

MSC

2.º
CICLO

FCUP
ICBAS
2021

U. PORTO

Enzymatic Technologies For Cleaning Petroleum
With Low CO₂ Emissions

Bruno Miguel Viveiros Araújo



U. PORTO



U. PORTO



Enzymatic Technologies for Cleaning Petroleum with Low CO₂ Emissions

Bruno Miguel Viveiros Araújo

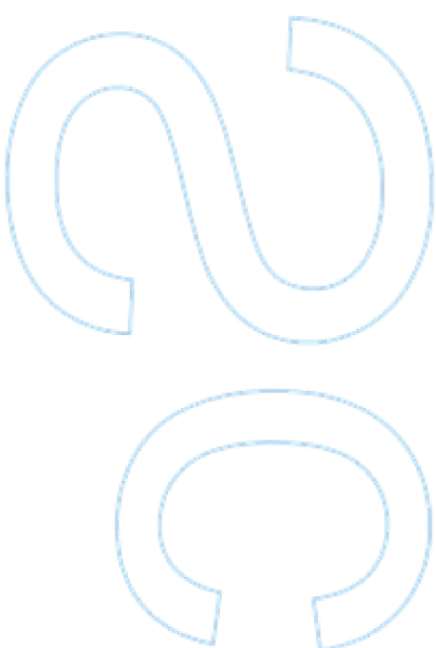
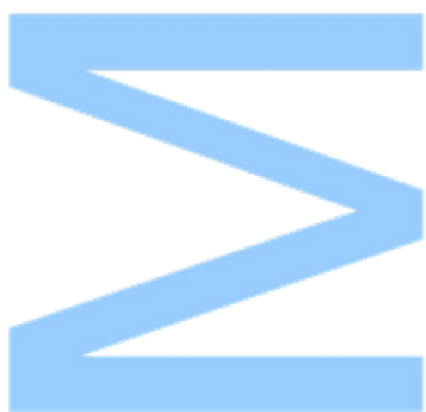
Dissertação de Mestrado apresentada à Faculdade de Ciências da
Universidade do Porto e ao Instituto de Ciências Biomédicas Abel
Salazar da Universidade do Porto em Bioquímica

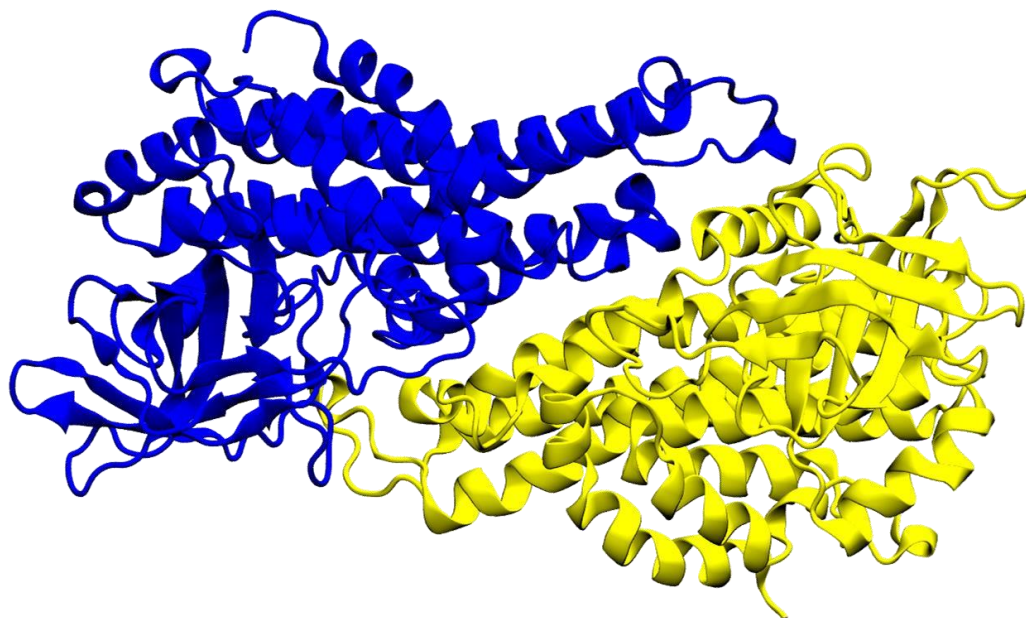
2021

U. PORTO



U. PORTO





Enzymatic Technologies for Cleaning Petroleum with Low CO₂ Emissions

Bruno Miguel Viveiros Araújo

Mestrado em Bioquímica

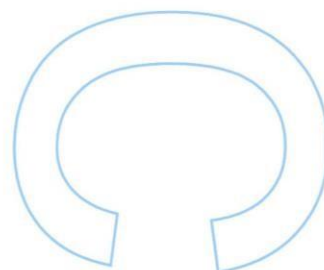
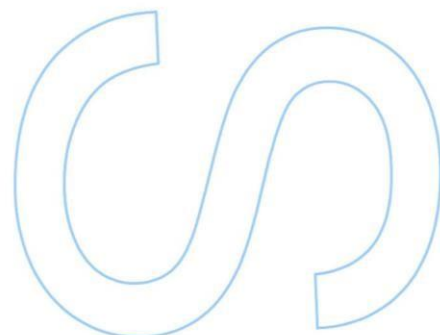
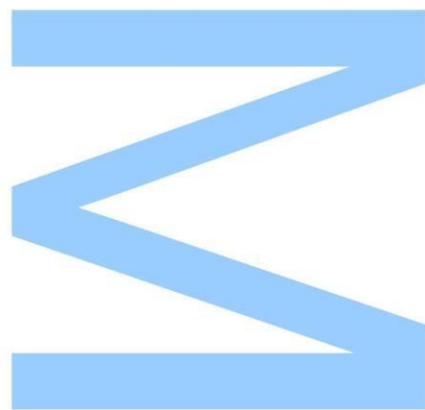
Departamento de Química e Bioquímica.
2021

Orientador

Professor Doutor Pedro Alexandrino Fernandes, Professor Associado,
Faculdade de Ciências da Universidade do Porto

Coorientador

Doutor Rui Pedro Pimenta das Neves, Investigador Júnior, LAQV/REQUIMTE,



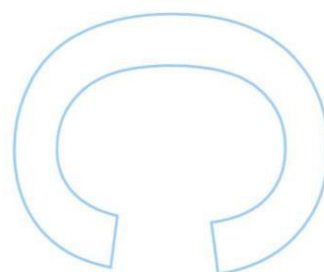
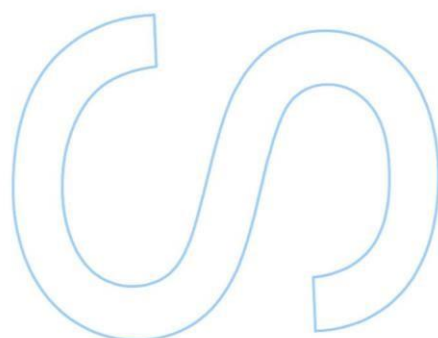
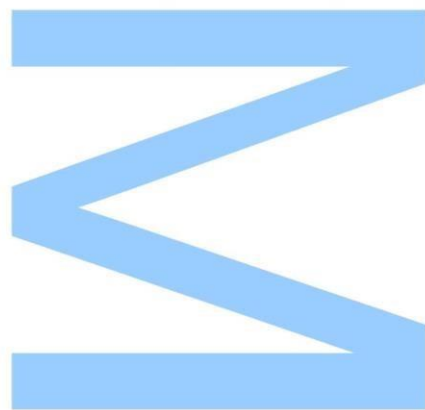


Faculdade de Ciências da Universidade do Porto

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



Agradecimentos

Em primeiro lugar, gostaria de agradecer ao Professor Doutor Pedro Alexandrino Fernandes, pelo apoio e entusiasmos ao longo desta jornada. O seu amor e paixão pela área são inigualáveis e inspiradores.

À Professora Doutora Maria João Ramos pela boa disposição, motivação e oportunidade de integrar no seu grupo de investigação. Este ano no seu grupo despertou-me uma curiosidade enorme pela Química Computacional.

Aos meus colegas de laboratório (João, Carola, Kris, Filipa e Rui) pela companhia e disponibilidade que demonstraram. Um especial obrigado ao Rui pela paciência e trabalho. A tua exigência inspira a que sejamos cada vez melhores e a apontar mais alto. Sem o teu acompanhamento não teria chegado até aqui.

Aos meus amigos Cristiana, Paula, Marta, Tiago, Iago, Sara e Filipa pelas aventuras e histórias destes últimos 6 anos. Ter-vos conhecido foi uma das melhores coisas que me aconteceu. Vou-vos levar sempre comigo para onde quer que vá.

Ao meu irmão Sérgio, pela companhia e inspiração ao longo de todo este percurso. Estar ao teu lado faz-me querer ser melhor.

Aos meus pais, por fazerem tudo isto ser possível. Pelas oportunidades fornecidas, apoio e motivação. Querer fazer-vos orgulhosos é a força que me carrega pela vida.

À Filipa, a minha namorada e companheira de vida/ futura colega de casa e de trabalho. Pelo carinho e cuidado, mesmo quando não merecidos, pelo amor e amizade que me suportam pelo djá, pelos momentos incríveis que vou levar para sempre no coração. Por teres tornado um sonho realidade. Espero poder estar ao teu lado para te fazer sorrir quando e sempre que precisares.

Abstract

One of the major concerns among environmental agencies includes the emission of sulfurous gases into the environment. As such, oil companies are constantly searching for processes to reduce sulfur content in fuels. The current technologies frequently employed are high-cost processes. In the past few years, tremendous advances have been made in elucidating the mechanisms of biodesulfurization through the study of the DszABC enzymes responsible for the process. However, attempts to implement this technology in the industry have failed due to the low reaction kinetics of these enzymes and the high degree of enzymatic self-inhibition by the end-product. This work focuses on the studying of the inhibition of DszC, which is one of the DszABC enzymes, by the end-products of the 4S pathway. We began by selecting an appropriate DszC structure from the Protein Data Bank (PDB). Then, we modelled different complexes containing the enzyme's substrate and cofactor in the catalytic site, and the 2-HBP/HBPS inhibitors in putative binding sites on the enzyme's surface. A combined molecular docking protocol with nanosecond time scale molecular dynamics, followed by a contact analysis, allowed the identification of the most promising binding sites. Subsequently, a cluster analysis uncovered the best inhibitor poses. The two most promising binding sites identified were observed to be shared by both 2-HBP and HBPS and were formed by: (1) residues Arg274, Thr277, Arg278, Thr293, Thr298, Ile299, Tyr302, Gly303, and Thr306 of chain A; and (2) residues Ile107, Val138, Trp141, Phe237, Val238, Phe241, Ile242, Pro413, Gly414, and Phe415 of chain F, and included the conserved residues Thr277, Ile299 and, Gly303. MM/PBSA and a residue fluctuation analysis concluded that the mechanism of inhibition probably involves the displacement of an α -helix (constituted by residues 394404) upon binding of the 2-HBP or HBPS.

The results presented in this thesis might be of value when it comes to further studies on DszC and can help to make viable the use of biodesulfurization by the oil refining industry.

Key words: Biodesulfurization; 4S Pathway; DszC; Non-Competitive Inhibition; Molecular Mechanics.

Resumo

Uma das maiores preocupações entre as agências ambientais inclui a emissão de gases sulfurosos para o ambiente. As actuais tecnologias frequentemente utilizadas são processos de alto custo. Como tal, têm sido desenvolvidas técnicas biotecnológicas que permitem a remoção específica do enxofre dos combustíveis fósseis. Nos últimos anos, foram feitos enormes progressos na elucidação dos mecanismos de biodesulfurização através do estudo das enzimas DszABC responsáveis pelo processo. Contudo, as tentativas de implementar esta tecnologia na indústria falharam devido à baixa cinética de reacção destas enzimas e ao elevado grau de inibição enzimática por parte do produto final da via. Este trabalho centra-se no estudo da inibição da DszC pelos produtos finais da via 4S. Começámos por seleccionar uma estrutura apropriada de DszC a partir da base de dados do protein data bank PDB, e modelámos diferentes complexos contendo o substrato da enzima e o cofactor no local catalítico e os inibidores 2-HBP/HBPS num conjunto hipotético de locais de ligação presentes na superfície da enzima. Uma análise de contactos dos 200ns simulados permitiu a identificação dos locais de ligação mais promissores. Subsequentemente, uma análise de *clustering* revelou as melhores poses dos inibidores. Os dois locais de ligação mais promissores identificados são partilhados tanto pelo 2-HBP como pelo HBPS e são formados pelos resíduos: (1) Arg274, Thr277, Arg278, Thr293, Thr298, Ile299, Tyr302, Gly303 e Thr306 da cadeia A; e (2) resíduos Ile107, Val138, Trp141, Phe237, Val238, Phe241, Ile242, Pro413, Gly414, e Phe415 da cadeia F, e incluiu os resíduos conservados Ile299, Thr277 e Gly303. MM-PBSA e uma análise da flutuação do *backbone* dos resíduos concluiu que o mecanismo de inibição provalvemente envolve o deslocamento de uma α -helice (constituída pelos resíduos 394-404) após a ligação do 2-HBP ou HBPS.

Os resultados apresentados nesta tese podem contribuir para futuros desenvolvimentos em trabalhos relacionados com a DszC e também ajudar a viabilizar a aplicação da biodesulfurização na indústria de refinamento de petróleo.

Palavras-chave: Biodesulfurização, via 4S; DszC; Inibição Não-Competitiva; Mecânica Molecular;

List Of Contents

Introduction.....	1
Sulfur in Crude Oil.....	2
Current Methods of Desulfurization	4
Hydrodesulfurization	5
Extraction Desulfurization	6
Adsorptive desulfurization.....	8
Oxidative desulfurization.....	10
Biodesulfurization	11
Mechanism of DBS	12
Ring destructive metabolic pathway.....	12
Sulfur specific metabolic pathway	13
4S Pathway.....	15
Enzymes of the 4S pathway	15
DBT monooxygenase (DszC).....	15
DBT-sulfone monooxygenase (DszA)	17
HBPS sulfinolyase (DszB).....	17
NADH-FMN oxidoreductase.....	18
Feedback inhibition mechanism.....	19
Theoretical background.....	21
Computational Protein-Ligand Docking.....	21
Searching Algorithms.....	21
Rigid-Body Search Algorithms	22
Flexible Docking Algorithms.....	22
Scoring Functions	22
Force-field scoring function.....	22
Empiric scoring function.....	23

-	
Knowledge-based scoring function	23
Molecular Mechanics	24
Force Fields	24
Amber Force Field	25
Molecular Dynamics	28
Ensembles	30
Periodic Boundary Conditions	30
MMPBSA.....	31
Computational Methods.....	34
Structure Selection and Preparation	34
Parametrization of the C ^{4a} OOH, FMN, DBT, 2-HBP, and HBPS	35
Modeling of DBT	35
Docking of 2-HBP and HBPS.....	36
Classical MD simulations of the complex.....	37
System preparation.....	37
Minimization, annealing, equilibrium, and production protocols	38
Results and Discussion	40
Binding mode identification for HBP and HBPS.....	40
Structure selection and preparation	40
Parametrization of small ligands	41
Modeling of DBT	43
Docking of 2-HBP and HBPS.....	45
Catalytic site docking	45
Blind Docking	46
Blind Docking 2-HBP.....	47
Blind Docking HBPS	48
Characterization of the interaction between the 2-HBP and HBPS and DszC.....	50

-	
Molecular Dynamics	51
Contact Analysis.....	53
Selection and characterization of the most promising binding sites	60
Common binding sites between protocol 1 and 2.....	60
Common binding sites between concentrated and diluted systems	64
Common binding sites between 2-HBP and HBPS inhibitors	66
Most promising binding sites	69
Characterization of the interactions between residues within binding sites and 2-HBP/HBPS inhibitors.....	70
Mechanism of Inhibition.....	76
Interactions at the catalytic site.....	77
Structural changes in DszC upon binding of 2-HBP/HBPS	80
Conclusion.....	84
Future Perspectives.....	85
Bibliography.....	86
Annexes	93

List of Figures

Figure 1 - CO ₂ emissions by fuel type. Annual carbon dioxide emissions from different fuel types, measured in tonnes per year. Retrieved from ourworldindata.com	2
Figure 2 - Representation of some of the major sulfur-containing constituents in crude oil.	3
Figure 3 - Schematic representation of ODS of several organosulfur compounds	11
Figure 4 - Representation of the DBT molecule with the numbering of the carbon atoms.	13
Figure 5 - Schematic representation of the 4S pathway. It starts with DszC performing the double oxidation of the DBT to DBTO and DBTO ₂ , respectively, followed by the first carbon-sulfur bond cleavage performed by DszA, and finally the formation of 2-HBP and SO ₃ ²⁻ by the action of DszB.	14
Figure 6 - Normalized desulfurization enzyme activity at different inhibitor concentrations for the four major inhibitory interactions identified in the 4S pathway. (A) DszA inhibition by 2-HBP; (B) DszB inhibition by 2-HBP; (C) and (D) DszC inhibition by HBPS and 2-HBP respectively. Source: “Exploring the Mechanism of Biocatalyst Inhibition in Microbial Desulfurization” ⁶⁵	19
Figure 7 - Catalytic site of DszC in “Ball-and-stick” with the most important distances identified for the reaction mechanism. Retrieved from Barbosa et. al. ⁵³	20
Figure 8 - Schematic representation of bonded and non-bonded interactions.	27
Figure 9 - Representation of the two chains selected from the DszC enzyme structure, PDB ID: 3x0y. Chains A and F are colored in blue and orange, respectively.	41
Figure 10 - Downloadable file obtained from PROPKA webserver. It can be seen information such as hydrogen bonding, location, solvent accessibility, and Coulombic interactions for each residue. Residues are ordered alphabetically and the represented image is truncated.	42
Figure 11 - (A) and (B) Atom names for HBPS and 2-HBP, respectively, according to parametrization. (C) and (D) atomic charge distribution according to atom types for HBPS and HBP, respectively, by averaging the values depicted in Table 3.	44

Figure 12 - Proposed mechanism of DszC. (A) Oxygen activation; (B) Oxidation of DBT by distal oxygen; (C) Regeneration of FMN. Image retrieved from Barbosa et. al⁵³. 45

Figure 13 Visual representation of the two catalytic sites modeled for the “hybrid” model. On the left, it is represented the catalytic site containing the DBT molecule (ice blue) and the C_{4a}OOH intermediate (tan). On the right, it is represented the catalytic site containing only the reduced FMN (magenta). In New Cartoon, the chains A (blue) and F (orange) are represented. 46

Figure 14 - Above, a representation of the entire complex, with chains A and F colored in orange and blue. Zoomed in are represented the 2-HBP molecules (cyan sticks) docked in the catalytic site in the presence (left) and absence (right) of the substrate (shown in pink). 47

Figure 15 –Image showing DszC enzyme with the 2-HBP molecules spread across putative cavities on its surface. In sticks are displayed the FMN (brown), C_{4a}OOH (purple), DBT (pink), and the 2-HBP (cyan) around the protein. 48

Figure 16 – A representation of the DszC enzyme with the HBPS inhibitor docked in putative cavities across the surface. In sticks a representation of the FMN (brown), DBT (pink), C_{4a}OOH (purple), and HBPS (green) around the protein. 50

Figure 17 - Representation of the resulting system after selection of the best poses for the 2-HBP and HBPS. In surface the DszC enzyme. Inside it is possible to observe chains A (blue) and F (orange). In brown sticks, representation of the FMN cofactor; in purple, the C_{4a}OOH intermediate, and in pink, the DBT substrate. Docked in the enzyme surface are 2-HBP (blue) and HBPS (green) molecules. 51

Figure 18 - Graphical representation of the RMSd value of the protein backbone over time. In the xx axis it is represented the number of frames and on the yy axis the RMSd value. The first 5 frames correspond to the initial structure and to the four steps of the energy minimization protocol. Frames 5 to 110 correspond to the annealing protocol. Frames 110 to 383, demonstrate the enzyme behavior during the restraint portion of the equilibration phase. The remaining corresponds to the RMSd value of the backbone during the unrestrained equilibration. 52

Figure 19 – Graphical representation of the backbone RMSd fluctuation of the concentrated and control systems upon 200ns simulation. In the legend, “p1” and “p2” refer to systems submitted to protocol 1 and protocol 2, respectively. 54

- Figure 20 – Graphical representation of the percentage of contacts between the 2-HBP inhibitor and the DszC enzyme in the concentrated systems, submitted to protocol 1 (left) and 2 (right). In blue are represented the contacts up to 3.5 Å and in red the contacts up to 6 Å, representative of the close range and long-range interactions, respectively. ... 56
- Figure 21 Graphical representation of the contacts between HBPS inhibitor and the DszC enzyme in the concentrated systems, submitted to protocol 1 (left) and 2 (right). In blue represented the contacts up to 3.5 Å and in red contacts up to 6 Å, representative of the close range and long-range interactions, respectively. 58
- Figure 22 - Graphical representation of the contacts between the 2-HBP inhibitor and the DszC enzyme in the diluted system, submitted to protocols 1 (left) and 2 (right). In blue represented the contacts up to 3.5 Å and in red contacts up to 6 Å, representative of the close range and long-range interactions, respectively. 59
- Figure 23 - Graphical representation of the contacts between HBPS inhibitor and the DszC enzyme in the diluted system, submitted to protocol 1 (left) and 2 (right). In blue represented the contacts up to 3.5 Å and in red contacts up to 6 Å, representative of the close range and long-range interactions, respectively. 60
- Figure 24 – Common residue interactions between concentrated system 2-HBP molecules M2 (blue and orange) and M3 (gray and yellow) from protocols 1 and 2 respectively. In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow. 62
- Figure 25 – Graphical representation of the distance between 2-HBP molecules M2 and M3 from concentrated system protocol 1 and 2 respectively and the geometric center of the shared binding site. In orange represented the 2-HBP molecule from the concentrated system submitted to protocol 2 and in blue the 2-HBP M2 of the same system submitted to protocol 1. 63
- Figure 26 - Common residue interactions between concentrated system 2-HBP molecules M4 from protocol 1 (blue and orange) and 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow. 64
- Figure 27 - Common residue interactions between concentrated system 2-HBP molecules M4 from protocol 1 (blue and orange) and 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow. 64

