while with 100 clusters, the selected representatives were 10. In total, we selected 11 compounds with similar binding score (Figure 62).

	SCORE
WARD - 50C	
Compound 1	61.772
Compound 2	67.638
Compound 3	61.044
Compound 4	55.626
Compound 5	64.214
Compound 6	60.334
Compound 7	59.238
DIVERSE SET - 50C	
Compound 8	63.630
Compound 9	65.118
Compound 10	53.610
Compound 11	56.428

Figure 62 - Final Compound selection for the cross-referencing step, for Ward Clustering with 50 clusters and Diverse Setting with 50 clusters.

Conclusions

This work was conducted with the aim to obtain a deeper knowledge on the dynamics and catalytic mechanism of LAAO from *B.atrox*, and to identify potential new inhibitors against its enzymatic activity.

Making use of Homology Modeling techniques and using as template X-ray structure of the LAAO from *C. rhodostoma*, we were able to construct a complete 3D model for this protein. The superimposition of the obtained models against the initial structure of the protein (PDB ID. 5TS5) showed RMSD values lower than 1 Å, noting virtually similar tridimensional structures.

Thereafter, a Molecular Docking studies predicted the binding of the substrate to the enzyme. A favorable pose for the reaction to occur was obtained, similar to the one obtained by Moustafa et al. At this step, we could construct a complex of the protein and the substrate (L-Phe) in presence of the FAD cofactor.

To shed light on the dynamics of the two major states of the enzyme, i.e, the two states that mimic the two half reactions of the mechanism, we setup two independent systems that were carefully parametrized. In a first system, the enzyme was complexed with a zwitterionic substrate, in presence of a fully oxidized FAD molecule. The second system was assembled with the protein, the L-Phe in the iminoacid form and a semi-reduced FADH molecule.

Results of Molecular Dynamics calculations confirmed the importance of the residues Arg90, Tyr372 and Gly464 and the FAD molecule to trap the substrate in the catalytic site. Additionally, the importance of these residues to the stability of the substrate was highlighted. Furthermore, to accommodate the iminoacid, when in presence of the FADH, the His223 and Arg322 side chain experienced rotations, as proven by the dihedral angles shifts. These results are in line with the proposal of Ullah et al., confirming that the oxidized and the semi-reduced systems undergo conformational changes to better accommodate the charges and the intermediate. Finally, a distancing between the His223 and L-Phe was observed in the reduced system.

Consequently, the study of the mechanistic mechanism for the protein was initiated over a QM/MM protocol. The QM layer was composed by 115 atoms, comprising the isoalloxazine ring of the FAD molecule, the imidazole ring of His223 and the L-Phe. Other residues (completed or not) included in the layer were Phe 227, Tyr 372, and a water molecule (residue 501). Overall, this layer consisted in 115 atoms, with a net charge of

+1. The QM layer was treated under a higher theoretical level, the DFT level with a hybrid functional – B3LYP, in a combination with the 6-31G(d) basis set. The remainder system was treated in MM embedding applying the Amber ff14SB force filed. In the first step of the catalysis, occurs the transference of the NH_3^+-L-Phe to the His 223, making the last double protonated. The estimated energy barrier for this step was 0.34 kcal/mol, noting a spontaneous reaction. As future work, more QM/MM steps need to be conducted, such as Potential Energy Scan between coordinates of proposed steps for this reaction. Moreover, geometry optimizations, zero-point, thermal and entropic corrections would allow to obtain free energies for all the obtained species. This obtained free energies could be further confirmed through new calculation with higher theoretical levels, or larger basis sets.

Finally, a drug screening protocol was used, a SBVS, namely though a drug repurposing study. Here, a pre-compiled chemical library, made of *circa* 10.000 drugs, was screened against the target protein and the top250 compounds clustered according to their chemical diversity. For each cluster generated, it was retrieved a representative compound. Simultaneously, we tried to understand the amino acids that are more likely to contribute for the drug binding by analyzing a pattern of H-bonds and pi-stacking interactions among the 250 obtained compounds.