Figure 28 - Common residue interactions between diluted system's HBPS molecules M2 from protocol 1 (blue and orange) and 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow.	65
Figure 29 – Relative positions of the inhibitors 2-HBP and HBPS regarding each other and the enzyme. In surface represented the DszC, in sticks the 2-HBP (blue) and HBPS (green).	66
Figure 30 Common residue interactions between diluted system's 2-HBP molecule M1 from protocol 1 (blue and orange) and concentrated system's 2-HBP molecule M1 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow.	67
Figure 31 - Common residue interactions between concentrated system's 2-HBP molecule M5 from protocol 1 (blue and orange) and diluted system's HBPS molecule M2 from protocol 1 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow.	68
Figure 32 - Common residue interactions between concentrated system's 2-HBP molecule M3 from protocol 2 (blue and orange) and diluted system's HBPS molecule M2 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow.	68
Figure 33 - Common residue interactions between concentrated system's 2-HBP molecule M4 from protocol 1 (blue and orange) and concentrated system's HBPS molecule M2 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow.	69
Figure 34 - Common residue interactions between concentrated system's 2-HBP molecule M4 from protocol 2 (blue and orange) and concentrated system's HBPS molecule M4 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6 Å are represented in orange and yellow.	70
Figure 35 – Cluster sizes clustering analysis on the inhibitors binding on binding site 1. On the xx axis represented the cluster id for cluster of 2-HBP (blue) and HBPS (orange), in the yy axis the number of poses found per cluster.....	72
Figure 36 – Graphical representation of the clustered structures over time for binding site 1.	72

-
Figure 37 - Representation of binding site 1 (residues Arg259, Thr262, Arg263, Thr278, Thr283, Ile284, Tyr287, Gly288, and Thr291) and the most representatives binding poses for 2-HBP. In sticks, the residues and the 2-HBP inhibitor are represented in blue and cyan. 73

Figure 38 - Representation of binding site 1 (residues Arg 259, Thr 262, Arg 263, Thr 278, Thr 283, Ile 284, Tyr 287, Gly 288, and Thr 291) and the most representatives binding poses for HBPS. In sticks are represented the residues (blue) and the HBPS inhibitor (green). 74

Figure 39 – Representation of binding site 1 in the absence of 2-HBP or HBPS inhibitors. When no inhibitors are in proximity, the binding site is closed.	74
Figure 40 – Representation of the entire DszC complex. In cartoon, a representation of DszC. In sticks, the 2-HBP inhibitor (cyan) and HBPS (green), the C ^{4a} OOH intermediate (purple), DBT substrate (pink), and the FMN cofactor (brown). Highlighted in yellow, the loop portion that interacts with the catalytic sites, bound to the same motive as binding site 1.	75
Figure 41 - Cluster sizes clustering analysis on the inhibitors binding on binding site 2. On the xx axis represented the cluster id for cluster of 2-HBP (blue) and HBPS (orange), in the yy axis the number of poses found per cluster.....	75
Figure 42 - Graphical representation of the clustered structures over time for binding site 2.	76
Figure 43 - Representation of binding site 2. In lines are represented the DszC's residues (blue); in sticks the 2- HBP (cyan) and HBPS (green).	76
Figure 44 - Representation of the entire DszC complex. In cartoon a representation of DszC chain's A (blue) and F (orange); in blue lines the residues that make up binding site 1; in sticks the 2-HBP inhibitor (cyan) and HBPS (green), the C ^{4a} OOH intermediate (purple), DBT substrate (pink) and the FMN cofactor (brown).	77
Figure 45 – Representation of the catalytic site of DszC as well as the most important distances for the reaction mechanism. Retrieved from Barbosa at. al.	79
Figure 46 – Residue interaction for C ^{4a} OOH depicted for the different concentrated systems.	79
Figure 47 – Representation of the catalytic site of DszC. To the left, the catalytic site after 200 ns simulation of 2-HBP concentrated system from protocol 2; to the right, the catalytic site after 200 ns simulation of the control system. In lines, the catalytic residues, and in sticks, the C ^{4a} OOH intermediate. Dotted lines in red and blue represent hydrogen bonds, and dotted lines in gray represent the distance between the most important residues and the C ^{4a} OOH.	80
Figure 48 – – Backbone RMSf values for concentrated 2-HBP system from protocol 2 (top) and concentrated and diluted HBPS systems from protocol 2 (bottom).	82

Figure 49 – Binding energy of the interface residues for the 2-HBP concentrated system from protocol 2 upon subtraction of the binding energy of the control system. 83

Figure 50 - Graphical representation of the backbone RMSd fluctuation of the diluted systems upon 200ns simulation. In the legend, “p1” and “p2” refer to systems submitted to protocol 1 and protocol 2, respectively. 95

Figure 51 - C^{4a}OOH contacts with the most relevant residues for the reaction mechanism.. 95

List of Tables

Table 1 - Summary of the systems constructed for each inhibitor and the number of molecules in each system.	38
Table 2 - Below are depicted the histidine residues of chain A and which tautomer was chosen for each. Only one chain is represented because the DszC structure selected is a homodimer, and therefore the two chains are equal.	42
Table 3 - Parameterized charges for each atom of 2-HBP (right) and HBPS (left).	43
Table 4 — Summary of the most representative binding sites found after docking the 2HBP molecule, with detailed information about the residues in each cavity. In curve brackets is the information about the residue chain.	49
Table 5 - Summary of the most representative binding sites found after docking the 2HBP molecule, with detailed information about the residues in each cavity. In curve brackets is the information about the residue chain.	50
Table 6 – Composition of each simulated system in terms of Inhibitor molecules, water molecules, cofactor, intermediate state, and substrate (included in the Additional Ligands row).	53

List of Equations

Equation 1 – General equation used in empirical scoring functions.....	24
Equation 2 – Mean Force Potential	25
Equation 3 – Bonding stretching energy.....	27
Equation 4 – Angle bending energy	27
Equation 5 – Energy associated with improper dihedrals	28
Equation 6 – Torsion energy	28
Equation 7 – Electrostatic interaction energy	28
Equation 8 – Van der Waals energy	28
Equation 9 – General equation for calculation of the binding free energy	33
Equation 10 – Free energy decomposition	33
Equation 11 – Binding energy of the solute and decomposition of MM energy.	33
Equation 12 – Solvent contribution for the binding free energy	33

List Of Abbreviations

2-HBP: 2-hydroxybiphenyl
3-hydroxy-2-formyl Benzothiophene: HFBT
ADS: Adsorption Desulfurization
AMBER: Assisted Model Building Energy Refinement
Biodesulfurization: BDS
C^{4a}OH: C^{4a}-hydroxyflavin
DBT: Dibenzothiophene
DBTO: Dibenzothiophene Sulfoxide
DBTO₂: Dibenzothiophene Sulfone
EDS: Extractive Desulfurization
EPA: Environment Protection Agency
ESP: Eletrostatic Potential
FEP: Free Energy Perturbation
FF: Force-field
FMNH₂: FMN-flavin Mononucleotide
FMN_{ox}: Oxidized FMN
GAFF: General AMBER Force-field
HBPS: 2-hydroxyphenyl-2-sulfonic acid
HDS: Hydrodesulfurization
MD: Molecular Dynamics
MM: Molecular Mechanics
MM/PBSA: Molecular Mechanics/Poisson-Boltzmann Surface Area
MOFs: Metal-organic Frameworks
NMR: Nuclear Magnetic Resonance
NO_x: Nitrogen Oxides
NPT: Isothermal-isobaric Ensemble
NVE: Microcanonical Ensemble
NVT: Canonical Ensemble
ODS: Oxidative Desulfurization
PBC: Periodic Boundary Conditions
PCET: Proton-coupled Electron Transfer Mechanism
PES: Potential Energy Surfaces
PMF: Potential of Mean Force
RESP: Restrained Electrostatic Potential
SASA: Solvent Accessible Surface Area
TI: Thermodynamic Integration, I
VDW: van der Waal

Introduction

Oil and gas have been used by Humankind to some extent for thousands of years before the modern era, such as in lamps or material for construction, with the earliest known oil wells emerging in 347 AD. Nonetheless, the real history of the oil and gas industry began in 1847, with the discovery of natural petroleum in a coal mine, which led to the creation of major oil companies in the late 18th and early 19th centuries that still dominate the oil and gas industry today¹. Ever since its first appearance, oil became one of the most important commodities in the world, playing a crucial part in some of the biggest industries, ranging from chemicals, transport, power production, petrochemicals, etc². Its application became so large that it is now hard to imagine a product that does not contain some trace of crude oil, from gasoline and petroleum-derived products such as tar or asphalt to fertilizers, perfumes, plastics, and more. The wide application of crude oil and derivatives results from its high energy density, which means that even a small amount of oil can produce large amounts of energy.

However, there are also some major setbacks regarding the use of crude oil. The main one being greenhouse gas emission. Burning oil releases carbon dioxide which has been sequestered, for millions of years, into the atmosphere leading to the global warming effect. Another problem of crude oil usage lays in its composition. Between 83% to 87% of oil is carbons, 10 to 14% is hydrogen, 0.1 to 2% nitrogen, 0.5 to 6% sulfur. Additionally, it may contain heavy metals such as copper, nickel, vanadium, and iron³. This means that upon burning, alongside CO₂ emissions, fossil fuel combustion also releases hazardous sulfur dioxide (SO₂) and nitrogen oxides (NO_x), with a significant impact on the environment and human health and welfare. Government policies around the world are recognizing the urge to reduce not only carbon but also non-carbon gas emissions, as stated in the 2030 Climate and Energy Policy Framework or the Paris Agreement^{4,5}.

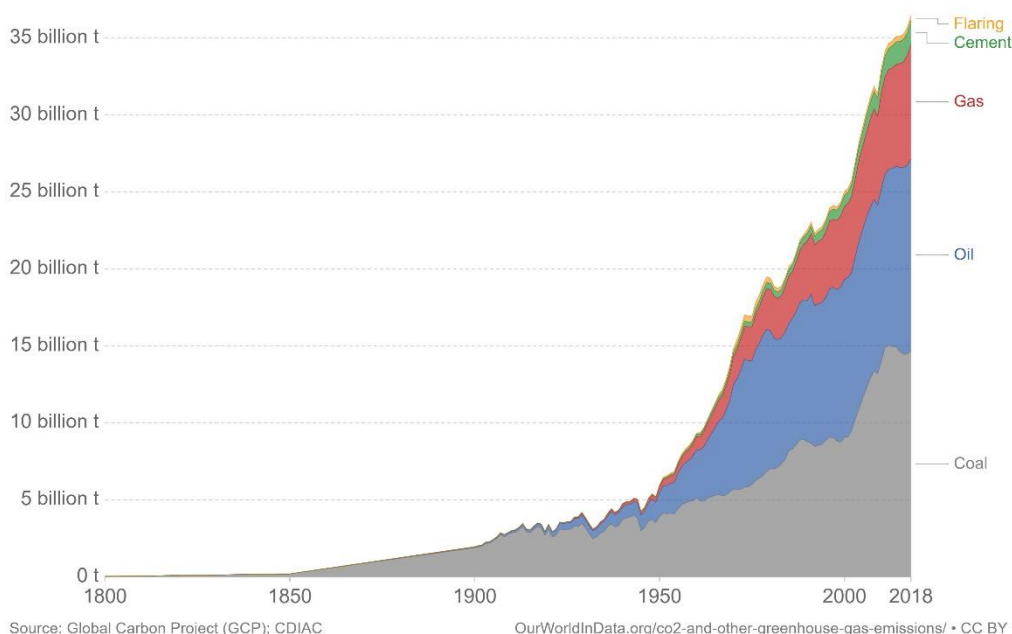


Figure 1 - CO₂ emissions by fuel type. Annual carbon dioxide emissions from different fuel types, measured in tonnes per year. Retrieved from ourworldindata.com

Although greener alternatives have emerged over the last few years, crude oil fuel and crude oil derivatives are still in high demand. In fact, in 2018 over 80% of global energy production was based on non-renewable carbon sources⁶, with crude oil and coal being the main source of global energy production⁷. As a result, over 35 billion tons of carbon dioxide were emitted into the atmosphere in that same year⁸. Despite the obvious impact of fossil fuel burning on the environment and human health, carbon-based energy sources are expected to continue to be explored in the near future⁹. However, large efforts are being made to mitigate the harmful consequences of high CO₂ emissions.

Sulfur in Crude Oil

Crude oil varies vastly in its components, containing a complex mixture of thousands of compounds. Most of these compounds are hydrocarbons (organic compounds composed of hydrogen and carbon atoms). However, it also contains other elements in a significant amount – so-called hetero elements – most notably sulfur. The oil is classified based on its sulfur concentration, as sweet or sour if the weight-to-weight ratio (Wt%) is below or above the 0.5% threshold, respectively¹⁰. These sulfur compounds can be found in either an organic or inorganic forms. Inorganic sulfur compounds such as, H₂S, and pyrite are found in a dissolved or suspended form, while organic sulfur compounds such as thiols, sulfides, and thiophenic compounds represent the main source of sulfur in crude oil. Some of the most relevant classes of organic sulfur compounds are shown in Figure 2.

Sulfur is considered undesired, as it releases toxic, harmful sulfur oxides upon combustion. During the combustion process, the fuel-bound sulfur is rapidly oxidized to sulfur dioxide (SO₂). A small fraction of this SO₂ can be further oxidized to SO₃, which may react with water, giving origin to devastating acid rains. Consequently, most refined petroleum products have a limit on how much sulfur they can contain, making sulfur removal an important part of the overall refining process¹¹.

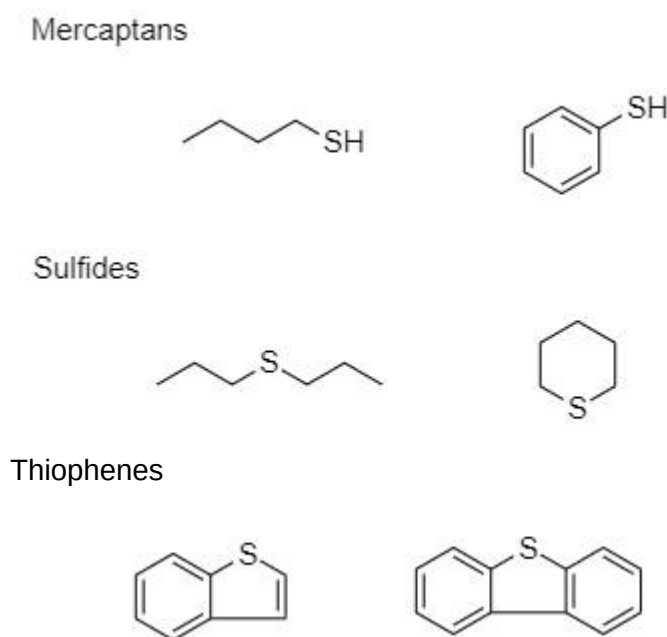


Figure 2 - Representation of some of the major sulfur-containing constituents in crude oil.

Several countries demand lower concentrations of sulfur in fossil fuels. The last update on the European Parliament and the Council of the European Union's regulation established that sulfur content in fossil fuels should not exceed 10 mg/kg⁴. Many oil refining companies have encountered economic difficulties meeting the increasingly stricter legislation. These costs increase with the concentration of sulfur in the oil. As a result, sour oil is more expensive to refine than sweet oil. Moreover, the majority of the world's oil reservoirs contain increasingly higher amounts of sour oil¹², mainly due to the preference for the lighter sweeter deposits, which have, consequentially, been almost depleted. Aside from the economic challenges, more restrictive environmental quality requirements for oil products also have a high impact on CO₂ emissions. Part of the technological development of the oil refining industry worldwide is justified by the fact that it is an industrial activity with high fossil fuel consumption¹³. Between 7% and 15%

of the crude oil input is used by the refinery processes, as these processes are energy-intensive, requiring a considerable amount of direct or indirect heat¹⁴. The high amount of oil consumed by the refineries and the shift to sour oil have been increasing CO₂ emissions due to the higher number of chemical processes required to clean the crude oil.

The cost of desulfurization can also be seen in the barrel price discrepancy between sweet and sour, which can be as high as 15€ per barrel¹⁵. The value of crude oil is determined by the value of the products that can be refined from it. As a rule of thumb, the market pays more for light sweet crude than for heavy and sour grade, which is thicker, has more sulfur, and is harder to refine. Technologically, simple refineries can turn light oil into valuable products like gasoline and diesel. However, if those simple refineries ran heavier oil, they would generate far less gasoline and diesel, and far more fuel oil, which is the heavier, least expensive fuel used for power generation.

Current Methods of Desulfurization

Presently, refineries meet the changing societal needs by upgrading existing technologies and developing more advanced technologies. Adjustments in refining processes are made in response to external driving forces, such as new regulations¹⁶. Environmental restrictions regarding the quality of the produced crude oil derivatives and the emissions from the refinery itself are important issues to consider, as well as the costliest to meet. The recently proposed regulations by the European Parliament and the Environment Protection Agency (EPA) aim to reduce the amount of sulfur content in transportation fuels to a maximum of 10 ppm. The CO₂ emissions emitted by the refinery itself are limited by the Kyoto protocol, which dictates that CO₂ emissions should be 5% below the levels of 1990.¹⁷.

Refineries nowadays are highly integrated industrial plants that aim to produce valuable products of variable composition from crude oil. These industrial plants employ different physical and chemical processes such as distillation, hydrogenation, cracking, amongst others, to convert crude oil into higher-value products. The main commodities converted from crude oil are petroleum gas, gasoline, jet and diesel fuels, and several other petrochemicals. Sulfur present in fuel ends up as SO_x pollutants in the atmosphere, after combustion by the engine.

Organosulfur compounds are presented in almost all fractions of crude oil distillate, with varying reactivities depending on their structure. Usually, higher boiling point fractions contain relatively higher sulfur levels, and the sulfur tends to be of higher molecular

weight. The lower boiling point fractions contain mainly the aliphatic compounds, such as mercaptans, sulfides, and disulfides (Figure 2), which are very reactive and can be easily removed from the fuel. In higher boiling point fractions, the organosulfur compounds present are essentially thiophenic rings, including thiophenes, benzothiophenes, and derivatives, from which sulfur removal is more difficult¹⁶. Compounds containing only a single sulfur atom without any other heteroatoms, such as benzothiophenes and dibenzothiophenes, are the dominant sulfur compounds in crude oil. The ease of sulfur removal from crude oil varies greatly depending on the structure of the sulfur compound. Generally, acyclic sulfur-containing compounds such as disulfides and thiols are extremely reactive and can be easily removed, even under mild operating conditions. Six-membered saturated and aromatic sulfur compounds are usually reactive too, as well as highly condensed ring structures (four or more). However, the reactivity of compounds in which the sulfur atom is incorporated into a five-membered aromatic system reactivity is much lower, and it decreases as ring structures become more condensed. Hence there are several different pathways in which sulfur can be removed from a molecule. Such pathways change for different compound structures¹⁸. Therefore, several methods are available to separate sulfur-containing compounds from crude oil.

Hydrodesulfurization

Hydrodesulfurization (HDS) is the most common technology employed. In general, HDS is performed by co-feeding crude oil and H₂ to a fixed-bed reactor packed with an appropriate HDS catalyst. During this process, the sulfur in organosulfur compounds is converted to H₂S, which is eventually converted into elemental sulfur by a modified Claus process¹⁹. Under mild conditions, H₂S can react with olefins within the reactor to create mercaptans (refer to Figure 2), which can be linear or branched thiols with typically five to twelve carbons. The formation of these mercaptans leads to sulfur retention, limiting the effectiveness of the process²⁰.

The selection of the catalyst is application-dependent. There are two main categories available: i) the hydrogenolysis catalyst (NiMo), which catalyzes a carbon-carbon or carbon-heteroatom single bond cleavage, or lysis by hydrogen and ii) a hydrogenating catalyst (CoMo), which catalyzes the reaction between molecular hydrogen and an atom or compound. Consequently, CoMo-catalysts are preferred for fractions requiring extreme hydrogenation, and NiMo-catalysts are usually preferred for the HDS of unsaturated hydrocarbon streams. The actual conditions vary with the degree of desulfurization intended and the nature of the sulfur compounds present in the processed

crude. Hence, the reaction through which HDS proceeds can be one of the following: the first is called direct hydrodesulfurization and the sulfur atom is removed from the structure and replaced by hydrogen without hydrogenating any other carbon-carbon double bond; the second is called hydrogenative route and assumes that at least one aromatic ring adjacent to the sulfur-containing ring is hydrogenated before the sulfur atom is removed and replaced by hydrogen²⁰. The partially hydrogenated intermediates formed via the hydrogenative route have lower equilibrium constants at higher temperatures and thus, this route is limited at low pressure and high temperature. The direct route involves the insertion of a metal atom on the surface of the catalyst into the sulfur-carbon bond in the sulfur-containing compound. The resultant metal-sulfur bond is energetically favorable, and after several other steps, the sulfur is released as hydrogen sulfide²¹.

While HDS is an extremely effective process to remove the more reactive aliphatic sulfur compounds, the same does not apply to the desulfurization of thiophenic rings, where the resonance stabilization energy of these compounds is around 120-130 kJ mol⁻¹ due to the interaction of the lone electron pair from the heteroatom with the conjugated C=C system's π -electron structure.

Resonance stabilization of thiophene also prevents cracking¹ and is the main reason why these compounds end up forming cokes². Hydrocracking facilitates aromatic hydrogenation, enabling desulfurization by cracking and hydrogenation. It is noteworthy to mention that this process is not to improve HDS efficiency, but rather to active selective ring opening to improve distillate quality. Even though HDS is employed industrially to upgrade heavy oils, its effectiveness is undermined by a high metal content in the oil, causing deposits; fouling propensity from the coking, resulting in catalyst deactivation; molecular size, which limits access to catalyst pores; and steric hindrance of thiophene sulfurs, slowing HDS adsorption.