Finally, the representative compounds that obey to the most frequent interactions were selected to be tested *in vitro*. At this point, 11 compounds were considered to have higher probability to be potential inhibitors candidates against the ***Bothrops atrox's*** LAAO.

References and Bibliography

1. Feliciano, P.R., Nonato, M.C. 5TS5: Crystal structure of L-amino acid oxidase from Bothrops atrox. 2016; Available from: https://www.rcsb.org/structure/5TS5.
2. BioRender. Available from: https://biorender.com/.
3. Sunagar, K. and Y. Moran, The Rise and Fall of an Evolutionary Innovation: Contrasting Strategies of Venom Evolution in Ancient and Young Animals. PLoS Genet, 2015. **11**(10): p. e1005596.
4. Sunagar, K., et al., Deadly Innovations: Unraveling the Molecular Evolution of Animal Venoms, in Venom Genomics and Proteomics, P. Gopalakrishnakone and J.J. Calvete, Editors. 2016, Springer Netherlands: Dordrecht. p. 1-27.
5. Gong, E., et al., The birdlike raptor Sinornithosaurus was venomous. Proceedings of the National Academy of Sciences, 2010. **107**(2): p. 766-768.
6. Fox, R.C. and C.S. Scott, First evidence of a venom delivery apparatus in extinct mammals. Nature, 2005. **435**(7045): p. 1091-3.
7. Jay S. Keystone, D.O.F., Phyllis E. Kozarsky, Bradley A. Connor, Hans D. Nothdurft, Travel Medicine (Third Edition). 2013. 413-424.
8. Williams, D., et al., The Global Snake Bite Initiative: an antidote for snake bite. Lancet, 2010. **375**(9708): p. 89-91.
9. Fry, B.G., From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res, 2005. **15**(3): p. 403-20.
10. Fry, B.G., et al., Evolution and diversification of the Toxicofera reptile venom system. J Proteomics, 2009. **72**(2): p. 127-36.
11. Yang, Z., et al., Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 2000. **155**(1): p. 431-49.
12. Uetz, P., Freed, P, Aguilar, R. & Hošek, J., The Reptile Database. 2022.
13. Tasoulis, T. and G.K. Isbister, A Review and Database of Snake Venom Proteomes. Toxins (Basel), 2017. **9**(9).
14. Hui Wen, F., et al., Snakebites and scorpion stings in the Brazilian Amazon: identifying research priorities for a largely neglected problem. PLoS Negl Trop Dis, 2015. **9**(5): p. e0003701.
15. Alcântara, J.A., et al., Stepping into a dangerous quagmire: Macroecological determinants of Bothrops envenomings, Brazilian Amazon. PLoS One, 2018. **13**(12): p. e0208532.