Extractive Desulfurization

Another well-known methodology for sulfur removal in crude oil is through extractive desulfurization (EDS), in which the extraction relies on the solubility of the organosulfur compounds towards certain solvents. In a separation chamber, the solvent containing

¹ The process of transforming heavy crude oil fractions (primarily gas oils), which have high natural yields in most crude oils, into light refinery streams that go to gasoline, jet fuel and petrochemical feedstocks.

² Coking refineries are a special class of conversion refineries. Coking units destroy the heaviest and least valuable crude oil fractions (**cokes**) by converting them into lighter streams that serve as additional feed to other conversion processes.

sulfur compounds is separated from the hydrocarbons, and post-processing of the solvent allows the removal of the organosulfur compounds and recycling of the solvent back into the mixing tank²².

EDS is a highly attractive method because the crude oil fed into the tank does not undergo any chemical transformations, as it is a purely physical process. As a result, it is easily converted into an industrial process, it does not require H₂ intake, and the mixing tank can be operated at near-ambient conditions. Additionally, EDS extracts the sulfur compounds from fuel oil while preserving its remaining composition, which allows the use of the extracted compounds as raw materials²³.

There are several solvents available, such as acetone, ethanol, and polyethylene glycols, that showed a 50-90% desulfurization efficacy, depending on the number of extraction cycles. Proper physical separation requires the solvent and crude oil to be immiscible, therefore the viscosity of the two liquids must also be as low as possible. This often presents as a problem with heavy crude oil, which requires extraction at higher temperatures to lower viscosity, placing a volatility limit on solvent choice, and extraction may have to occur under high pressure. Additional limitations include the need to use solvents with different boiling points from those of the extracted solvent, and the recovered solvent may contain compounds that are extracted from the oil but cannot be aptly separated by distillation²⁴. Consequently, during recycling, these compounds accumulate, increasing their concentration in the solvent, which must then be disposed of after a few cycles. As a result, despite the obvious advantages of the method, its efficacy is severely limited to the solubility of the organosulfur to the solvent, and the choice of solvent is of utmost importance.

An interesting alternative to overcome some of the problems associated with the traditional organic solvents is the EDS of fuels, such as diesel oil, by ionic liquids, which was observed to provide ultra-clean diesel oils. Because it is a physical method, ionic liquids process works at, or around, ambient temperature and pressure conditions. However, some drawbacks prevent its large-scale implementation. Although many ionic liquids present reasonable distribution coefficients for model sulfur compounds and model mixtures, namely benzothiophene (BT) and dibenzothiophene (DBT), the distribution coefficient for multiple alkylated dibenzothiophenes in real crude oil distillates is relatively poor. This happens because the distribution coefficient depends on the concentration and the structure of all sulfur compounds present in the oil, which results in reduced extraction efficiency and a higher number of extraction steps. In that regard,

a study has suggested that the oxidation of sulfur compounds, namely DBT, to sulfoxides and sulfones could improve extractability by ionic liquids in about 3.6 fold²⁵.

Another disadvantage of using ionic liquids is that the recovery and recycling of the solvent are extremely difficult, namely because ionic liquids have a high boiling point relative to that of organosulfur compounds. Some strategies described in the literature include^{23,24}: (1) direct removal of sulfur-containing compounds from ionic liquids by distillation, (2) extraction from the ionic liquid by transferring them into another solvent, and (3) extraction of sulfur compounds by adding water to the ionic liquids. Those strategies become much more efficient when the organosulfur compounds are oxidized. An important step in ionic liquid extraction is the removal of water from the ionic liquid before recycling, however, this is also the most energy-consuming step because it requires multi-step evaporation at different temperatures and pressures, making the energy demand comparable to that of the traditional HDS.

Even though desulfurization through ionic liquid extraction has some advantages, there are no reports of its use in heavy oil, because the high cost and water sensibility of some ionic liquids erode the cost-effectiveness of extractive desulfurization when applied to heavy oil.

Adsorptive desulfurization

Desulfurization by adsorption is a recent desulfurization technique that uses the principles of adsorption. This method is based on the ability of some solid sorbents to selectively adsorb organosulfur compounds from the oil. This methods' efficacy depends on the properties of the sorbent material, namely its ability to selectively bind organosulfur materials, adsorption capacity, durability, and renewability. Adsorption desulfurization (ADS) can occur in several different forms, including reactive adsorption, polar adsorption, selective adsorption, integrated adsorption, and π -complexation. Amongst those, π -complexation and selective adsorption are widely used for adsorptive desulfurization²⁶. However, it is far more common to group ADS into reactive and nonreactive desulfurization. Nonreactive ADS occurs by physical adsorption, where, similarly to extraction desulfurization, the sulfur compound is not chemically transformed, and it is the least energy-demanding of the two adsorption desulfurization methods. Reactive ADS involves a reaction between the organosulfur compound and the sorbent surface to which sulfur usually binds as a sulfide. Subsequent regeneration of the sorbent can be obtained either by thermal treatment or by flushing the sorbent with desorbent, and sulfur is normally removed as H₂S or SO_x, depending on the process and the nature

of the crude oil. ADS has some advantages over HDS²⁷, namely: the requisite for low temperature, the absence or low consumption of hydrogen for nonreactive and reactive desulfurization, respectively, the ability to remove refractory sulfur, the chemical selectivity, and ease of integration into a portable fuel processor unit, and the absence of H₂S in nonreactive ADS. However, many of the problems associated with catalytic HDS are also inherent for adsorptive desulfurization as both methods rely on a solid surface.

Several sorbents were tested for desulfurization. Among those evaluated are activated charcoal, silica-alumina, zeolites, and metal-organic frameworks. Activated charcoals or activated carbons are a class of microcrystalline materials derived from biomass sources such as coal, wood, coconut shells, fruit nuts, amongst others, with a large surface area and pore volume. Despite having a high sulfur capacity, activated carbons have low adsorption capability due to the broad pore size distribution as well as mass transport limitations²⁸. Alternatively, silica-alumina or zeolites have good adsorptive properties and have been used for the removal of organic compounds from aqueous solutions. Materials containing Si and Al are generally used as catalysts, supports, or adsorbents due to their textural and acid properties since they possess mesopores and Brønsted and Lewis acid sites. These properties can vary depending on the chemical composition, arrangement of silica and aluminum network, and preparation method²⁹. The acid sites on these sorbents possess a high affinity towards soft bases such as thiophenes and derivatives, where the lone pair of electrons on the sulfur atom interacts directly with the Lewis acid site²⁸. The properties of these materials can be manipulated by modifying its structure and their Si/Al ratio³⁰. Metal-organic frameworks (MOFs) are assembled from organic linkers and metal ion cluster nodes. The structure of these MOFs can be customized through a thorough choice of metal-based constituents and organic linkers, forming an almost endless chemical and geometrical diversity³¹.

Despite the acceptable degree of desulfurization obtained under mild reaction conditions, the performance of the most efficient adsorbents is still insufficient for industrial implementation. ADS still faces two major challenges; the first is to develop an easy remunerable adsorbent with a high adsorption capacity for sulfur compounds. The second challenge is to find adsorbents that selectively bind aromatic sulfur compounds. Furthermore, while the adsorption process is highly effective, the adsorbents are difficult to regenerate and often require calcinations³ or solvent washing²⁰. Additionally, the

³ The process of subjecting a substance to the action of heat, but without causing fusion, for the purpose of causing some change in its physical or chemical constitution.

application of adsorbent desulfurization to heavy oils is both unpractical and inefficient, due to the poor accessibility of large molecules in the narrow pores and their steric hindrance.

Oxidative desulfurization

Oxidative desulfurization (ODS) is a process that intends to remove sulfur compounds from crude oil feedstocks through the oxidation reaction in the presence of a suitable oxidizing agent. This reaction may occur with or without a catalyst. In ODS sulfur compounds undergo sequential oxidation that yields the end-product sulfones, with the formation of a sulfoxide intermediate. This step consists of adding two oxygen atoms to the sulfur, without breaking any carbon-carbon bonds. The changes in polarity that come due to the chemical reduction of sulfur compounds make sulfones easily separable from the remaining hydrocarbons, and processable for further use³². Sulfones can be separated from the remaining fuel either by solvent extraction, thermal distillation, or adsorption. In solvent extraction, polar solvents can be used to extract sulfones from non-polar oils, rich in non-polar hydrocarbon compounds. Solvents that selectively adsorb sulfones are more desirable for the removal of sulfones through the adsorption method. Thermal distillation intends to separate sulfones from feedstocks based on their different boiling temperatures because sulfones have a greater boiling point than their parent sulfur compounds. Hence, by collecting only the lower boiling point fractions, sulfones can be retrieved as residues³³. In the past, solvent extraction was often utilized for the removal of sulfones because it is a relatively easy and practical process. A schematic representation of the ODS process combined with solvent extraction can be seen in Figure 3. According to the literature, the solvents used for the extraction of sulfones should have the following characteristics: (a) be immiscible with oil, (b) have the ability to dissolve oxidized sulfur compounds, (c) have a relatively low boiling point for easy separation of sulfones and oil, (d) inability to promptly form an emulsion with the oil, (e) substantial density difference compared to the oil for easy separation, (f) low price as compared to oil, and (g) ability to maintain the fuel properties of the oil³⁴.

The main advantage of oxidative desulfurization is that the strong affinity of sulfur for oxygen, and the lower oxidation potential of sulfur compounds compared to those of hydrocarbons, promote the selective oxidation of the sulfur-containing hydrocarbons without cleaving the C-C and C-S bonds. The more common oxidants used for ODS are organic peracids, H₂O₂ and O₂. The organic peracids can be prepared *in situ* during ODS by reacting an organic acid with H₂O₂. Based on the type of oxidant used, ODS can be classified into three types: two-phase liquid system using water-soluble oxidants,

singlephase liquid system with oil-soluble oxidants, and gas-liquid system with gas-phase oxidants³⁵. The mechanism of action, advantages, and disadvantages of the different types of oxidant are outside of the scope of this thesis and can be consulted in the literature^{27,36}. The order of reactivity varies with each oxidant, catalyst, and experimental condition, however, most organosulfur compounds are reactive to ODS³⁶. Nevertheless, there are still some drawbacks associated with the ODS process, such as the high loading of the oxidizing agent, deactivation of the catalyst, and waste management of the oxidized sulfur compounds. Some initiatives have been developed to mitigate the drawbacks found in the ODS method, including the development of a cost-effective, highly efficient, and recyclable catalyst, and a friendly and cheap oxidizing agent^{37,38}.

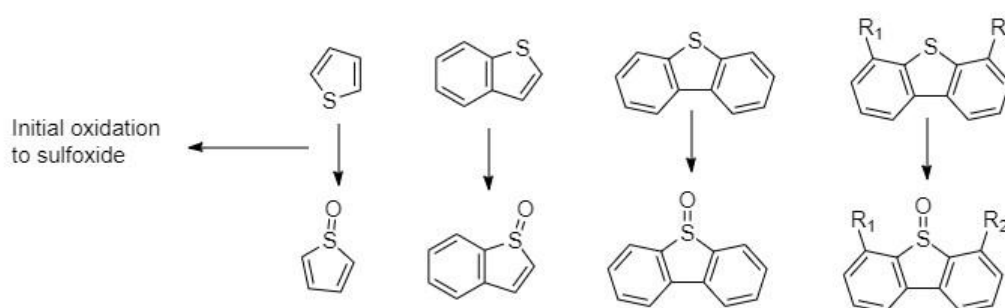


Figure 3 - Schematic representation of ODS of several organosulfur compounds

Biodesulfurization

Biotechnology is a process where microorganisms are used to transform or produce compounds for human application. It is nowadays a vital technique employed in numerous sectors such as pharmaceuticals, agricultural and food industries. Over the past years, it has prompted several innovations and breakthroughs, and it now provides an essential tool with applications on large industrial processes. Biodesulfurization (BDS), in particular, makes use of the ability of biological catalysts to operate at an ambient temperature and pressure, with a great capacity to remove sulfur compounds. Furthermore, BDS results in a low energy cost process, with a minimal generation of undesired side products and low emissions of sulfur-containing compounds and CO₂.

Biodesulfurization is a clean oil technology that aims to reduce emissions of particulate matter and SO_x and NO_x gases, in order to minimize negative environmental impacts of oil derivatives' combustion³⁹. Sulfur is present in bacterial cells, and it accounts for 0.51%

of its dry weight. In nature, several microorganisms are capable of altering the organosulfur molecule to use either the sulfur or the carbon skeleton, and thereby they can be used as an efficient method to remove organosulfur compounds from refinery streams^{40,41}.

As discussed above, petroleum refinement is a process that employs physicochemical methods to remove organosulfur compounds from crude oil. Due to the characteristics of the biotechnological method, it has been used progressively more as an alternative, or a complement to the already discussed methods. Several studies have been carried out on different microorganisms for application on almost all steps of the refining process^{42,43}, namely, biological demulsification, demetallation, denitrogenation, enhancement of oil recovery, conversion of heavy oil into light oil, and even reservoir sourcing⁴⁴. In BDS, bacteria remove sulfur heteroatoms, conserving the carbon skeleton, and bacterial enzymes can catalyze the reaction at a much cheaper cost than the HDS process²⁴, and with reduced greenhouse gas emissions⁴⁵. Therefore, it has been considered as an alternative to the HDS in oil refineries. In BDS, thiophenic compounds are used as sulfur sources by the microorganisms, while DBT is converted, after different steps of oxidation, into water-soluble compounds. However, the end-products of these metabolic pathways usually inhibit DBT oxidation and microbial growth. Despite its potential, the use of microorganisms or their enzymes in the oil industry is still relatively small, and it is mainly focused on bioremediation and not in the refining processes⁴⁶.

Mechanism of biodesulfurization

There are two main metabolic pathways for which the BDS of DBT and its alkylsubstituted derivatives can occur: ring destruction and sulfur specific. Among them, the sulfur specific pathway has gained more attention as it preserves the carbon skeleton, which is of great interest for refineries.

Ring destructive metabolic pathway

In ring destructive BDS, sulfur is not specifically removed from the parental organic molecule. This pathway is composed of three major steps, namely: hydroxylation, ring cleavage, and hydrolysis, and it was first proposed by Kodama⁴⁷. In the first step of the Kodama pathway, oxidation of one of the rings in DBT takes place and results in the destruction of a bond or fragment by enzymatic action. The initial oxidation of the ring occurs in a peripheric aromatic ring, followed by the destruction of the ring. The final products of oxidation are 3-hydroxy-2-formyl benzothiophene (HFBT) and pyruvic acid and were observed to inhibit oxidation and microbial growth upon accumulation⁴¹. This

biological desulfurization pathway is not extensively studied as it leads to the decrease of calorific value of the fuel due to the undesirable cleavage of C-C bonds^{47,48}. Additionally, because ring attack occurs at positions 1 and 7 of DBT, where alky- or arylsubstitutions are frequently found (Figure 4), several substituted DBT compounds are not a substrate for the Kodama pathway, rendering the pathway invaluable in fossil fuel desulfurization.

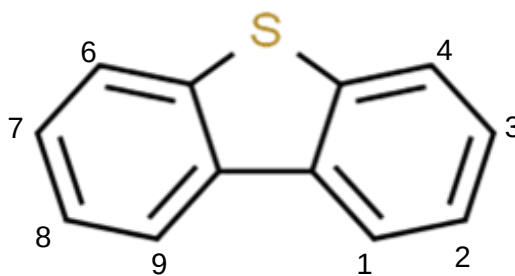


Figure 4 - Representation of the DBT molecule with the numbering of the carbon atoms.

Sulfur specific metabolic pathway

Within the sulfur-specific metabolic pathways, there are two sub-pathways by which this route can be divided: ametabolic pathways and specific oxidative.

Anaerobic metabolic pathway

In this pathway, organosulfur compounds are decomposed by certain organisms, in the absence of oxygen. For example, the reduction of DBT by a thermophilic mixed culture released H₂S and biphenyl as an end-product. In this case, the reductive C-S bond cleavage releases H₂S gas, and the production of organic sulfur compounds is accompanied by anaerobic desulfurization⁴⁹. The major advantage of this technique is that it is easily applicable to traditional desulfurization methods, such as hydrodesulfurization, it presents a low aeration cost and a very low formation of undesirable side-products. On the other hand, the anaerobic metabolic pathway has some drawbacks, such as the maintenance of the anaerobic process, low rate of reaction, and price issues. Due to this, there is no significant anaerobic biodesulfurization on a large or industrial scale⁵⁰. In contrast, the oxidative BDS, carried out by aerobic microorganisms, does not require so much infrastructure, has a higher rate of desulfurization, and can degrade a more significant number of sulfur compounds.

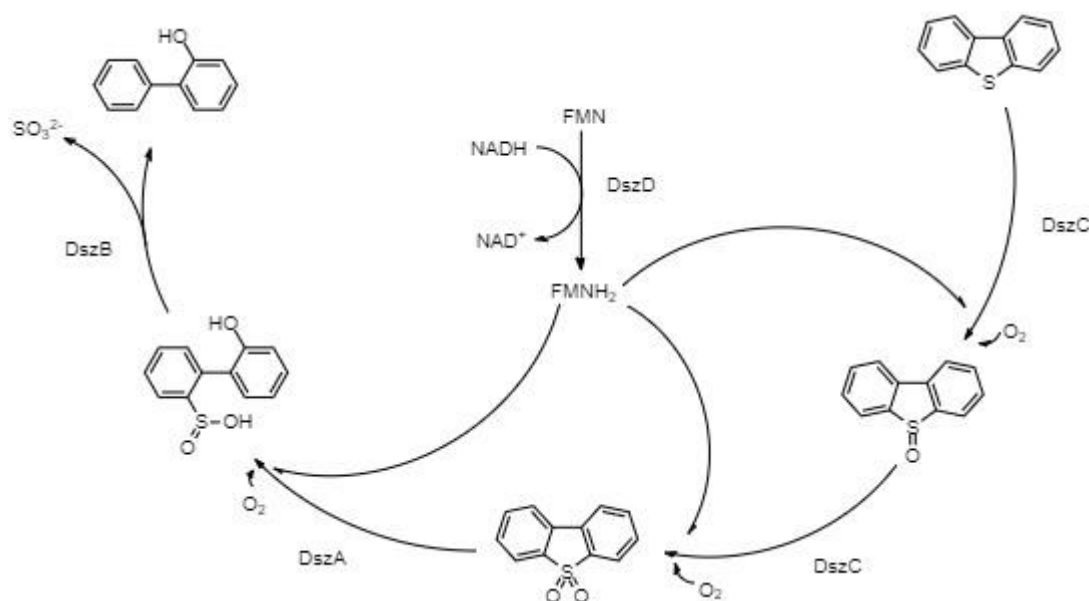


Figure 5 - Schematic representation of the 4S pathway. It starts with DszC performing the double oxidation of the DBT to DBTO and DBTO₂, respectively, followed by the first carbon-sulfur bond cleavage performed by DszA, and finally the formation of 2-HBP and SO₃²⁻ by the action of DszB.

Oxidative aerobic metabolic pathway

In oxidative aerobic BDS, microorganisms are able to oxidize sulfur atoms from DBT without breaking C-C bonds. A lot of attention was brought up to oxidative BDS at the end of the 20th century, with the discovery of the *Rhodococcus erythropolis*, which is able to convert DBT into sulfur-free 2-hydroxybiphenyl (2-HBP). In the reaction carried out by these microorganisms, sulfur is released in the aqueous phase as hydrogen sulfite. This pathway usually referred to as the 4S pathway (Figure 5), is a DBT sulfur-specific desulfurization pathway in which consecutive sulfur oxidation of DBT leads to the breaking of the C-S bonds and conservation of its carbon skeleton, resulting in the 2HBP product and sulfite formation. In the 4S pathway, three enzymes carry out a fourstep attack on the sulfur portion of DBT, selectively releasing the sulfur from the hydrocarbon. The complete reaction of desulfurization of DBT through the 4S pathway can be divided, from a chemical point of view, into three main steps: (1) activation of thiophene, by oxidation of the sulfur; (2) aromatic sulfinate formation, derived from thiophene ring cleavage; and (3) removal of the sulfinate group. The entire reaction is carried by the three enzymes, DszC, DszA, and DszB, encoded in the *dszABC* genes located in the mega-plasmid *pSOX*, and relies on a fourth chromosome-encoded DszD enzyme for cofactor supply. The series of reactions in the 4S pathway is reviewed extensively in the following section.

4S Pathway

The 4S pathway is an aerobic desulfurization pathway. The name of this process comes from the four produced intermediate compounds, namely dibenzothiophene sulfoxide (DBTO), dibenzothiophene sulfone (DBTO₂), 2-hydroxybiphenyl-2-sulfonic acid (HBPS), and sulfite (SO₃²⁻), represented in Figure 5. The reaction catalyzed by this pathway is energetically intensive and requires oxygen, producing sulfite and 2-hydroxybiphenyl as end products. The genes involved in the metabolic desulfurization of the 4S pathway were first reported in *Rhodococcus* sp. IGTS8. However, some thermophilic desulfurizing bacteria, such as *Bacillus subtilis* WU-S₂B, *Paenibacillus*, and *Pseudomonas*, can also grow in the presence of DBT, utilizing DBT as their sole source of sulfur. In particular, the DBT desulfurization enzymes, isolated from *Paenibacillus* (TszABCD) are shown to have higher heat stability than those from *Rhodococcus*, but their enzymatic activities are lower. The corresponding genes from *Bacillus subtilis* WU-S₂B have also been identified and the enzymes (BdsABCD) characterized, showing higher enzymatic activities compared to the *Rhodococcus* enzymes.

In *R. erythropolis*, the 4S pathway comprehends four enzymatic reactions catalyzed by three enzymes of the *DszABC* genes. Those enzymes are: DBT monooxygenase (DszC) that sequentially catalyzes the double oxidation of the DBT into DBT-sulfoxide and then into DBT-sulfone; the DBT-sulfone monooxygenase (DszA) that catalyzes the oxidation of the DBT-sulfone into HBPS; and the HBPS sulfinolyase (desulfinase) (DszB) that catalyzes the conversion of HBPS into sulfur-free 2-HBP. In addition to those three enzymes, the 4S pathway requires an additional enzyme (DszD). This is a NADH-dependent-FMN oxidoreductase that allows the regeneration of the reduced form of the FMN-flavin mononucleotide (FMNH₂) cofactor, needed for the reactions catalyzed by the DszC and DszA enzymes.

Enzymes of the 4S pathway

Below, the role of each of the four enzymes is described. Since DszA, DszB, and DszD are outside of the scope of this work, hence only DszC will be described with atomistic detail.

DBT monooxygenase (DszC)

DszC is the first enzyme in the 4S pathway, and it catalyzes the successive oxidation of DBT to DBTO and then to DBTO₂. The DszC X-ray structure from a *R.erythropolis* strain

IGTS8 showed that it is active as a homodimer, each one possessing two catalytic sites 50 Å apart⁵¹. The catalytic site is formed by the conserved residues His92, Tyr96, Asn129, Ser163, His 388, and His 391 from one monomer and a reduced flavinmononucleotide (FMNH₂) cofactor, which is also located at the interface of the homodimer and contributes to its formation through a network of hydrogen bonds formed between its phosphate moiety and residues from the second monomer in the homodimer (Arg282, Ala369, Arg370)⁵². The impact of each residue on the oxidation of DBT to DBTO₂ was assessed by site-directed mutagenesis and the activity of the mutants was evaluated through quantification of DBTO₂. The enzymatic assays revealed that the mutation of His92, Try96, or Ser163 led to a complete loss of DszC activity.

DszC follows an ordered sequential mechanism, comprising three major steps: (1) the activation of molecular oxygen, forming a C^{4a}-(hydro)peroxyflavin intermediate; (2) the oxidation of DBT to DBTO and DBTO₂; and (3) the deprotonation of the NH of the oxidized flavin to yield FMN and a water molecule. The first step occurs as soon as O₂ enters the catalytic site, with the deprotonation of His391 occurring with no energy barrier. This results in a hydroperoxyl radical. It was also shown that the negative charge in the FMNH was partially transferred to O₂, with O₂ exhibiting a partial superoxide charge. This proton-coupled electron transfer mechanism (PCET) has been observed for similar flavin-dependent (mono)oxygenases. Furthermore, the partial electron transfer precedes the deprotonation of His391. Subsequently, the formation of the C^{4a} – O_p bond occurs. Recent studies suggest that DBT enters the catalytic site once C^{4a}OOH is formed⁵³. In the second step, the nucleophilic attack on the DBT occurs, forming DBTO and C^{4a}-hydroxyflavin (C^{4a}-OH). Given the orientation of the nucleophilic sulfur of DBT toward the distal oxygen of C^{4a}OOH, the oxidation of DBT follows an irreversible nucleophilic S_N2 substitution with a sperimental observed activation barrier of 20.4 kcal mol⁻¹. Structurally, the inline orientation of the nucleophilic sulfur of DBT relatively to the C^{4a}OOH and the hydrogen bond between Tyr96 and the proximal oxygen of C^{4a}OOH are critical aspects for DBT oxidation, as, without these, there is an increase in activation barrier above 25 kcal mol⁻¹. In the last step of the catalytic mechanism of DszC, the FMN cofactor is regenerated from the C^{4a}OH intermediate. In order for this to occur, the C^{4a}OH bond has to be cleaved. This cleavage is preceded by deprotonation of N⁵ of C^{4a}OH. For this step, the His92 is critical. Here, Ser163 acts as a shuttle to assist the deprotonation of the N⁵ nitrogen of C^{4a}OH by His92, which propels the Tyr96-assisted dehydration of the C^{4a} carbon of C^{4a}OH to restore the FMN cofactor.

DBT oxidation is proposed to be the rate-limiting step of the reaction. During this step, a hydrogen bond between Tyr96 and the proximal oxygen (O_p) of C^{4a}OOH anchors it for the nucleophilic attack by DBT-sulfur and prevents steric impediment on DBT from the bulky side chain of Tyr96⁵³.

DBT-sulfone monooxygenase (DszA)

DszA is a flavin-dependent monooxygenase responsible for the oxidation of DBT-sulfone (DBTO₂) to HBPS, breaking the first carbon-sulfur bond. The reaction mechanism of DszA has not been defined yet. The most accepted hypothesis was proposed by Adak and Begley, in 2016. The authors suggested that after the general oxygen activation step, a peroxyhemiacetal would be formed between C^{4a}OOH and DBTO₂, to produce a sulfonestabilized semi-carbanionic structure. Then the proximal C^{4a}-oxygen would be protonated, subsequently attacking the intermediate to form a N⁵-oxide flavin and the HBPS product, and the N⁵-oxide flavin could be oxidized to FMN by a NADH molecule.

No X-ray structure of DszA exists; however, a structure of a DszA ortholog, BdsA from *Bacillus subtilis*, was published in 2017⁵⁴. The two ortholog enzymes possess 453 residues and share 79% sequence identity and 90.9% catalytic site conservation. As such, most of the structural aspects of BdsA are expected to be transferable to DszA. BdsA was reported active as a tetramer, more specifically a dimer of homodimers; however, DszA was reported to be active as a dimer through anion-exchange column chromatography. Each monomer of BdsA shows a TIM barrel fold with a large binding pocket and a reactive conformation requiring FMNH₂. This means that the enzyme should undergo Bayer-Villiger-like rearrangement between the C^{4a}OOH intermediate and the substrate, resulting in the C^{4a}-peroxyhemiacetal previously reported by Adak and Begley for DszA.⁵⁵

HBPS sulfinolyase (DszB)

The cleavage of the second carbon-sulfur bond, forming 2-HPB and SO₃²⁻, is performed by DszB, a desulfinase and the last enzyme in the 4S pathway. DszB is unique since it removes the sulfur from organosulfur compounds without the use of a cofactor, as usually happens with other desulfurization enzymes. The reaction proceeds through an electrophilic aromatic substitution mechanism, which involves the substitution of the sulfinate group of HBPS by a cysteine proton. It was also found that the transition state leads directly to SO₂ release without the formation of a σ-complex, following a one-step, concerted pathway⁵⁶.

Structurally, DszB's fold is similar to those of periplasm substrate-binding proteins. It is composed of two α/β domains⁵⁷. When bound to the substrate the substrate's biphenyl rings are recognized by extensive hydrophobic interactions with the catalytic site residues. Substrate binding is accompanied by structural changes of the catalytic site and His60 is recruited. The sulfinic group of the bound substrate forms hydrogen bonds with side chains of Ser27, His60, and Arg70, and were shown by site-directed mutagenesis to be essential for enzymatic activity⁵⁸. The reaction mechanism proceeds with Cys27 functioning as a nucleophile and it is activated by the sulfinic group of the substrate, whereas His60 and Arg70 orient the *syn* orbital of the sulfinic group to the sulfhydryl hydrogen of Cys27 and stabilizes the negative charge of the reaction intermediate. Cys, His, and Arg residues are conserved in putative proteins homologous to DszB⁵⁹.

NADH-FMN oxidoreductase (DszC)

As mentioned above, DszD is responsible for supplying FMNH₂ to DszC and DszA in the 4S pathway. DszD is a dimer of the class 1 flavin reductases, has 192 residues in each monomer, and catalyzes the reduction of FMN to FMNH₂ through NADH oxidation to NAD⁺. Although no crystal structure of DszD exists at the moment, the catalytic mechanism of DszD was studied on the X-ray structure of the homolog enzyme HpaCst, which has FMN and NADH co-crystallized. The mechanism was found to include two steps: 1) transfer of a hydride from the nicotinamide moiety of NADH to the central nitrogen in the isoalloxazine moiety of the oxidized FMN forming FMNH, identified as rate-limiting, and (2) proton transfer from a water molecule to the nitrogen atom near the ribityl tail of FMNH, resulting in the reduced FMNH₂ product⁶⁰.

The supply of FMNH₂ to DszA and DszC can limit the overall rate of the 4S pathway since it has been shown that overexpression of DszD increases the overall rate of desulfurization of *R. erythropolis*⁶¹. Studies have been conducted in order to increase DszD activity and therefore the amount of FMNH₂ available for DszA and DszC. One study found that replacing a threonine residue at position 62 for a positively charged asparagine resulted in a 5-fold increase compared to that of the wild-type enzyme, probably due to the accelerated electron transfer rate⁶².

Feedback inhibition mechanism

There are two major obstacles facing BDS commercialization: (1) biocatalysts' activity is usually much lower than what is needed to match HDS rates and (2) biocatalysts cannot maintain activity for a long period of time⁶³. Considerable efforts have been put into trying to understand and overcome the limitations to the desulfurization activity by the biocatalyst. Higher desulfurization activities can be easily achieved by enzyme immobilization or genetic engineering⁶⁴. The reduction in activity is typically correlated to an increase in 2-HBP in the medium during BDS. In fact, as can be seen in Figure 6,

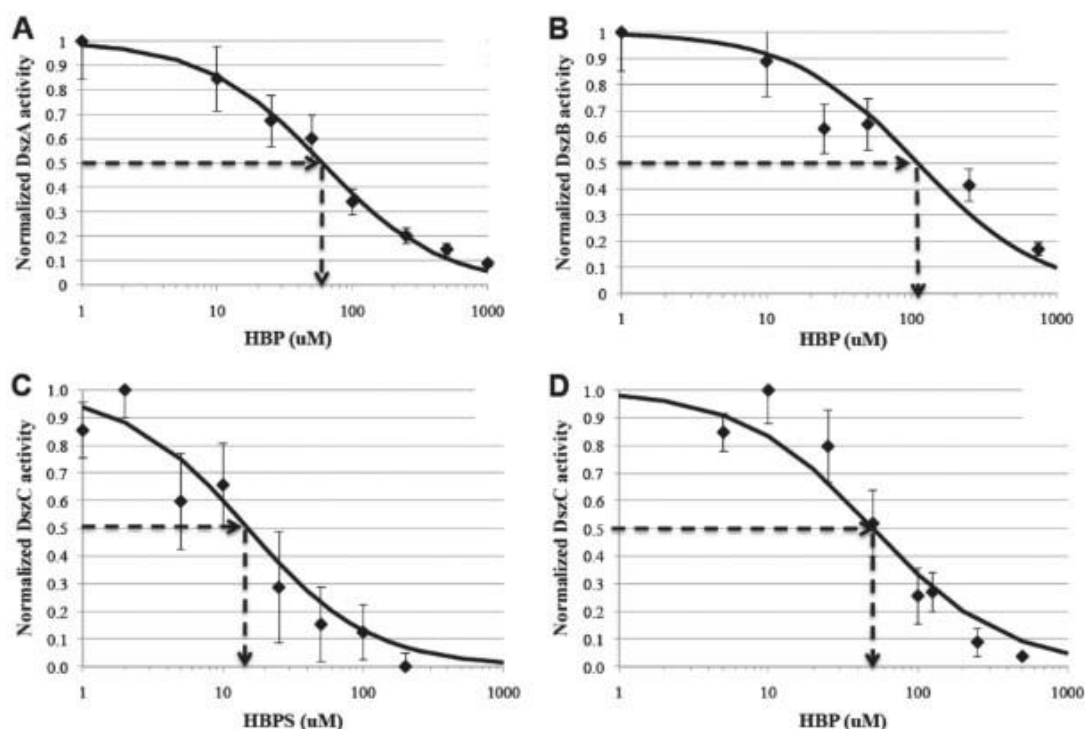


Figure 6 - Normalized desulfurization enzyme activity at different inhibitor concentrations for the four major inhibitory interactions identified in the 4S pathway. (A) DszA inhibition by 2-HBP; (B) DszB inhibition by 2-HBP; (C) and (D) DszC inhibition by HBPS and 2-HBP respectively. Source: "Exploring the Mechanism of Biocatalyst Inhibition in Microbial Desulfurization"⁶⁵

2-HBP can inhibit three of the Dsz enzymes. DszABC has been shown to be inhibited noncompetitively by 2-HBP. Regarding DszB, it was shown that 2-HBP binds the interface of the two α/β domains. Additionally, DszB has a lower catalytic efficiency when compared to other Dsz enzymes. Hence, several studies have been conducted on this enzyme, as it is an appealing target for enhancement through protein engineering⁶⁶. Out of all the Dsz enzymes, DszC is one of the most severely affected by the feedback inhibition caused mostly by 2-HBP and HBPS products of DszB and DszA, respectively. Additionally, studies have shown that together with DszB, DszC is one of the least efficient enzymes in the 4S-pathway, with a k_{cat}/K_M of $1.3 \mu\text{M}^{-1}$. Therefore, DszC is one of the most appealing enzymes for optimization through enzyme engineering. For

instance, our group's data suggest that reducing the number of positively charged residues in the immediacy of the C^{4a}OOH-hydroperoxyl may lower the activation barrier of DBT oxidation in up to 2 kcal mol⁻¹. This would correspond to an increase in the reaction rate of approximately 30-fold for a single mutant.

Alternatively, studies have also been conducted to target the mechanism of feedback inhibition in DszC by the HBPS intermediate and the 2-HBP end-product. In one study, desensitization engineering and DszC overexpression were used to develop double mutants 13.8-fold faster than the wild-type DszC. The authors also indicated that the mutation of the Trp327 was important for increased efficiency. However, the enzyme patches responsible for the feedback inhibition still remain unknown. In this regard, we are currently developing combined molecular docking and MD simulations protocols to identify these regions and pave ways to develop rational mutants with improved efficiency.

In summary, the commercialization of BDS would be a positive turn-over as it may be regarded as a greener alternative. However, there is sufficient data to suggest that the presence of 2-HBP/ HBPS reduces the overall activity of the biocatalyst, especially the DszC enzyme. In this work, computational methods will be used to identify hot-spot regions that may bind the 2-HBP or HBPS molecules and to analyze those regions to identify the phenomena responsible for inhibiting the DszC enzyme.

Theoretical background

Protein:Ligand Docking

Molecular docking is widely used to study protein-ligand interactions. This method is a computational procedure that attempts to predict the non-covalent binding of macromolecules, or more frequently, of a macromolecule, the receptor, and a small molecule, the ligand. Docking starts with the unbound structures obtained from MD simulations, crystallographic structures, homology modeling, etc. The goal is to quickly predict the bound conformation and the binding free energy of the ligand to the target receptor.

The prediction of the binding of small molecules to a given protein is very important and it can be used to screen virtual libraries of drug-like molecules in order to identify new inhibitors for drug development⁶⁷. However, docking can also be used to try to predict the bound conformation of known binders, when the experimental *holo*-structures are unavailable. The docking process involves two basic steps: i) prediction of the ligand position and orientation within the binding site, usually referred to as *pose* and, ii) assessment of the binding affinity. These two steps involve search algorithms and scoring functions, which will be discussed further below.

Knowing the location of the binding site can significantly increase the docking efficiency and rate of success⁶⁸. However, in cases where the binding site is unknown, information about those sites can be discovered through comparison of the target protein with a family of proteins sharing a similar function, or with proteins co-crystallized with similar ligands⁶⁷. In the absence of such knowledge, webserver such as POCKET⁶⁹, SurfNet⁷⁰, GRID⁷¹, or others, can be used to detect cavities and identify putative sites within the protein.

Searching Algorithms

Searching algorithms perform the task of sampling, that is, generate different poses of the ligand into the selected binding site. This is done by incrementally modifying structural parameters of the ligand, such as torsions, additionally, other parameters like translational and rotational degrees of freedom are also incremented. Regarding the way searching algorithms treat the ligand, they can be grouped into three categories: rigid-body methods, flexible ligand docking methods, and flexible target methods.^{72,73}

Rigid-Body Search Algorithms

Rigid-body algorithms consist of those that do not modify the bond angles, bond lengths nor torsion angles at any stage of the protein-ligand complex generation. These algorithms usually produce results with low specificity and accuracy and are only able to identify accurately ligand binding sites for quasi-reactive protein conformations. Rigidbody algorithms consider only six degrees of freedom (three translational and three rotational) and, therefore, are faster and require less computational power when compared to flexible docking⁷⁴.

Flexible Docking Algorithms

Flexible docking algorithms, as opposed to rigid-body, consider the conformational degrees of freedom of either the ligand, receptor or both at the same time. Molecular docking with flexible ligand and flexible receptor is the most correct approximation to *in vivo* complexes formation, especially when conformational changes preceding the binding are present. However, due to the high computational power required, most of the methods only consider the ligand conformational space, treating the receptor as rigid. This later approach is usually a good trade-off between accuracy and computational power⁷².

Scoring Functions

Searching algorithms predict several conformations and orientations (the pose) for the ligand inside the binding site. Scoring functions then perform the evaluation and ranking of the poses through some approximate mathematical function that predicts the binding affinity between the ligand and the receptor. In general, these functions can be grouped into three major classes, according to how they were derived: force-field, empirical, and knowledge-based⁷⁵.

Force-field scoring function

Force-field (FF) scoring functions are developed based on the physical properties of atomic interactions, including van der Waals (VDW) interactions, electrostatic interactions, and intramolecular forces namely, stretching, bending, and torsion. Parameters for FF scoring functions are derived from both experimental data and *ab initio* quantum mechanical calculations according to the principles of physics. Despite the computational efficiency of the FF scoring functions the distance-dependent dielectric factor of the electrostatic interactions cannot account for desolvation, an important

solvent effect in which charged groups interact favorably with aqueous environments and non-polar groups tend to interact more in non-aqueous environments⁷⁴. The desolvation energy depends on specific geometric and chemical surrounding environments of the considered solute atoms. If the desolvation effect were to be ignored, a scoring function would be biased on Coulombic electrostatic interactions and therefore would tend to select highly charged ligands. Since entropic contributions from solvation cannot be accounted for, several techniques are used in post-docking refinement⁷⁵. Those include techniques that use explicit or implicit water molecules.

Empirical scoring function

Empirical scoring functions are developed to replicate experimental affinity data, based on the idea that it is possible to correlate the free energy of binding to a set of uncorrelated variables. A general expression is shown below⁷⁵:

$$\Delta G = \sum_i w_i \Delta G_i \quad (1)$$

Equation 1 – General equation used in empirical scoring functions

where ΔG_i represents different energy terms, namely, VDW, electrostatic, hydrogen bond, desolvation, ligand entropy, etc. The corresponding coefficients, w_i are determined by fitting the binding affinity data of a training set of protein-ligand complexes with known three-dimensional structures. Compared with FF scoring functions, empirical scoring functions are much faster at calculating the binding score due to their much simpler energy terms given that they are much easier to calculate⁷⁶.

Knowledge-based scoring function

Knowledge-based scoring functions employ energy potentials that are derived from structural information embedded in experimentally determined structures. In other words, knowledge-based functions are based on the statistical analysis of interacting atom pairs from protein-ligand complexes with available three-dimensional structures. These pairwise-atom data are then converted into a pseudopotential, the potential of mean force (PMF), that describes the preferred geometries of the protein-ligand pairwise atoms⁷⁷. In simpler terms, the PMF score is calculated as a sum over all protein-ligand atom pair interaction free energies $A_{ij}(r)$ as a function of the atom pair distance r . The equation can be seen below⁷⁸.

$$PMF = \sum A_{ij}(r) \quad (2)$$

$$ijr < r_{cutoff}$$

Equation 2 –Potential of Mean Force

In summary, molecular docking is a fast approach that enables the discovery of hot-spot regions that may bind a given molecule. However, this method is still far from giving accurate results, and therefore, to validate the findings attained using this procedure, molecular dynamics simulation are generally performed, and the behavior of the docked molecules is analyzed over the simulated time⁷⁹. The theory behind such simulations is described below in detail.

Molecular Mechanics

Molecular mechanics (MM) deals with the classical description of molecular systems. In MM, nuclei and electrons are merged into single atom-like particles, treating the atoms and the bonds between them as spheres and springs. These simplified assumptions and approximations enable the use of MM methods for a wide number of systems, ranging from low-molecular-weight molecules to large biomolecular complexes.

Force Fields

The MM approach is a generalization of these models via mathematical expressions. A set of those expressions together with their numerical coefficients is usually referred to as a force field and it can be used to estimate the forces between atoms and molecules in a given user-defined system. This method is defined by an energy function and by a set of parameters associated with the energy contribution of the different terms of that potential energy function. Nowadays, several force fields exist with different purposes; however, their general application is to calculate the total energy of a system as a function of its atomic coordinates. The most common biomolecular force fields are AMBER⁸⁰, CHARMM⁸¹, OPLS⁸², and GROMOS⁸³. These apply similar potential energy functions with minor variations amongst each other. Simply put, force fields describe molecules in terms of bonded atoms that usually have their geometry distorted from an ideal (equilibrium) geometry, due to a set of nonbonded interactions between the atoms of the same molecule and others. Force fields currently used in atomistic molecular simulations differ in the way they are parametrized. These parameters are not necessarily interchangeable, and not all force fields allow the representation of all molecule types.

Amber Force Field

The term AMBER (Assisted Model Building Energy Refinement) is a classical molecular mechanic's FF, primarily designed for the simulation of biomolecules. It contains a set of parameters that describe the most common components of biomolecular simulations, namely the most naturally occurring solvents, ions, amino acids, carbohydrates, and lipids. Additionally, the general AMBER force field (GAFF) also allows the parametrization of most organic molecules. Charges in AMBER were calculated using an Electronic Surface Potential (ESP) using the 6-31G* basis set; however, this leads to an over-polarization due to the gathering of partial charges on single atoms. Later the HF/6-31G* basis set variant was developed. In this model the over polarization of the atoms was corrected by calculating the charges in a vacuum, making a more correct approximation to the self-polarization of molecules and/or the condensed phase⁸⁰. Later on, the improved Restrained Electrostatic Potential (RESP) was created. RESP forces equivalent atoms to have equal charges (for example, the three hydrogen atoms in the amine group of a lysine).