16. Common Lancehead (Bothrops atrox). 09.08]; Available from: https://www.reptilesofecuador.com/bothrops_atrox.html#46.
17. Zavaleta A, S.M., Ofidismo: envenenamiento por mordedura de serpientes. Emergencias en medicina interna., ed. L.-B.R. Martínez-Villaverde JR, Vidal-Neira L, Losno-García R 1996.
18. Integrated Taxonomic Information System. 10.08]; Available from: https://www.itis.gov/.
19. Saldarriaga MM, O.R., Núñez V, Toro MF, Díaz A, Gutiérrez JM, Ontogenetic variability of Bothrops atrox and Bothrops asper snake venoms from Colombia. Vol. 42. 2003, Toxicon. 405-411.
20. Kuch U, M.D., Gutiérrez JM, Freire A Biochemical and biological characterization of the Ecuadorian pitvipers (genera Bothriechis, Bothriopsis, Bothrops, and Lachesis). 1996(34): p. 714-717.
21. DA, W., Snakebites in Central and South America: epidemiology, clinical features, and clinical management. The Venomous reptiles of the Western Hemisphere, 2004: p. 709-761.
22. Sachett JAG, V.F., Alcantara JA, Cubas-Vega N, Montenegro CS, da Silva IM, de Souza TG, Santana MF, Ferreira LCL, Monteiro WM Bothrops atrox snakebite: how a bad decision may lead to a chronic disability: a case report. Wilderness & Environmental Medicine, 2020.
23. Santos Barreto GN, d.O.S., Dos Anjos IV, Chalkidis HM, Mourao RH, Moura-da-Silva AM, Sano-Martins IS, Goncalves LR Experimental Bothrops atrox envenomation: efficacy of antivenom therapy and the combination of Bothrops antivenom with dexamethasone. PLoS Neglected Tropical Diseases 2017(11).
24. Gutiérrez, J.M.e.a., Snakebite envenoming. Nat. Rev. Dis.Primers 2017. **3**(17063).
25. Pardal PPO, S.S., Monteiro MRCC, Fan HW, Cardoso JLC, França FOS, et al., Clinical trial of two antivenoms for the treatment of Bothrops and Lachesis bites in the north eastern Amazon region of Brazil. Trans R Soc Trop Med Hyg, 2004(98): p. 28-42.
26. Álvaro Segura, M.H., Mariángela Vargas, Mauren Villalta, Alfredo Uscanga-Reynell, Guillermo León, JoséMaría Gutiérrez, Preclinical efficacy against toxic activities of medically relevant Bothrops sp. (Serpentes: Viperidae) snake venoms by a polyspecific antivenom produced in Mexico. Revista de Biología Tropical, 2017. **65**(1).

27. da-Silva, W., Guidolin, R., Raw, I., Higashi, H., Caricati, C., Morais, J., al, e., Crossreactivity of Horse Monovalent Antivenoms to Venoms of Ten bothrops species. Memórias do Instituto Butantan. 1999.
28. Monteiro, W.M., et al., Bothrops atrox, the most important snake involved in human envenomings in the amazon: How venomics contributes to the knowledge of snake biology and clinical toxinology. Toxicon X, 2020. **6**: p. 100037.
29. Moretto Del-Rei, T.H., et al., Functional variability of Bothrops atrox venoms from three distinct areas across the Brazilian Amazon and consequences for human envenomings. Toxicon, 2019. **164**: p. 61-70.
30. Global Snakebite Initiative. Available from: https://www.snakebiteinitiative.org.
31. Dennis E. Vance, J.E.V., Biochemistry of Lipids, Lipoproteins and Membranes Fifth Edition ed. 2008.
32. MARITZ, E.A.Z.a.A., Helv. Chim. Acta, 1944(27): p. 1888-1903.
33. Izidoro, L.F., et al., Snake venom L-amino acid oxidases: trends in pharmacology and biochemistry. Biomed Res Int, 2014. **2014**: p. 196754.
34. Page, D.S. and R.L. Vanetten, L-amino acid oxidase. II. Deuterium isotope effects and the action mechanism for the reduction of L-amino acid oxidase by L-leucine. Biochim Biophys Acta, 1971. **227**(1): p. 16-31.
35. Ullah, A., Structure-Function Studies and Mechanism of Action of Snake Venom L-Amino Acid Oxidases. Front Pharmacol, 2020. **11**: p. 110.
36. Ande, S.R., et al., Induction of apoptosis in yeast by L-amino acid oxidase from the Malayan pit viper Calloselasma rhodostoma. Yeast, 2008. **25**(5): p. 349-57.
37. Takatsuka, H., et al., Molecular characterization of L-amino acid oxidase from Agkistrodon halys blomhoffii with special reference to platelet aggregation. Biochim Biophys Acta, 2001. **1544**(1-2): p. 267-77.
38. Georgieva, D., et al., The structure of a native l-amino acid oxidase, the major component of the Vipera ammodytes ammodytes venomic, reveals dynamic active site and quaternary structure stabilization by divalent ions. Mol Biosyst, 2011. **7**(2): p. 379-84.
39. Pawelek, P.D., et al., The structure of L-amino acid oxidase reveals the substrate trajectory into an enantiomerically conserved active site. Embo j, 2000. **19**(16): p. 4204-15.
40. Moustafa, I.M., et al., Crystal structure of LAAO from Calloselasma rhodostoma with an L-phenylalanine substrate: insights into structure and mechanism. J Mol Biol, 2006. **364**(5): p. 991-1002.