FFs usually decompose molecules in bonded interactions, which can be described with harmonic potentials such as Hooke's law, and non-bonded interactions which include electrostatic interactions and van der Waals interactions⁸⁴. The bonded and non-bonded interactions are discussed in detail further below.

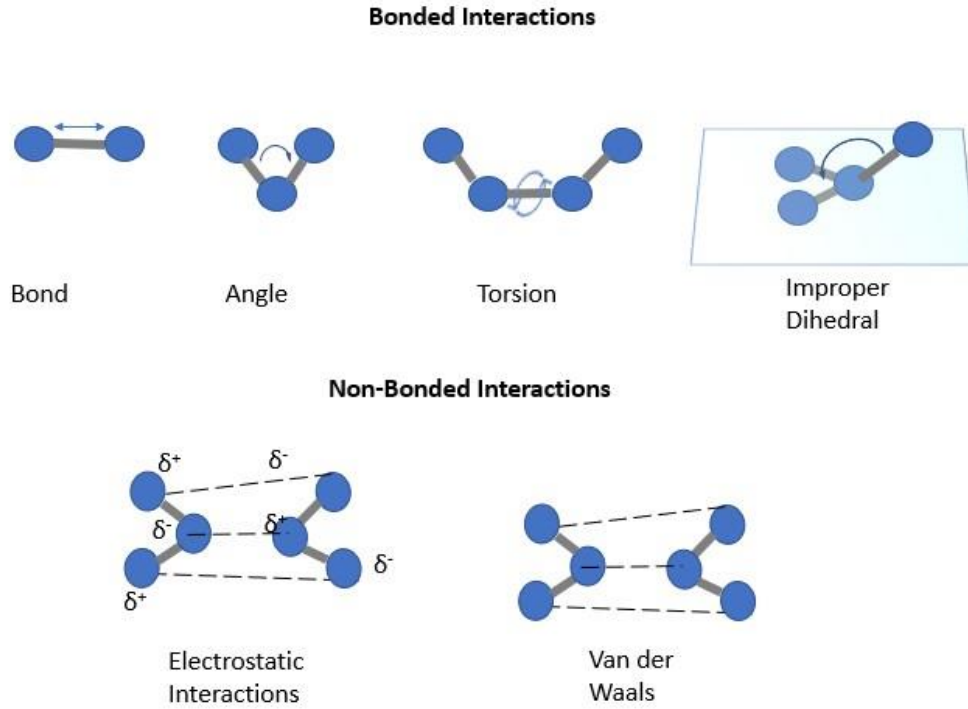


Figure 7 - Schematic representation of bonded and non-bonded interactions.

Bonded Interactions

The potential energy function comprises three different types of bonded interactions: bond stretching, angle bending, and torsion. Regarding the bond stretching and angle bending, these are treated as harmonic oscillators, as a function of each bond length and valence angle, respectively. In order to calculate the contributions of these terms, only two parameters are required for each: the equilibrium or reference value (b_0 and θ_0) and the force constant (K_b and K_θ).

$$\sum_{\text{Bonds}} K_b (b - b_0)^2 \quad (3)$$

Equation 3 – Bond stretching energy

$$\sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 \quad (4)$$

Equation 4 – Angle bending energy

Equation 4 only describes angle movements centered on the same atom. Those are said to be in-plane angles. An equation of the same form can be used to describe out-of-plane angles. These are oscillations of two bonds not centered on the same plane, relatively to

the angle formed between them, and are generally called improper angles (Equation 5). Here the potential energy is described by a harmonic potential that is a function of the out-of-plane angle ϕ , formed by three atoms which are all attached to one "central" atom. Finally, the torsional energy is described by a sum of cosine function with multiplicity $n=1,2,3,4\dots$ and amplitude $K_{\phi n}$. The phases δ_n are usually constrained at 0° or 180° so that the energy surface of achiral molecules is symmetric, and the enantiomers have the same energetic properties.

$$\sum_{\text{Improper Dihedral}} K_{\phi}(\phi - \phi_0)^2 \quad (5)$$

Equation 5 – Energy associated with improper dihedrals

$$\sum_{\text{Dihedral torsion}} K_{\phi n}(1 + \cos(n\phi - \delta_0)) \quad (6)$$

Equation 6 – Torsion energy

Additionally, sometimes, two or more periodicities are used for a given torsion angle to describe more accurately the change in energy of the system as a function of rotation around the central bond in the dihedral.

Non-bonded Interactions

Non-bonded interactions refer to all atom-pairs not bonding to each other or neighboring atoms up to two adjacent covalent bonds. Each atom-pair interaction is represented by two terms: a sum of electrostatic and van der Waals interactions.

$$\sum_{ij} k \frac{q_i q_j}{r_{ij}} \quad (7)$$

Equation 7 – Electrostatic interaction energy

$$\sum_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \quad (8)$$

Equation 8 – Van der Waals energy

As shown in equation 5, the electrostatics in these force fields are handled by a Coulomb potential between fixed point charges q_i and q_j . These are also known as partial charges.

This type of treatment of electrostatic interactions is referred to as “additive” because the charges are independent of each other, and individual charge interactions may be summed to yield the total electrostatic energy of the system.

As for the van der Waals component, usually, a classical Lennard-Jones (LJ) 12-6 potential is applied (Equation 7), defined by the radius r_{ij} and experimental determined parameters A_{ij} and B_{ij} . The LJ directly relates to the distance between the atoms containing the charges. The majority of the force-fields utilizes the 12-6 expressions for the description of the van der Waals interactions; however, some replace it by the 12-10 one for the interactions of hydrogen atoms, in order to describe a sharper distance dependence in the most important area of the energy minimum corresponding to hydrogen bond formation⁸⁵.

The calculations of potential energy via these functions are used to search for a local minimum (atom positions that possibly correspond to a stable protein conformation), to construct and analyze multidimensional Potential Energy Surfaces (PES), to follow trajectories from molecular dynamics simulations, or to study the average thermodynamic properties and characteristics of a system.

Molecular Dynamics

Molecular dynamics (MD) simulations predict how a given system behaves over time, based on a general model of physics governing interatomic interactions, as described in the FF section above. These simulations can capture a wide variety of important phenomena, including conformational changes, ligand binding, and protein folding, by calculating the atom positions at a femtosecond temporal resolution. More importantly, these simulations can also predict how a given system will respond to a perturbation such as a mutation, phosphorylation, protonation, and addition or removal of ligands. MD simulations are usually used in combination with a wide range of experimental techniques like X-ray crystallography, cryo-electron microscopy, Nuclear Magnetic Resonance (NMR), *etc*⁸⁶.

The advancement in computing architectures spurred the extensive growth of classical MD simulation tools over the past decades. Such advancements enabled a more reliable prediction of the physical properties of proteins as well as their structure at an atomic level, and the observation of biological phenomena, such as those described above, that

may require thousands of nanoseconds. Additionally, classical MD has several other advantages, such as the possibility to treat larger systems without compromising computational efficiency, its ability to determine dynamic and thermodynamic properties of the system with greater statistical significance and to offer a visual perspective of the evolution of the system over time. The major downfall of classical MD simulations is that, since they do not explicitly include electronic effects, it is generally prohibitive to directly study chemical phenomena such as covalent bond breaking and formation⁸⁴.

Classical MD simulations are generally obtained by iteratively solving Newton's Second Law of Motion, in which the movement of an atom is determined by integration of Newton's equation, $F = m \cdot a$. Where F is the force constant on each atom, m is the mass of the atom and a the atom's acceleration. F can be obtained from the equation $F = -\nabla U$ where the force F in each atom can be determined from the gradient of the potential energy U along with its coordinates. Once the forces acting on individual atoms are obtained, the acceleration and velocity are calculated (velocity is the derivative of position, and acceleration is the derivative of velocity), and the atom positions are updated. Because it is analytically impossible to calculate the atom's position at any moment in time, an integration step is used⁸⁷.

The integration step is a user-defined variable that represents the timestep between each iterative calculation of the equations of motion. Between each iteration, it is assumed that the forces acting on the system remain constant during the integration step. The accuracy of the movement prediction is directly related to the size of the integration step. The shorter the integration step, the more accurate the positions of the atoms will be predicted. For an accurate simulation, it is considered that the integration step should be at least 10 times smaller than the fastest movement in the system, which corresponds to the vibration between covalent bonds involving hydrogen atoms⁸⁸. A complete cycle of shortening and stretching of that bond takes about 10 fs, and so the integration step is usually set to 1 fs. A form of increasing the integration step, and therefore reduce the number of steps required to simulate the same amount of time, is to resort to the SHAKE algorithm, which constrains the bonds between all heavy atoms and hydrogens at the beginning of the integration step. In this sense, the bonds between the second-row atoms become the fastest movement in the system, completing a cycle in 20 fs, and a 2 fs integration step can be used⁸⁸.

Ensembles

The role of thermodynamic ensembles is to derive the classical thermodynamics of materials in terms of the properties of their constituent particles and the interactions between them. In other words, thermodynamic ensembles provide a connection between the macroscopic properties of materials in thermodynamic equilibrium, and the microscopic behavior and motions occurring inside the simulation cell. When choosing an ensemble one must define the system (*i.e.* the portion of the ensemble that contains the process in consideration), the neighborhood, which contains everything else, and the boundary, the imaginary wall that separates the two⁸⁸. The system can interact with the neighborhood on several levels, hence, there are isolated systems that do not interact with their surroundings, and systems that are completely open, exchanging energy and matter. In terms of MD, there are three main ensembles that are used⁸⁹. The microcanonical ensemble (*NVE*) is completely isolated from its surroundings. In other words, the total number of particles, the total volume, and the total energy is constant. There is no defined temperature, but the energy of the system is constant. In the canonical ensemble (*NVT*) the volume and temperature of the system are fixed. This system can be described as being immersed in a heat bath much larger than the system itself. Since the system can exchange heat with the surrounding, the states of the system will differ only in total energy. In the isothermal-isobaric ensemble (*NPT*), the volume of the system can change, such that the internal pressure of the system matches the pressure exerted on the system by its surroundings. In this ensemble, the total number of particles is maintained, and constant temperature and pressure are applied⁸⁹.

It is important to note that the type of ensemble used depends on the biological process intended to replicate. As an example, if one pretends to mimic a reaction that occurs in an animal cell one would use the *NPT* ensemble. However, if the same reaction were to occur in a more rigid cell the *NPT* ensemble might not be adequate.

Periodic Boundary Conditions

Correct treatment of the boundaries and boundary effects is crucial for simulation methods, as it enables the calculation of the macroscopic properties from simulations, using appropriate model systems. However, in simulations, the molecules near the boundary experience different conditions than those in the center.

One way to overcome that problem is by applying Periodic Boundary Conditions (PBC). In this approach, the system is replicated an infinite number of times, each of those replicas are called cells. By applying PBC, the particles within the system experience

forces as if they were in a continuous bulk fluid. The movement equations are solved for each particle in the unit cell and the movements are reproduced in the replicate cells. Should a particle leave the unit cell during the simulation, it is replaced by an imaged particle that enters from the opposite side⁹⁰. Thus, the number of particles in the unit cell remains constant. For this approach to be effective, the distance between the protein and the limit of the cell should be high enough to discard the possibility of non-bonded interactions between imaged atoms of neighboring image cells.

Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PBSA)

The theoretical prediction of binding affinities is one of the most important problems in computational chemistry. The most rigorous methods used to calculate binding affinities involve alchemical or structural transformations, such as Free Energy Perturbation (FEP)^{91,92} and thermodynamic integration (TI)^{93,94}. The accuracy of these methods relies on equilibrium sampling of the entire transformation path, from an initial to a final state. Hence, due to the high computational demand, FEP and TI are most amenable for determining binding affinities between similar ligands. End-point free energy methods, where only the initial and final states of the system are evaluated, present a desirable alternative to perturbation simulations. They are less computationally expensive making them suitable for a greater variety of systems and problems. They are typically based on partitioning the free energy into a sum of enthalpic and entropic contributions. Additionally, end-point free energy methods that use implicit solvent approximations reduce computational demands even further. In such methods, the solvent is represented as a continuous medium instead of individual explicit solvent molecules⁹⁵. Probably the most well-known end-point implicit solvent-free energy method is the Molecular Mechanics/ Poisson-Boltzmann Surface Area (MM/PBSA)⁹⁶. This approach can be used to evaluate docking poses, determining structural stability, and predicting binding affinities and hot spots. As a further advantage, MM/PBSA allows the analysis of individual contributions by means of free energy decomposition, which gives additional insights into the energetics of the investigated system.

In MM/PBSA, the free energy of a molecule is estimated as the sum of its gas-phase energy, the solvation free energy, and a contribution due to the configurational entropy of the solute. The gas-phase energy is approximated by the molecular mechanics' energy of the molecule, determined from a force field comprising bonded and non-bonded energy contributions. In the calculation of the solvation free energy, two contributions are considered, a polar and a non-polar one⁹⁶. For the polar contribution, the change in the free energy upon transfer of a charged molecule from gas-phase (modeled as a

homogeneous medium with dielectric constant $\epsilon=1$) to solvent (modeled as a homogeneous medium with $\epsilon=78$ or 80)⁹⁷ is estimated by an implicit solvent model. Using this method, the binding free energy is calculated using equation (9), where $G_{Complex}$, $G_{Protein}$ and G_{Ligand} are absolute free energies for the complex, protein, and ligand, respectively⁹⁸.

$$\Delta G_{bind} = G_{Complex} - G_{Protein} - G_{Ligand} \quad (9)$$

Equation 9 – General equation for calculation of the binding free energy

The free energy for any molecule could be divided into a contribution from the solute and a contribution from the solvent:

$$G = G_{solute} + G_{solvent} \quad (10)$$

Equation 10 – Free energy decomposition

The contribution from the solute is composed of the gas phase MM energy (E_{MM}) and the conformational entropy ($-TS$), commonly calculated by normal-mode analysis⁹⁹. E_{MM} can be obtained by averaging the molecular mechanics' energy from the collected snapshots during an MD simulation, and it includes the ΔE_{int} comprising bond, angle, and dihedral energies, the ΔE_{ele} representing the electrostatic interaction energy and the ΔE_{vdW} representing the van der Waals interaction energy. The contribution from the solvent is the sum of the electrostatic solvation energy, both the polar contribution ΔG_{PB} and the non-polar solvation component ΔG_{SA} . The polar contribution is calculated using the PB model (described in literature¹⁰⁰), while the non-polar contribution is estimated by Solvent Accessible Surface Area (SASA).

$$G_{solute} = E_{MM} - TS \quad (11)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (11.1)$$

Equation 11 – Binding energy of the solute and decomposition of MM energy.

$$\Delta G_{sol} = \Delta G_{PB} + \Delta G_{SA} \quad (12)$$

$$\Delta G_{SA} = \gamma \cdot SASA + b \quad (12.1)$$

Equation 12 – Solvent contribution for the binding free energy

To compute all the ensemble averages in Equations (7-9) for the binding affinity calculations, the MM/PBSA method can be performed using either a single trajectory

approach using an MD simulation of the entire system or the multi-trajectory approach using an independent MD simulation for the complex, receptor, and ligand. A snapshot of a structure is taken at different points in the production run of the MD simulation, and these snapshots are then used to calculate the average values and thermodynamic quantities, namely the interaction free energy. In the single trajectory strategy, the ΔE_{int} is canceled between ligand, protein and complex. The non-polar contribution, ΔG_{SA} , is calculated from the SASA, with the surface tension proportionally constant γ and b generally set to $0.0072 \text{ kcal} \cdot \text{mol}^{-1} \text{ \AA}^{-2}$ and $0.00 \text{ kcal} \cdot \text{mol}^{-1}$ respectively⁹⁷.

Computational Methods

DszC is one of the key enzymes in the 4S pathway, because it has been referred to as one of its kinetic bottlenecks. X-ray crystallography studies have provided essential structural and mechanistic insights. DszC consists of a homodimer, with one catalytic site per monomer. The catalysis takes place close to the interface of each monomer and involves the molecular oxygen, the FMNH₂ cofactor, the DBT substrate, and the respective monomer residues (His92, Tyr96, Asn129, Ser163, His388, and His391).

It is also known that DszC is inhibited by 2-HBP, the final product from the 4S pathway, by a negative feedback mechanism. We now set up to develop mutants that are less affected by the negative feedback mechanism. Additionally, ideally, the proposed mutants should not alter the catalytic efficiency of the enzyme.

Structure Selection and Preparation

The system was modeled from an X-ray structure of FMN-bound DszC¹⁰¹. Several structures of the DszC enzyme exist in the PDB Database. A general comparison of the existing structures was performed, where parameters such as resolution, mutations, domain completeness, and presence of ligand, cofactor, and crystal waters molecules were assessed. The selected crystal structure (PDB ID: 3X0Y) contained two tetramers in the asymmetric unit, with four equivalent catalytic sites. Each tetramer was composed of two catalytically competent homodimers with 412 amino-acid residues per monomer. Since DszC is active as a homodimer, only two chains, comprising a homodimer, were selected. The resulting protein complex was then submitted to the PROPKA webserver, which utilizes a very fast empirical method to predict pK_a values of titrable residues and is successful at predicting unusual pK_a values¹⁰². A recent study showed that PROPKA was the most accurate method overall, to predict protein's pK_a results¹⁰³. This software uses a heuristic method to compute the pK_a perturbations due to desolvation, hydrogen bonding, and charge-charge interactions. However, in the current versions, PROPKA does not consider the contribution of heteroatoms such as bound ions or ligands for pK_a determination. The Xleap module of AMBER18 was used to protonate the protein, accordingly to the PROPKA results. Additionally, a visual inspection of each residue was performed in VMD using as reference the physiological pH value. Special attention was given to the histidine residues because those can be protonated to give rise to two different tautomers. The decision of which histidine tautomer to model was made by taking into consideration the surrounding chemical environment of each histidine. The proton was preferably placed on the N atom with the most hydrogen bond acceptors

within a 3 Å radius. If none existed, then the unprotonated N would be the one with more hydrogen bonds donors in its vicinity. In cases where both were true, the hydrogen bonds that were closer and those with bond angles closer to 180° were prioritized. If those criteria were not enough for an informed decision, an assessment of the chemical environment around the heavy atoms (O and N) surrounding the His residue was performed, and the same criteria as described above were applied. For His residues closer to the surface that could not be fit in the criteria described, solvent exposure was used as the determinant factor. The exception was for His residues 92, 388, and 391, for which the protonation state was already described in the work of Barbosa et al⁵³.

Parametrization of the C^{4a}OOH, FMNH₂, DBT, 2-HBP, and HBPS

As mentioned above, the AMBER force field contains a set of parameters for amino acids and nucleic acids. However, small molecules such as ligands must be parametrized in the force field before any simulation can be performed. The parameters for the reduced FMN and the C^{4a}OOH intermediate were already available from a previous work developed in the group⁵³. Parameters for 2-HBP and HBPS molecules were taken from the General AMBER Force Field (GAFF). We followed the parametrization procedure of the AMBER force field. Coordinates for 2-HBP and HBPS were determined from a geometry optimization in vacuum at the HF/6-31* level. The atomic charges were calculated at the same level of theory with the restrained electrostatic potential (RESP) fitting, using the antechamber tool.

Modeling of DBT

It is known that the inhibition of the DszC enzyme by the 2-HBP and HBPS molecules is non-competitive, as it is observed a reduced $V_{\max}$ without change in K_m as inhibitor concentration is increased.⁶⁵. However, it is not known whether the inhibition occurs before or after the entrance of the substrate into the catalytic site. Hence, the substrate was placed on only one of the two available catalytic sites of the DszC dimer, creating a “hybrid” model. This allows the study of the inhibition on different protein states (*holo* and *apo*). The catalytic site containing the DBT simulates the inhibition at the rate-limiting step, and the catalytic site containing only the FMNH₂ cofactor simulates the inhibition by blocking the entrance of the substrate into the catalytic site. The DBT substrate was then modeled into the DszC enzyme, after superposition of a thermophilic dibenzothiophene desulfurizing enzyme (PDB ID: 5XDE) to our model DszC homodimer, resorting to the *Pymol* software^{104–106}.

Docking of 2-HBP and HBPS

The objective of this work is to assess the mechanism of inhibition of DszC by the 2-HBP and HBPS. Therefore, the presence of those molecules is of the utmost importance to this work. As there are no structures of DszC or other similar enzymes with 2-HBP or HBPS bound, it was necessary to dock those molecules on the protein surface. In addition, there is no information available as to where the interaction between the inhibitors (2-HBP and HBPS) and the enzyme may occur. So, in order to obtain knowledge on possible binding sites in the protein surface, a blind docking assay was performed alongside a run in the DEPTH webserver.

Blind docking refers to the process of enlarging the searching stage to the whole protein surface, due to lack of prior knowledge of the location of the binding pocket. This method involves several trials/runs and several energy calculations before a favorable proteinligand complex is found. Additionally, when it comes to blind docking, most of the docking tools are quite limited¹⁰⁷. In this work, the Autodock Vina Software was used to perform the blind docking simulations. Several attempts were made by changing three variables in the input instruction file, to control parameters, such as the energy difference between the best binding pose and the worst and the maximum number of different poses generated.

Classical MD simulations of the complex

Both the 2-HBP and HBPS molecules were docked on the same protein structure, however, further studies were not conducted in a system where both molecules existed simultaneously. That is due to the fact that in a simulation conducted in a system containing both the 2-HBP and HBPS molecules, they would be competing with each other for the same binding cavity; the discovery of the binding affinity of each molecule is outside of the scope of this work and has been described before⁶³. Hence, different systems were created where DszC was complexed with either 2-HBP or HBPS molecules.

The procedures mentioned below are transferable for both.

System preparation

After docking the 2-HBP/HBPS molecules in the protein surface, the best cavities were selected based on the abundance of poses *per* cavity and the binding energy of each pose. Additionally, cavities that bound both molecules were also selected. Those cavities were then divided into two systems for each inhibitor type (4 systems total), distributed as shown in Table 1 below. The diluted systems contain the cavities that are shared between both molecules. Additionally, a control system was created without any 2HBP/HBPS molecules, simulating the protein behavior in the absence of product/inhibitor molecules.

Table 1 - Summary of the systems constructed for each inhibitor and the number of molecules in each system.

	Concentrated	Diluted	Control
2-HBP	5	2	0
HBPS	4	2	0

The decision of dividing the different cavities into different systems was made because it allowed the study of a greater number of cavities as well as the effect of product concentration on protein inhibition. The difference in the number of poses in the concentrated system with 2-HBP and HBPS is because docking of 2-HBP revealed an additional interesting binding pocket than docking of HBPS.

Once divided, the different systems were loaded into Xleap, along with the necessary FF modifiers describing the small molecules within the system. A total of 33 Na⁺ ions (35 or 37 for the systems containing HBPS) were added to neutralize the system global charge. The resulting complex was surrounded by a rectangular box of TIP3P water molecules

distancing a minimum distance of 16 Å from the solute. Note that the distance between the edge of the box and the protein exceeded the usual 12 Å, because the system may contain small molecules that are not bound nor interacting strongly with the protein and may diffuse in solution.

Minimization, annealing, equilibrium, and production protocols

Once all the systems were modeled, solvated, and neutralized via *Xleap*, they were submitted to energy minimization. This method is used to find a set of coordinates representing the minimum energy conformation for a given structure. Apart from its role in finding the nearest local minima, energy minimization is also an indispensable tool in correcting structural anomalies such as bad stereochemistry or short contacts. The energy minimization was performed in four steps, as follows: (1) an energy minimization was first applied to the solvent molecules, with restrains on all other groups of the system; (2) restraints were removed from the small ligands (FMNH₂, C_{4a}OOH, DBT, and 2HBP/HBPS), minimizing their energy; (3) next, the energy of the protein side chains was minimized; and (4) the energy of the entire system was minimized. Step (1) was performed using the steepest descent algorithm, while steps (2), (3), and (4) were performed with the conjugated gradient, as this methodology provides the best and more accurate results.

In the annealing step, energy is provided to the system by increasing its temperature. In this step, the systems were heated from 0 K to 300 K during 100 ps, using the Langevin thermostat, with constant *NV*. Next, the systems were submitted to a small 20 ns equilibration in the *NPT* (300 K, 1 bar) ensemble, before proceeding to the production run. The change to the *NPT* ensemble was necessary to best mimic the conditions in which the inhibition should happen in nature, as most chemical reactions occur under constant pressure. After transitioning to the *NPT* ensemble, the resulting structures were submitted to an additional equilibration. In this step, two different protocols were applied to test out different conditions, as well as produce additional sampling. The steps used in the previously mentioned protocols were:

- (1) Equilibration of the solvent over 10 ns, with restrains on the remaining groups;
- (2) The restrains on the protein were removed and the system simulated for an additional 10 ns;
- (3) All remaining restrains were removed, and the entire system was simulated for 10ns.

In protocol 1, steps (1) and (3) were used. For protocol 2, all three steps were used. Hence, all systems submitted to protocol 1 were equilibrated for a total of 20 ns, and systems submitted to protocol 2 were simulated for a total of 30.

We argue that keeping the 2-HBP/HBPS molecules restrained and releasing the protein, on protocol 2, should favor a hypothesis where the protein is more adapted to the presence of the inhibitor product on its surface. Afterward, all systems were submitted to 200 ns production runs in an *NPT* ensemble (300 K, 1 bar).

Results and Discussion

Binding mode identification for 2-HBP and HBPS

Structure selection and preparation

Since it is described in literature that the DszC enzyme is active as a homodimer, only two chains were selected from the PDB file obtained. The additional chains were removed in order to reduce the computational power required to simulate the system. Hence, only chains A and F were selected for further analysis making up a total of two catalytic sites (Figure 9).

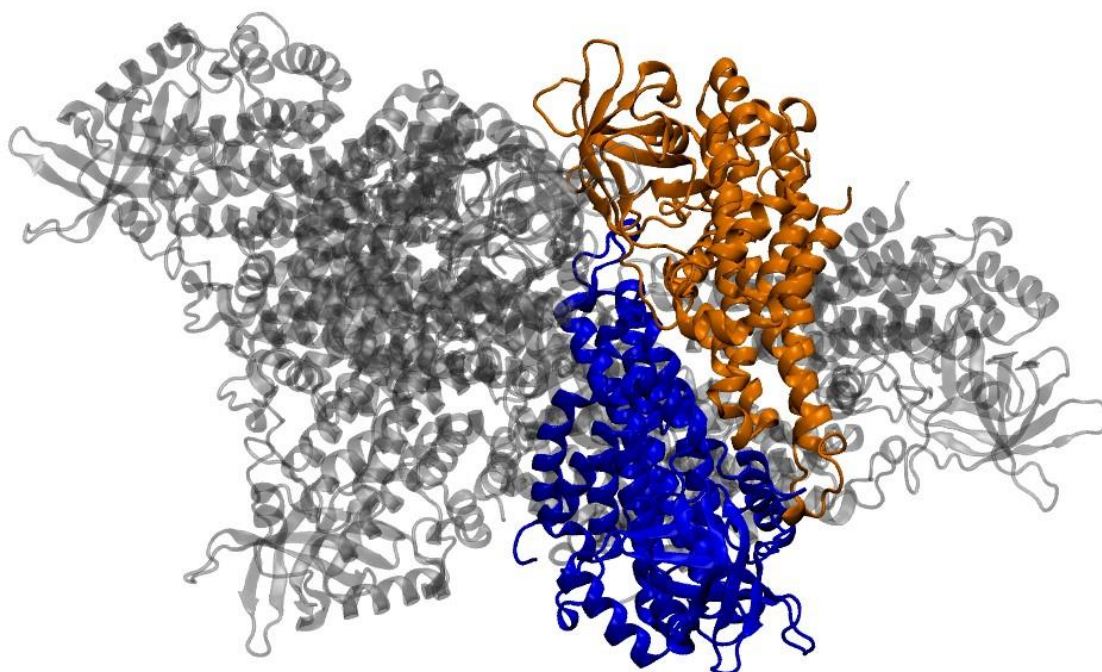


Figure 8 - Representation of the two chains selected from the DszC enzyme structure, PDB ID: 3x0y. Chains A and F are colored in blue and orange, respectively.

The two chains were submitted to the PROPKA webserver in order to analyze their residues' pK_a values. PROPKA returns a downloadable file with statistics on residue, hydrogen bonding, location, solvent accessibility, and Coulombic interactions. An example of the file can be seen in Figure 10. Based on the results given by PROPKA, the Xleap module of AMBER was used to protonate the protein, assuming a physiological pH. PROPKA was then used to verify the protonation state of the residues. As mentioned in the Computational Methods section, histidine residues were given special attention, as these residues can easily form different species upon protonation.

A list of each tautomer is depicted in Table 2.

Table 2 - Below are depicted the histidine residues of chain A and which tautomer was chosen for each. Only one chain is represented because the DszC structure selected is a homodimer, and therefore the two chains are equal.

His_δ	His92, His391
His_ε	His97, His115, His137, His160, His221, His 267, His321, His373, His388, His404, His391

For histidine residues 391, 388, and 92, the protonation was made based on the work described by Barbosa *et al*⁵³.

Parametrization of small ligands

All atoms in the system should be correctly parameterized before running any MD simulation. Extensively validated amino acid parameters are already available in the AMBER force field. In addition, the reduced FMN, C4aOOH, and DBT molecules have been already parametrized in previous works by the group. However, the end-products of DszA and DszB, HBPS and 2-HBP, respectively, needed to be parameterized. The atomic charges were calculated at the *ab initio* HF level, with the 6-31G(d) basis set and the remaining parameters were determined using the ANTECHAMBER module from the AMBER package and using the GAFF force - field. The parameterized atomic charges are presented in Table 3.

Table 3 – Parameterized charges for each atom of 2-HBP (right) and HBPS (left).

2-HBP			HBPS		
Atom name	Atom type	charge	Atom name	Atom type	charge

O1	oh	S1	s4	0.254
C1	ca	O1	oh	-0.621
C2	ca	O2	o	-0.553
C3	ca	O3	o	-0.553
C4	ca	C1	cp	0.376
C5	ca	C2	cp	-0.360
C6	cp	C3	ca	-0.151
C7	cp	C4	ca	-0.319
C8	ca	C5	ca	0.455
C9	ca	C6	ca	0.011
C10	ca	C7	ca	0.111
C11	ca	C8	ca	-0.102
C12	ca	C9	ca	-0.324
H1	ho	C10	ca	-0.329
H2	ha	C11	ca	-0.250
H3	ha	C12	ca	-0.103
H4	ha	H1	ha	0.148
H5	ha	H2	ha	0.140
H6	ha	H3	ha	0.091
H7	ha	H4	ha	0.117
H8	ha	H5	ha	0.165
H9	ha	H6	ha	0.151
H10	ha	H7	ha	0.137
		H8	ha	0.124
		H9	ho	0.388

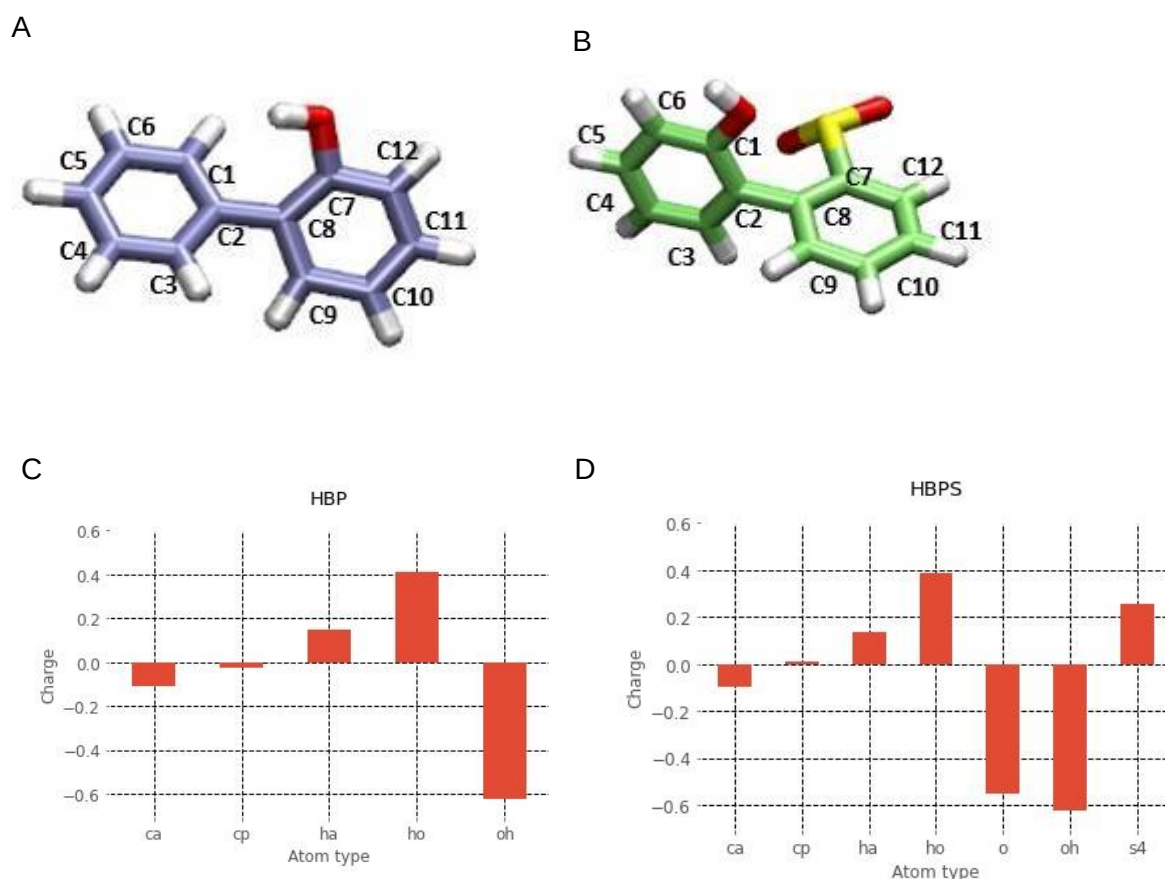


Figure 9 – (A) and (B) Atom names for 2-HBP and HBPS, respectively, according to parametrization. (C) and (D) atomic charge distribution according to atom types for 2-HBP and HBPS, respectively, by averaging the values depicted in Table 3. ca - sp² C, aromatic; cp - bridge aromatic C.

The distribution of the computed atomic charges is depicted in Figure 11. Charge distribution is an important parameter as it dictates the type of interaction between these molecules and protein residues. Therefore, a correct parametrization of the small molecules is utterly important. As expected, the most negative charges were placed on the oxygen atoms in both the 2-HBP and HBPS molecules. Additionally, the HBPS molecule has a small partial positive charge on the sulfur atom. The total atomic charges for each molecule were 0 for the 2-HBP and -1 for the HBPS.

Modeling of DBT

Once the pK_a of all residues had been resolved, the DBT molecule, was modeled into the catalytic cavities of chain A and F. Because the inhibition mechanism of the DszC enzyme is not yet solved, it was decided that a “hybrid” model would answer a larger number of questions. The resulting model has one catalytic cavity composed of the DBT substrate and the C^{4a}OOH intermediate (Figure 13), which corresponds to the limiting

step of the reaction catalyzed by the DszC and one probable target in known inhibition mechanisms. Another possible target is the beginning of the reaction and precedes the entrance of the substrate into the catalytic cavity of the enzyme. This situation is depicted in the second catalytic cavity, where only the FMNH₂ cofactor is present (Figure 13). The “hybrid” model was based on the reaction mechanism previously proposed by Zhang and co-workers¹⁰⁸, and recently validated by Barbosa and co-workers⁵³ (Figure 12).

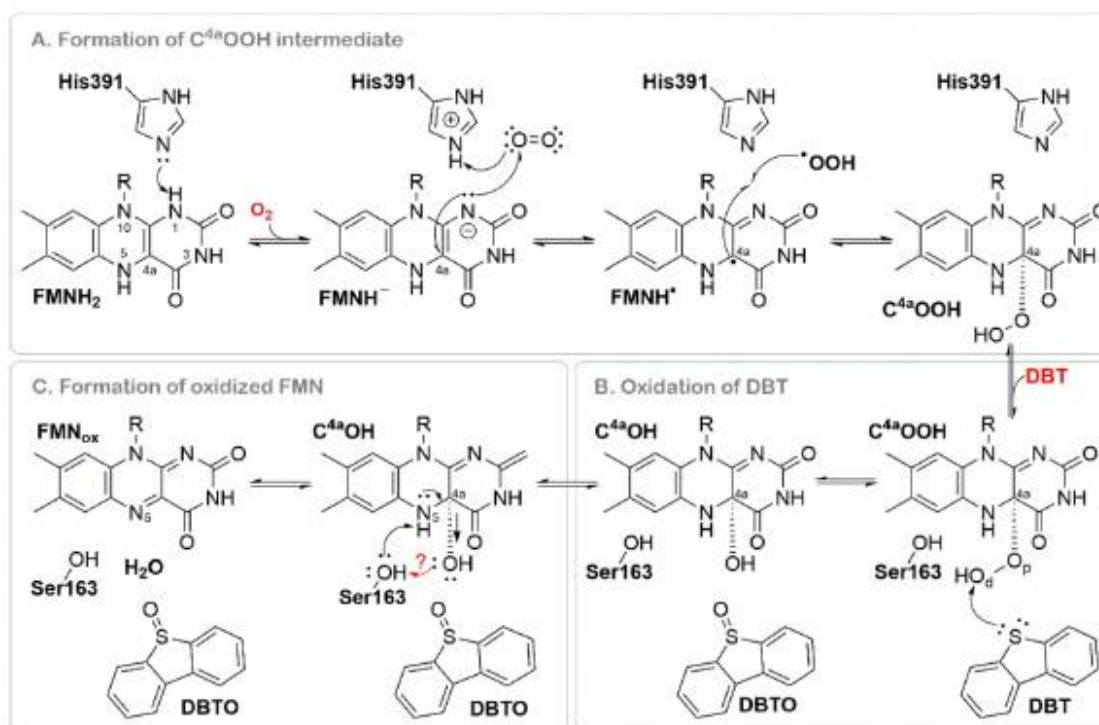


Figure 10 - Proposed mechanism of DszC. (A) Oxygen activation; (B) Oxidation of DBT by distal oxygen; (C) Regeneration of FMN. Image retrieved from Barbosa et. al⁵³.

It is important to notice that the inhibition by the 2-HBP/HBPS may affect any of the steps shown in Figure 10. However, simulating the different steps singularly would not bring any additional significant information, because the enzyme does not undergo any conformational changes as the reaction develops.

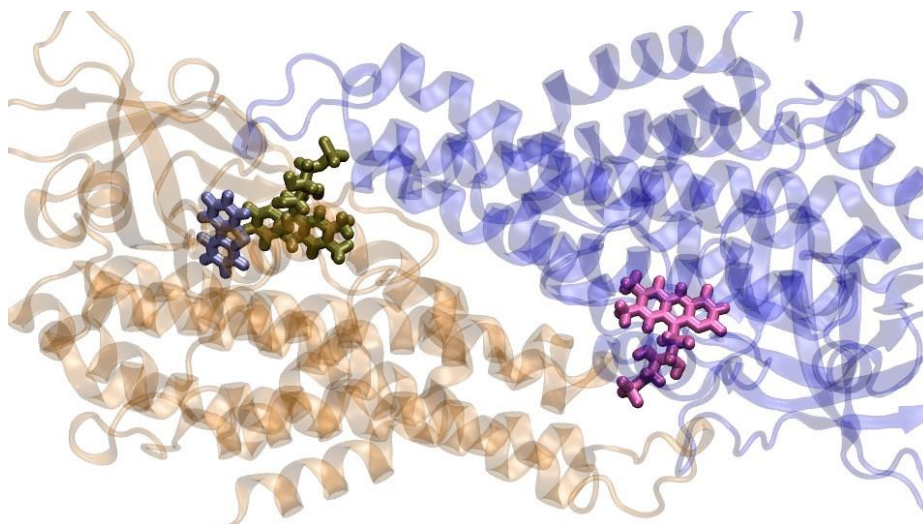


Figure 11 - Visual representation of the two catalytic sites modeled for the "hybrid" model. On the left, it is represented the catalytic site containing the DBT molecule (ice blue) and the C^{4a}OOH intermediate (tan). On the right, it is represented the catalytic site containing only the reduced FMN (magenta). In Cartoon, the chains A (blue) and F (orange) are represented.

Docking of 2-HBP and HBPS

As mentioned above, the 2-HBP and the HBPS molecules inhibit non-competitively the DszC complex. However, the binding site for either molecule is unknown. Therefore, a blind docking run was performed using both molecules to generate binding poses. Moreover, because a non-competitive binding can also include binding of the inhibitor inside the catalytic site (on a different position than the substrate), another docking run was performed, where the inhibitor molecules were docked inside the catalytic site, alongside the cofactor and substrate.

Catalytic site docking

To perform the docking of the 2-HBP and HBPS molecules inside the catalytic site, the GOLD software was used. A searching area of 8 Å around the cofactor was defined, and polar hydrogens were added to both the protein and the ligand. Non-polar hydrogens are merged with the carbon atoms to which they are bonded. The partial charge of each nonpolar hydrogen is added to the charge of its carbon, and the hydrogen is "removed" from the residue. This model simplifies docking by reducing the overall number of atoms to be considered.

Both inhibitors were docked in the catalytic sites of the DszC dimer: one binding the FMNH₂, and the other binding both the C^{4a}OOH intermediate and the DBT substrate. In non-competitive inhibition, the inhibitor reduces the activity of the enzyme and should bind equally well to the enzyme whether it is already bound to the substrate or not¹⁰⁹.

Hence, if the catalytic cavity was the binding site of 2-HBP or HBPS, it would be expected that the inhibitor would bind the catalytic site equally well in the presence or absence of the substrate. However, from Figure 14 it is possible to observe that when the DBT is inside the catalytic site, neither inhibitor molecule is able to enter the catalytic site. Without the DBT in the catalytic site, several different poses were obtained for both the 2-HBP and HBPS molecules. This indicates that the cavity where the inhibition of the DszC complex occurs is most likely not located inside the catalytic site.

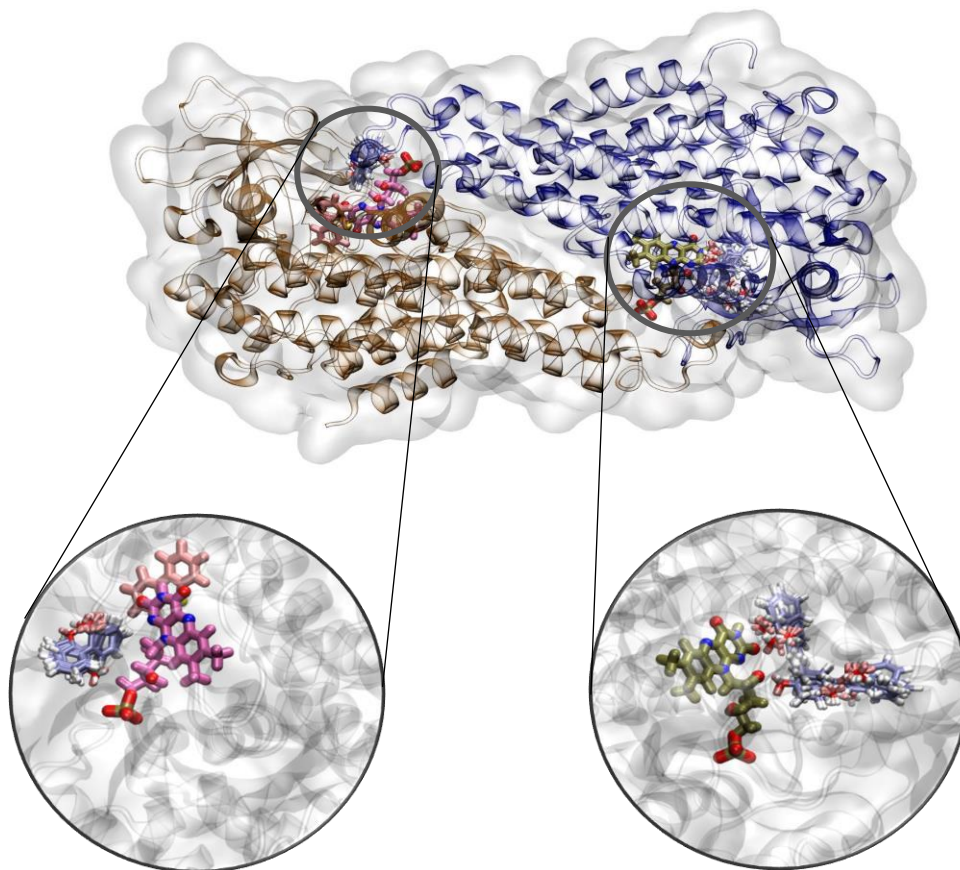


Figure 12 - Above, a representation of the entire complex, with chains A and F colored in orange and blue. Zoomed in are represented the 2-HBP molecules (cyan sticks) docked in the catalytic site in the presence (left) and absence (right) of the DBT substrate (shown in pink).