41. Massey, V. and B. Curti, On the reaction mechanism of Crotalus adamanteus L-amino acid oxidase. J Biol Chem, 1967. **242**(6): p. 1259-64.
42. Fitzpatrick, P.F., Carbanion versus hydride transfer mechanisms in flavoprotein-catalyzed dehydrogenations. Bioorg Chem, 2004. **32**(3): p. 125-39.
43. Zhang, H., et al., Purification, partial characterization, crystallization and structural determination of AHP-LAAO, a novel L-amino-acid oxidase with cell apoptosis-inducing activity from Agkistrodon halys pallas venom. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 5): p. 974-7.
44. Lionta, E., et al., Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem, 2014. **14**(16): p. 1923-38.
45. Lodola, A. and M. De Vivo, The increasing role of QM/MM in drug discovery. Adv Protein Chem Struct Biol, 2012. **87**: p. 337-62.
46. Lonsdale, R., J.N. Harvey, and A.J. Mulholland, A practical guide to modelling enzyme-catalysed reactions. Chem Soc Rev, 2012. **41**(8): p. 3025-38.
47. Baker, D., A surprising simplicity to protein folding. Nature, 2000. **405**(6782): p. 39-42.
48. Jooyoung Lee, P.F., Yang Zhang, Ab initio protein structure prediction, Protein Structure to Function with Bioinformatics, ed. D.J. Rigden. 2017, London: Springer.
49. Martí-Renom, M.A., et al., Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct, 2000. **29**: p. 291-325.
50. Fiser, A. and A. Sali, Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol, 2003. **374**: p. 461-91.
51. Jumper, J., et al., Highly accurate protein structure prediction with AlphaFold. Nature, 2021. **596**(7873): p. 583-589.
52. Hooft, R.W., et al., Errors in protein structures. Nature, 1996. **381**(6580): p. 272.
53. International Journal of Biochemistry & Cell Biology, 2018. **104**: p. 94-113.
54. Kuntz, I.D., et al., A geometric approach to macromolecule-ligand interactions. J Mol Biol, 1982. **161**(2): p. 269-88.
55. Teague, S.J., Implications of protein flexibility for drug discovery. Nat Rev Drug Discov, 2003. **2**(7): p. 527-41.
56. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, Protein-ligand docking: current status and future challenges. Proteins, 2006. **65**(1): p. 15-26.
57. Ingo Muegge, M.R., Small molecule docking and scoring. 2001: Rev Comput Chem

58. Böhm, H.J., The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des, 1992. **6**(1): p. 61-78.
59. Rarey, M., et al., A fast flexible docking method using an incremental construction algorithm. J Mol Biol, 1996. **261**(3): p. 470-89.
60. Ewing, T.J.A. and I.D. Kuntz, Critical evaluation of search algorithms for automated molecular docking and database screening. Journal of Computational Chemistry, 1997. **18**(9): p. 1175-1189.
61. Mizutani, M.Y., N. Tomioka, and A. Itai, Rational Automatic Search Method for Stable Docking Models of Protein and Ligand. Journal of Molecular Biology, 1994. **243**(2): p. 310-326.
62. Welch, W., J. Ruppert, and A.N. Jain, Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. Chem Biol, 1996. **3**(6): p. 449-62.
63. Miller, M.D., et al., FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. J Comput Aided Mol Des, 1994. **8**(2): p. 153-74.
64. Taylor, R.D., P.J. Jewsbury, and J.W. Essex, A review of protein-small molecule docking methods. J Comput Aided Mol Des, 2002. **16**(3): p. 151-66.
65. Trosset, J.-Y. and H.A. Scheraga, Prodock: Software package for protein modeling and docking. Journal of Computational Chemistry, 1999. **20**(4): p. 412-427.
66. Abagyan, R., M. Totrov, and D. Kuznetsov, ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. Journal of Computational Chemistry, 1994. **15**(5): p. 488-506.
67. Liu, M. and S. Wang, MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. J Comput Aided Mol Des, 1999. **13**(5): p. 435-51.
68. Hart, T.N. and R.J. Read, A multiple-start Monte Carlo docking method. Proteins, 1992. **13**(3): p. 206-22.
69. McMartin, C. and R.S. Bohacek, QXP: powerful, rapid computer algorithms for structure-based drug design. J Comput Aided Mol Des, 1997. **11**(4): p. 333-44.
70. Morris, G.M., et al., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
71. Clark, K.P. and Ajay, Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. Journal of Computational Chemistry, 1995. **16**(10): p. 1210-1226.