Blind Docking

Given the results from the docking of the 2-HBP and HBPS on the catalytic site, it is possible to assume that the catalytic site is unlikely to bind the inhibitors. That means that the inhibitory site is probably an allosteric site located elsewhere in the enzyme. Since there is no literature on the topic, blind docking was performed with both 2-HBP and HBPS.

Blind Docking 2-HBP

The blind docking results for the 2-HBP molecules suggest that the binding site is made up of mostly hydrophobic residues such as (Ile, Val, and Phe).

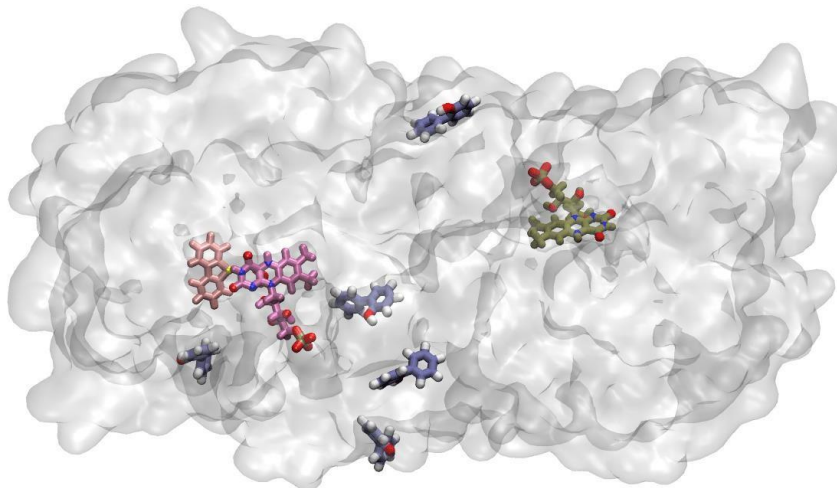


Figure 13 –Image showing DszC enzyme with the 2-HBP molecules spread across putative cavities on its surface. In sticks are displayed the FMNH₂ (brown), C^{4a}OOH (purple), DBT (pink), and the 2-HBP (cyan) around the protein.

This result was expected since the 2-HBP molecule is mostly non-polar. The most populated site was located at the entrance of the catalytic site. Since all poses presented similar binding energies (ranging from 5.4 to 6.5 kJ mol⁻¹), the cavities that bound the largest number of poses were prioritized. The table below summarizes the most populated sites, with a more in-depth analysis of the surrounding residues within 4 Å of the 2-HBP molecule, ordered by a decreasing number of poses docked in each. The sites described in the table below correspond to those that will be further investigated by MD simulations. The remaining sites were set aside due to a low number of docking poses.

Table 4 — Summary of the most representative binding sites found after docking the 2-HBP molecule, with detailed information about the residues in each cavity. In curved brackets is the information about the residue chain (only for the S4 site).

Designation	Chain	Residues
S1 & S5	F	ILE242; VAL138; ILE187; VAL238; PHE237; PHE241; PRO413; TRP141; PHE415; THR416; GLY414; ASN134; SER136;
S2	F	GLU275; ARG32; PHE378; TYR376; TYR276; PHE381; ALA272; ARG380; THR279;
S3	A	TYR376; TYR276; GLN280; THR279; ALA272; PHE378; ARG380; GLU275; VAL36;
S4	A & F	ARG(F)211; ARG(F)40; PRO(F)374; ASP(F)379; ARG(A)40; ARG(A)211; ILE(A)208; ALA(A)206; ALA(A)207

Blind Docking HBPS

Regarding HBPS, docking results revealed that most of the cavities found were composed of hydrophilic residues, such as Arg, Ser, and Asp. The composition of the cavities found was expected since the HBPS molecule is not only polar, but also carries a negatively charged sulfinate group. Similarly, to the procedure applied when choosing the appropriate cavities and poses for the study of the 2-HBP inhibition, the cavities that bound a higher number of poses were prioritized. The results are expressed in Table 5. and the poses can be seen in Figure 16.

Table 5 - Summary of the most representative binding sites found after docking the HBPS molecule, with detailed information about the residues in each cavity. In curved brackets is the information about the residue chain (only for the S'1 and S'4 sites).

Designation	Chain	Residues
S'1	A & F	ARG(A)370; HIS ϵ (F)160; PHE(F)161; SER(F)217; ASP(F) 204; TRP(F)205; ARG(A)275 ; HIS ϵ (A)373; ARG(A)375;
S'2	A	THR306; ARG274; ARG278; THR293; THR298; ILE299; TYR302;
S'3	F	ARG278; THR293; THR277; ILE299; TYR302; ARG274; THR306
S'4	A & F	GLU(A)304; SER(A)301; GLU(A)365; SER(F)395; ILE(F)398; VAL(F)394; THR(F)353; ASN(A)354

Since the DszC enzyme is inhibited by both the 2-HBP and HBPS molecules and both molecules are chemically similar, there is a chance that both those molecules bind the same pocket. Hence, all cavities that bound

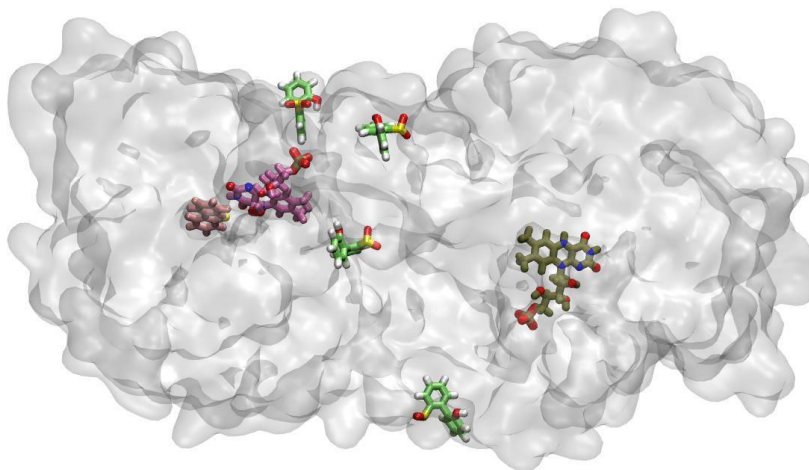


Figure 14 – A representation of the DszC enzyme with the HBPS inhibitor docked in putative cavities across the surface. In sticks a representation of the FMN (brown), DBT (pink), C^{4a}OOH (purple), and HBPS (green) around the protein.

both molecules, whether they were located in the same dimer or in the equivalent pocket of the homologous dimer, were selected for further analysis via molecular dynamics simulations.

The resulting complex containing the cofactors (FMNH₂ and C^{4a}OOH), the DBT substrate, and the 2-HBP and HBPS inhibitors are shown below.

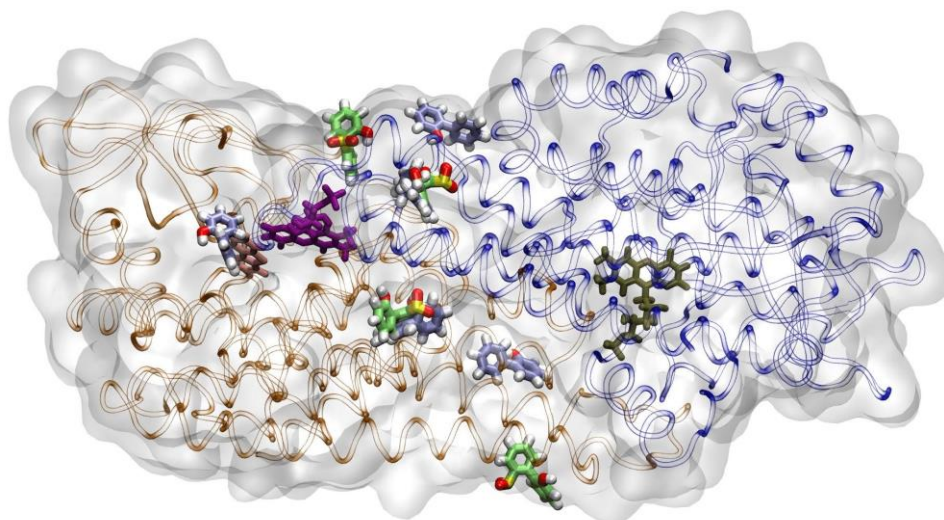


Figure 15 - Representation of the resulting system after selection of the best poses for the 2-HBP and HBPS. In surface, the DszC enzyme. Inside, it is possible to observe chains A (blue) and F (orange). In brown sticks, representation of the FMN cofactor; in purple, the C^{4a}OOH intermediate, and in pink, the DBT substrate. Docked in the enzyme surface are 2-HBP (blue) and HBPS (green) molecules.

Autodock Vina was used to dock the 2-HBP and HBPS molecules on putative cavities in the DszC enzyme's surface. However as mentioned previously, docking software does not always produce the most accurate results, especially in this case where the rigid body receptor approach was used. Hence, to validate the docked poses MD simulations of the systems were performed. Additionally, MD will allow the identification of the most relevant residues in the interaction between enzyme and inhibitor and how that binding may lead to inhibition, a piece of information that will be relevant to, ultimately, impair the feedback inhibition mechanism.

Characterization of the interaction between the 2-HBP and HBPS and DszC

Docking and selection of the best poses of 2-HBP and HBPS resulted in four different systems. For each inhibitor, two systems were created with different concentrations of the inhibitor molecule. Additionally, this procedure allows the testing of the concentration effect on enzyme inhibition. The number of inhibitor molecules in each system is depicted in Table 1.

Molecular Dynamics

The MD simulations were carried out on all four initial systems and a control system, without any inhibitor molecules. After applying the protocol described in “System preparation”, the root-mean-square deviation RMSd value was verified to confirm the systems’ structural stability (Figure 18). The figure below only depicts the systems submitted to protocol 1 (described in “Minimization, annealing, equilibrium, and production protocols”). For the systems submitted to protocol 2, the minimization and annealing steps shown in Figure 18 were shared by both protocols.

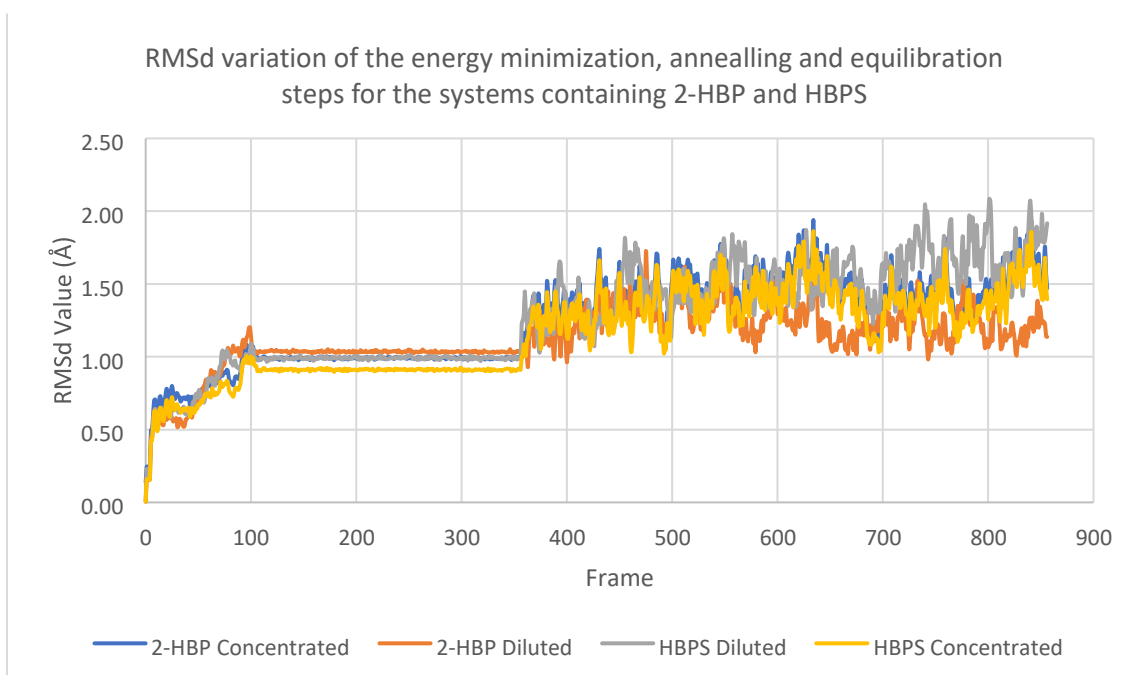


Figure 16 - Graphical representation of the RMSd of the protein backbone over time. In the xx axis it is represented the number of frames and on the yy axis the RMSd value. The first 5 frames correspond to the initial structure and the four steps of the energy minimization protocol. Frames 5 to 110 correspond to the annealing protocol. Frames 110 to 383, demonstrate the enzyme behavior during the restrained portion of the equilibration phase for 10 ns. The remaining corresponds to the RMSd value of the backbone during the unrestrained equilibration of 10 ns.

Analyzing Figure 16, it is possible to observe that all systems behaved as expected. The starting system (obtained from *Xleap*) was used as a reference for the RMSd measurement. Hence, all structures start with an RMSd value of 0 (zero). The linear portion of the graph (frame 5 to 115) is relative to the equilibration performed with restrictions on the enzyme backbone. Once those restrictions were lifted (frame 383) it is possible to see an increase in RMSd fluctuation. Even though the RMSd value does not give much information on the system’s behavior, it is possible nonetheless, to conclude that there are no relevant structural displacements occurring during the preparation of the systems for simulation.

Further information about the composition of each system is depicted in Table 6.

Table 6 – Composition of each simulated system in terms of Inhibitor molecules, water molecules, cofactor, intermediate state, and substrate (included in the Additional Ligands row).

Components \ Systems	2-HBP		HBPS		Control
	Concentrated	Diluted	Concentrated	Diluted	
Number of Inhibitors	5	2	4	2	0
Number of water molecules	29824	29529	30332	30752	41709
Additional Ligands	DBT, FMNH and c ^{4a} OOH	DBT, FMNH and c ^{4a} OOH	DBT, FMNH and c ^{4a} OOH	DBT, FMNH and c ^{4a} OOH	DBT, FMNH and c ^{4a} OOH

Each system was submitted to two different protocols of equilibration, and each protocol was submitted to a 200 ns production MD simulation. Overall, a total of nine systems were simulated (four containing 2-HBP; four containing HBPS, and a control system). Once again, relevant structural displacements were rejected by the RMSd analysis in Figure 19. The graphical representation of the RMSd values is only depicted for the concentrated and control systems. The information regarding the remaining diluted systems can be found in Figure 50 in the Annexes section.

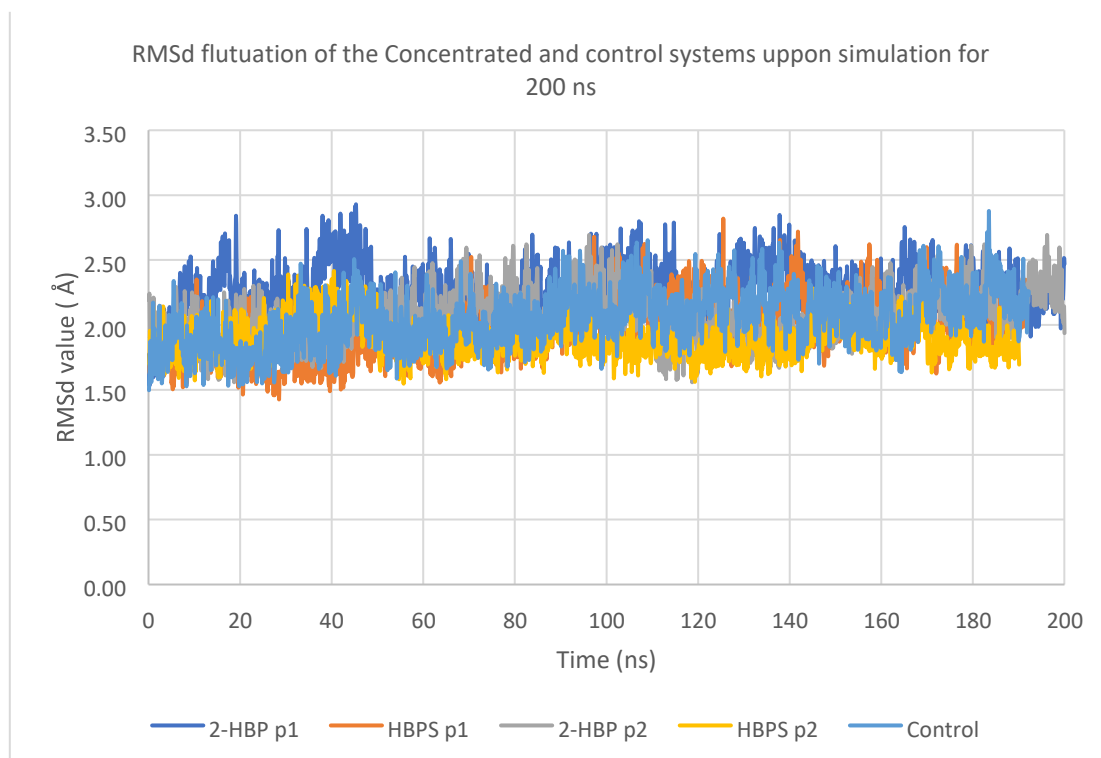
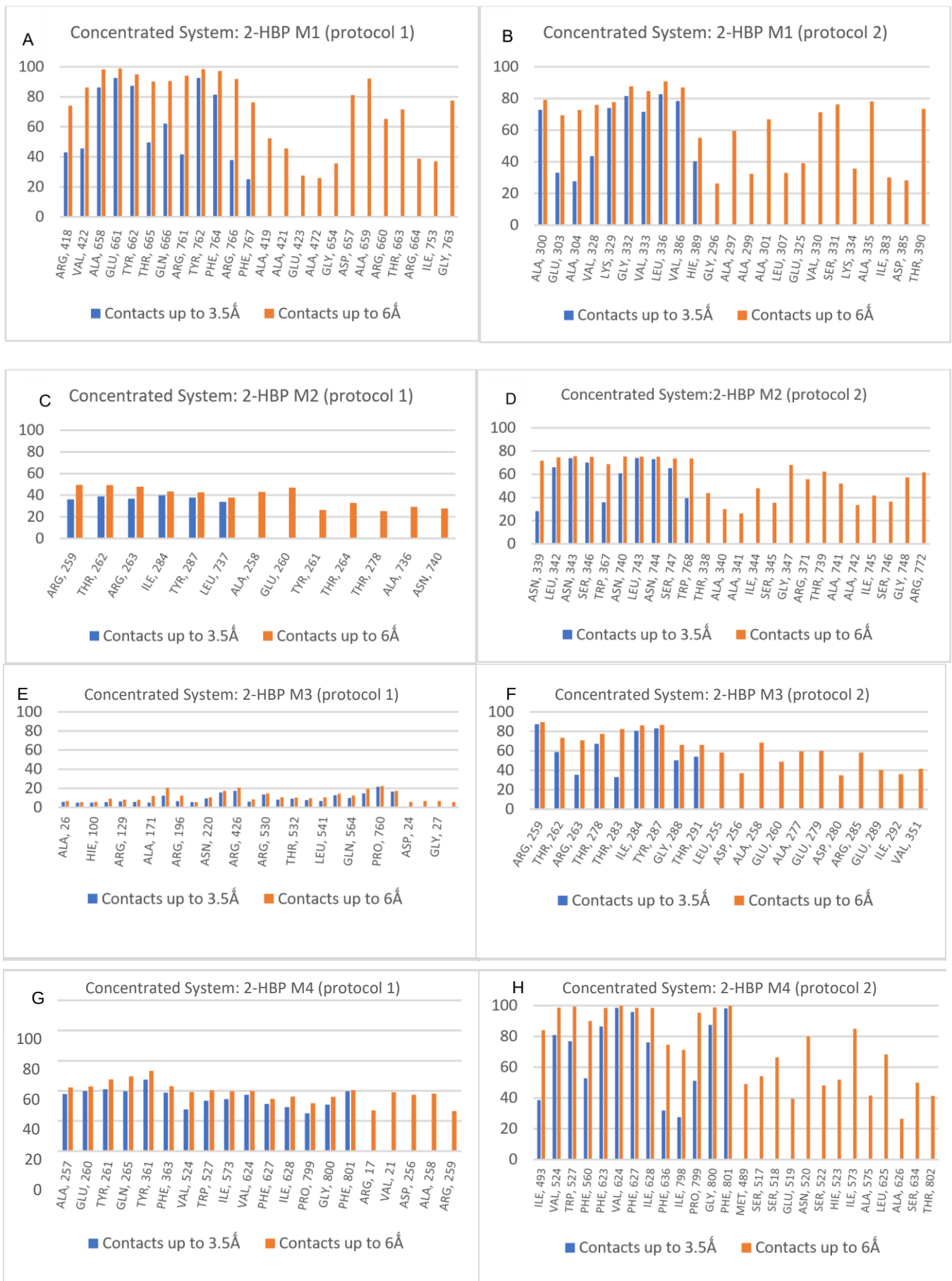


Figure 17 – Graphical representation of the backbone RMSd fluctuation of the concentrated and control systems upon 200 ns simulation. In the legend, “p1” and “p2” refer to systems submitted to protocol 1 and protocol 2, respectively.