72. Taylor, J.S. and R.M. Burnett, DARWIN: a program for docking flexible molecules. Proteins, 2000. **41**(2): p. 173-91.
73. Verdonk, M.L., et al., Improved protein-ligand docking using GOLD. Proteins, 2003. **52**(4): p. 609-23.
74. Baxter, C.A., et al., Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins, 1998. **33**(3): p. 367-82.
75. Murray, C.W., C.A. Baxter, and A.D. Frenkel, The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des, 1999. **13**(6): p. 547-62.
76. Kitchen, D.B., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov, 2004. **3**(11): p. 935-49.
77. Bissantz, C., G. Folkers, and D. Rognan, Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem, 2000. **43**(25): p. 4759-67.
78. Kramer, B., M. Rarey, and T. Lengauer, Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins, 1999. **37**(2): p. 228-41.
79. Böhm, H.J., Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. J Comput Aided Mol Des, 1998. **12**(4): p. 309-23.
80. Böhm, H.J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des, 1994. **8**(3): p. 243-56.
81. Eldridge, M.D., et al., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des, 1997. **11**(5): p. 425-45.
82. Carlson, H.A., Protein flexibility and drug design: how to hit a moving target. Curr Opin Chem Biol, 2002. **6**(4): p. 447-52.
83. Muegge, I. and Y.C. Martin, A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem, 1999. **42**(5): p. 791-804.
84. Muegge, I., Effect of ligand volume correction on PMF scoring. Journal of Computational Chemistry, 2001. **22**(4): p. 418-425.
85. Muegge, I., A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. Perspectives in Drug Discovery and Design, 2000. **20**(1): p. 99-114.

86. Gohlke, H., M. Hendlich, and G. Klebe, Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol, 2000. **295**(2): p. 337-56.
87. Wang, R., L. Lai, and S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des, 2002. **16**(1): p. 11-26.
88. Jones, G., et al., Development and validation of a genetic algorithm for flexible docking. J Mol Biol, 1997. **267**(3): p. 727-48.
89. Salomon-Ferrer, R., D.A. Case, and R.C. Walker, An overview of the Amber biomolecular simulation package. WIREs Computational Molecular Science, 2013. **3**(2): p. 198-210.
90. Brooks, B.R., et al., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry, 1983. **4**.
91. A D J MacKerell, B.B., C L I Brooks, L Nilsson, B Roux, Y Won, M Karplus, CHARMM: the energy function and its parameterization with an overview of the program, in The Encyclopedia of Computational Chemistry. 1998.
92. van Gunsteren, W.F. and H.J.C. Berendsen, Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. Angewandte Chemie International Edition in English, 1990. **29**(9): p. 992-1023.
93. 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. Journal of the American Chemical Society, 1988. **110**(6): p. 1657-1666.
94. Gordon, J.C., et al., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W368-71.
95. PropKa Online. Available from: https://www.ddl.unimi.it/vegaol/propka.htm.
96. Dupradeau, F.Y., et al., R.E.DD.B.: a database for RESP and ESP atomic charges, and force field libraries. Nucleic Acids Res, 2008. **36**(Database issue): p. D360-7.
97. Tian, C., et al., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. Journal of Chemical Theory and Computation, 2020. **16**(1): p. 528-552.
98. FADx Parameters. Available from: https://upjv.q4md-forcefieldtools.org/REDDB/projects/F-90/script1.ff.
99. Amber Parameter Database.

100. Parameters for FADH. Available from: http://amber.manchester.ac.uk/cof/FADH-.lib.
101. M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 09, Revision A.02. 2016, Wallingford CT: Gaussian, Inc.
102. Wang, J., et al., Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model, 2006. **25**(2): p. 247-60.
103. Hollingsworth, S.A. and R.O. Dror, Molecular Dynamics Simulation for All. Neuron, 2018. **99**(6): p. 1129-1143.
104. Leach, A.R., Molecular Modelling : Principles and Applications. 2001, New York: Prentice Hall:Harlow, England.
105. González, M.A., Force fields and molecular dynamics simulations. JDN, 2011. **12**: p. 169-200.
106. Sargsyan, K., C. Grauffel, and C. Lim, How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. J Chem Theory Comput, 2017. **13**(4): p. 1518-1524.
107. Méndez, R., et al., Assessment of blind predictions of protein–protein interactions: Current status of docking methods. Proteins: Structure, Function, and Bioinformatics, 2003. **52**(1): p. 51-67.
108. Young, D.C., Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems. 2001.
109. Cramer, C.J., Essentials of Computational Chemistry: Theories and Models. 2 ed. 2004: Wiley.
110. Sherril, C.D. An Introduction to Hartree-Fock Molecular Orbital Theory. 2000.

111. Wesołowski, T.A., Hohenberg-Kohn-Sham Density Functional Theory, in Molecular Materials with Specific Interactions – Modeling and Design, W.A. Sokalski, Editor. 2007, Springer Netherlands: Dordrecht. p. 153-201.
112. Foresman, J.B.F.A., Exploring Chemistry with Electronic Structure Methods. 1996, Pittsburgh: Gaussian, Inc.
113. Becke, A.D., Perspective: Fifty years of density-functional theory in chemical physics. The Journal of Chemical Physics, 2014. **140**(18): p. 18A301.
114. Jensen, F., Introduction to Computational Chemistry, 3rd Edition. 2017. 664.
115. 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.
116. Sousa, S.F., et al., Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2017. **7**.
117. Vreven, T., et al., Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. Journal of chemical theory and computation, 2006. **2**(3): p. 815-826.
118. Hratchian, H.P. and M.J. Frisch, Integrating steepest-descent reaction pathways for large molecules. The Journal of chemical physics, 2011. **134 20**: p. 204103.
119. Polyak, I., et al., A microiterative intrinsic reaction coordinate method for large QM/MM systems. Physical chemistry chemical physics : PCCP, 2013. **15 34**: p. 14188-95.
120. Puzari, U., P.A. Fernandes, and A.K. Mukherjee, Advances in the Therapeutic Application of Small-Molecule Inhibitors and Repurposed Drugs against Snakebite. Journal of Medicinal Chemistry, 2021. **64**(19): p. 13938-13979.
121. Ward's Linkage. Available from: https://www.statistics.com/glossary/wards-linkage/.
122. Jarvis-Patrick clustering | Chemaxon. 15.08]; Available from: https://docs.chemaxon.com/display/docs/jarvis-patrick-clustering.md.
123. Diverse Set selection | Chemaxon. 15.08]; Available from: https://docs.chemaxon.com/display/docs/diverse-set-selection.md.
124. Chemaxon. Available from: https://www.chemaxon.com.
125. Moustafa, I.M., Foster, S., Lyubimov, A.Y., Vrielink, A. 2IID :Structure of L-amino acid oxidase from Calloselasma rhodostoma in complex with L-phenylalanine. 2006 07.2022].
126. Protein BLAST: search protein databases using sequence alignment. 04.09].

127. Competitive Inhibition | Biochemistry - britannica. Available from: https://www.britannica.com/science/competitive-inhibition.
128. LigPlot home page. Available from: https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/.