Contact Analysis

Having simulated the different systems, it became important to characterize the interaction of the inhibitors with the DszC enzyme. To that effect, the movement of each inhibitor in each system was followed and any contact established with enzyme residues within 3.5 or 6.0 Å was registered. The results were then expressed in a graphical manner, by percentage of contacts per residue and averaged over the simulation time (Figure 18). The 3.5 and 6.0 Å were chosen for contact analysis since those were considered as proper maximum limits for close-range (electrostatic interactions) and long-range (van der Waals interactions), respectively.



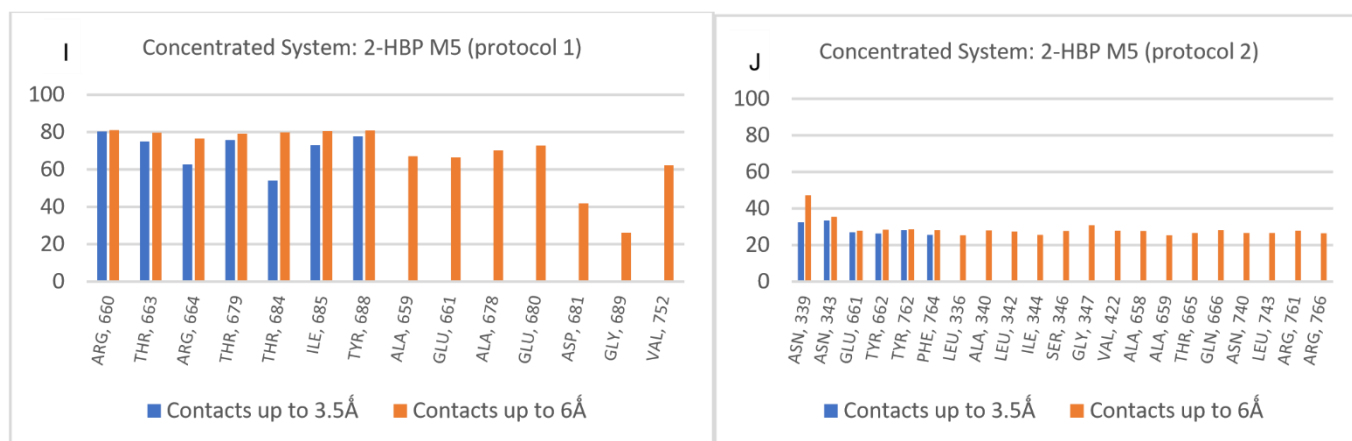
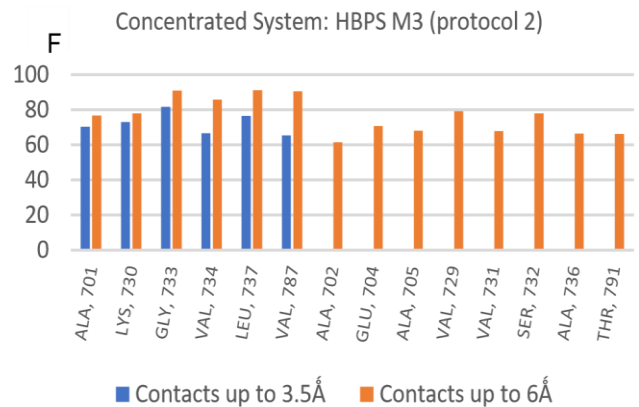
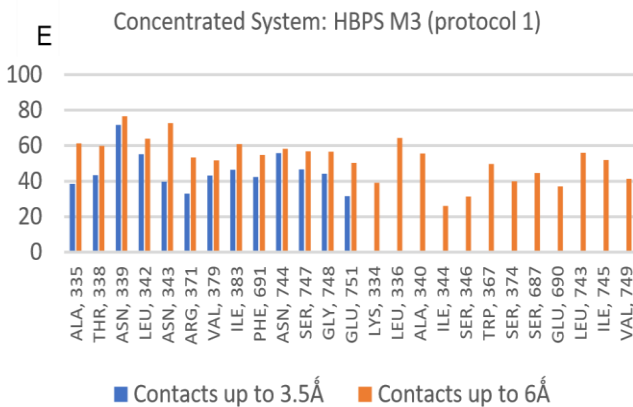
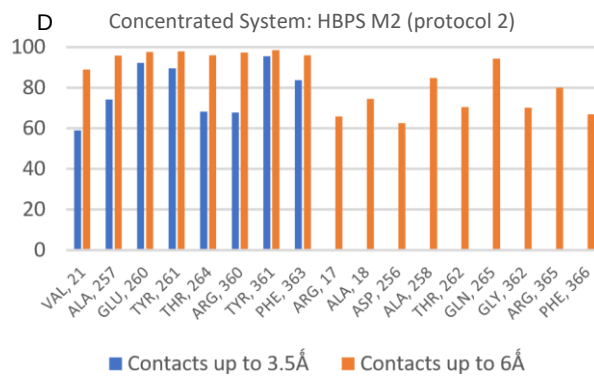
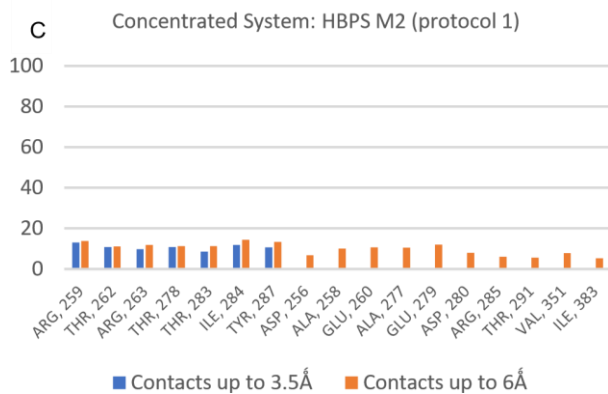
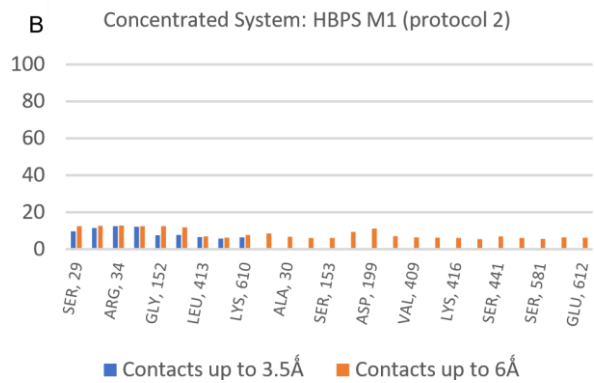
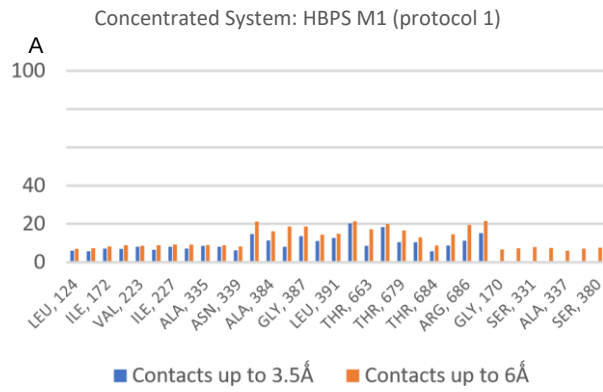


Figure 18 – Graphical representation of the percentage of contacts between the 2-HBP inhibitor and the DszC enzyme in the concentrated systems, submitted to protocol 1 (left) and 2 (right). In blue are represented the contacts up to 3.5 Å and in red the contacts up to 6.0 Å,

From the analysis of Figure 18, it is possible to see that most of the inhibitor molecules present in the concentrated system submitted to protocol 1 remained in contact with the DszC enzyme. Out of the 5 molecules present in the system, one (2-HBP M3) lost interaction with the enzyme, having interacted with it for approximately 25% of the simulated time. 2-HBP molecules M2 and M4 interacted with DszC for about 40% of the simulated time. The best poses were 2-HBP molecules M1 and M5 that remained in the same cavity for more than 60% of the time. Results from protocol 2 show overall, the 2-HBP molecules had a higher contact percentage. The only exception was 2-HBP molecule M5 which had higher contact percentage when submitted to protocol 1. A more in-depth comparison of the poses obtained from both protocols will be performed below.



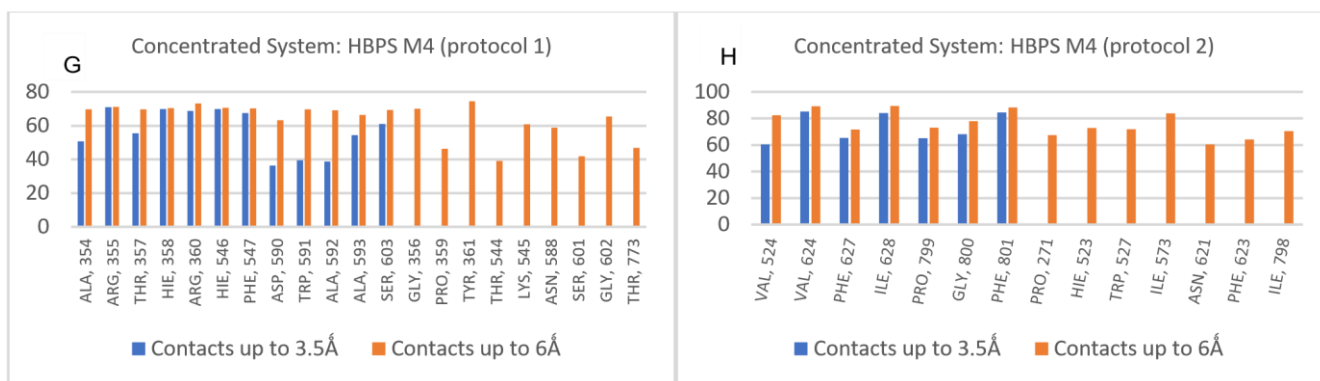


Figure 19- Graphical representation of the contacts between HBPS inhibitor and the DszC enzyme in the concentrated systems, submitted to protocol 1 (left) and 2 (right). In blue represented the contacts up to 3.5 Å and in red contacts up to 6.0 Å, representative of the close range and long-range interactions, respectively.

Analysis of Figure 19 shows that out of the four HBPS molecules present in the concentrated system submitted to protocol 1, two (HBPS molecules M1 and M2) did not show any relevant interaction with the DszC enzyme (under 25% contact percentage over time). However, the HBPS molecules M3 and M4 remained within the same binding cavity for an extended period of time (over 60%). As observed for 2-HBP molecules, HBPS shows higher contact percentages with DszC when submitted to protocol 2. The exception happened for the HBPS molecule M1, which had an even lower contact percentage than the one observed in protocol 1. In addition, a bad interaction between the HBPS molecule and the DszC enzyme in both protocols 1 and 2 may be a good indicator that the cavity in which the molecule was placed may not be suitable for binding.

Regarding the diluted systems, which possessed 2 molecules of either 2-HBP or HBPS each (see Table 6), it is possible to observe that the behavior of the inhibitor molecules was similar to the behavior observed for the concentrated systems. Molecules from protocol 2 displayed higher contact percentage overall.

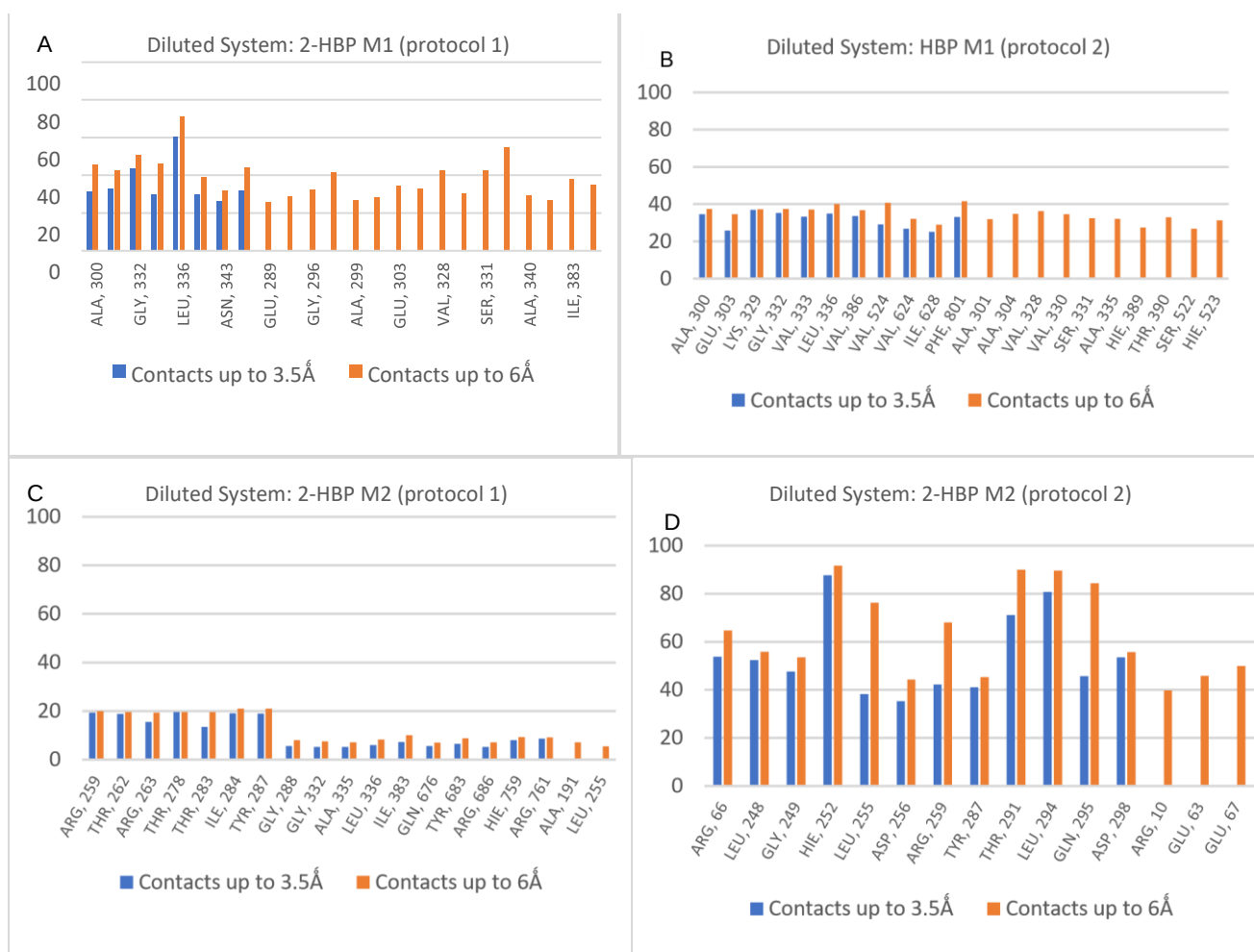


Figure 20 - Graphical representation of the contacts between the 2-HBP inhibitor and the DszC enzyme in the diluted system, submitted to protocols 1 (left) and 2 (right). In blue represented the contacts up to 3.5 Å and in red contacts up to 6.0 Å, representative of the close range and long-range interactions, respectively.

As shown in Figure 20, of the two 2-HBP molecules in the diluted system, 2-HBP M1 shows greater affinity (approximately 50% contact percentage over time) for its cavity compared to the 2-HBP molecule M2 (approximately 20% contact percentage over time). Molecules from the systems submitted to protocol 2 show in comparison to the 2-HBP molecules from the system submitted to protocol 1, higher contact percentages for both molecules. It is important to notice that the residues interacting with 2-HBP molecule M2 in the system submitted to protocol 1 are different than the one interacting with its counterpart from protocol 2.

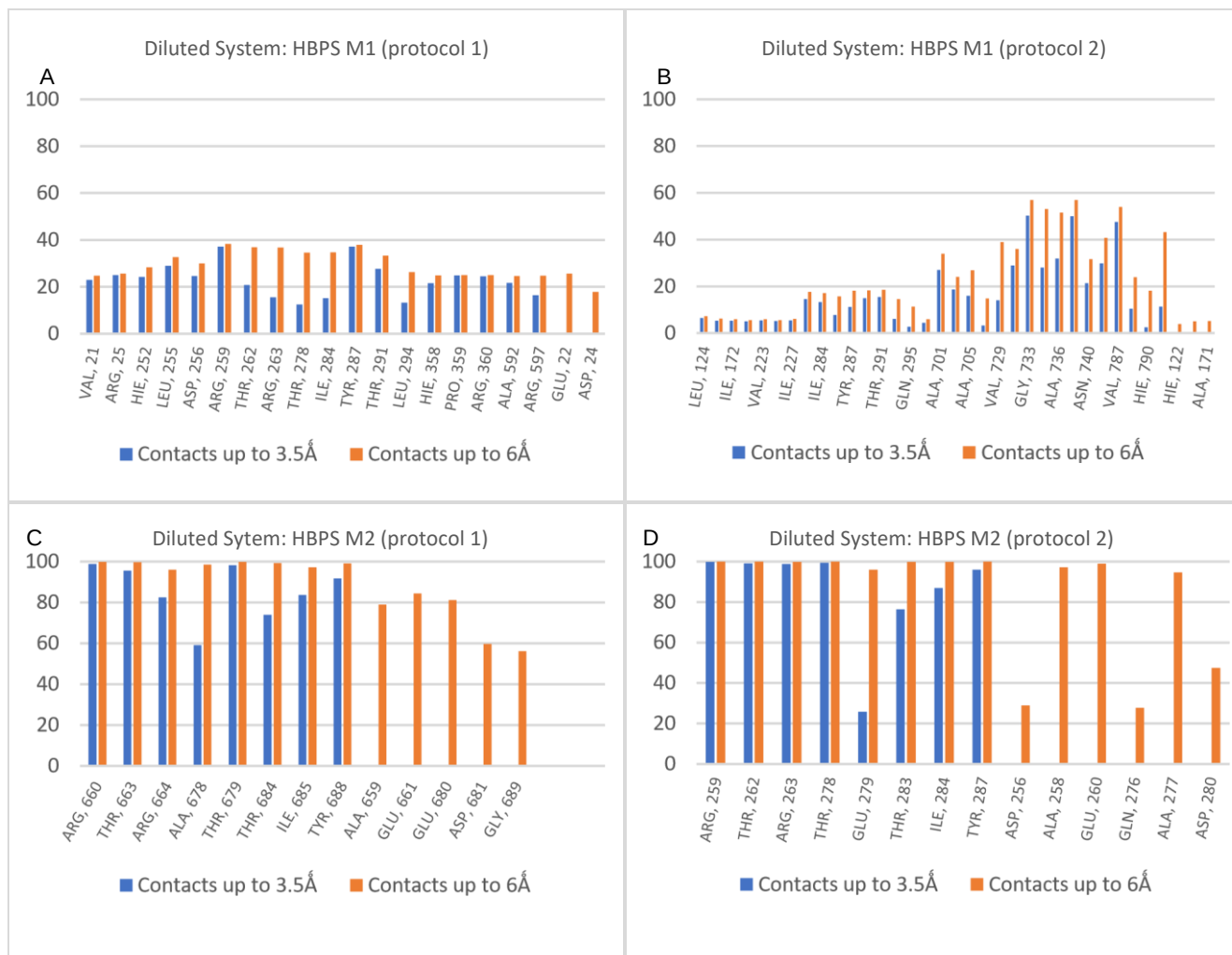


Figure 21 - Graphical representation of the contacts between HBPS inhibitor and the DszC enzyme in the diluted system, submitted to protocol 1 (left) and 2 (right). In blue represented the contacts up to 3.5 Å and in red contacts up to 6.0 Å, representative of the close range and long-range interactions, respectively.

Analysis of Figure 21, containing information regarding the HBPS inhibitors from both protocols 1 and 2 of the diluted systems, it is possible to see that the HBPS molecule M1 has a low affinity for the cavity in which it was docked, with contact percentages of approximately 40% and 50% in protocol 1 and 2, respectively. HBPS molecule M2 shows high contact percentages in both protocols, even though the residues with which it interacts are not the same.

Submitting the different systems to either protocol 1 or protocol 2 had several effects on the behavior of the inhibitor molecules. As mentioned previously, protocol 2 should allow the enzyme to adapt better to the presence of the inhibitor molecule on its surface. Hence, most inhibitor molecules on protocol 2, in general, showed a higher contact percentage than their counterparts on protocol 1. The major differences between both protocols are explored further below. Crossing information regarding the behavior of the inhibitor molecules in both protocols provided valuable information allowing for a more accurate and valid binding pocket selection. Once the results were obtained, a

comparison of the inhibitor poses was performed against the poses obtained after the docking. This was used as a measurement of the docking accuracy. For the poses found through protocol 1, most remained within 6 Å of the predicted binding site. However, for the poses obtained through protocol 2, most occupied different binding sites than the one predicted by the docking software. The differences found between both protocols can be explained by the lower repulsive forces felt by the inhibitor molecules that may have been caused by close distances resulting from the docking. Better accuracy of results could probably have been obtained using a flexible receptor, or an ensemble docking strategy; however, discussing docking approaches, efficacy and accuracy are outside the scope of this work.

Selection and characterization of the most promising binding sites

Having characterized the interactions of each inhibitor with DszC's residues, the contact percentage over time, and the overall behavior of each inhibitor molecule, it is now important to characterize more extensively the most promising binding sites. To understand which binding sites were more relevant and which were not, a crossreference of all data was performed. To that effect, the data regarding the residues interacting with each inhibitor molecule was filtered based on four distinct criteria:

- (1) whether or not the contacts existed in both protocol 1 and 2;
- (2) the abundance/contact percentage higher than 80%;
- (3) whether the contacts existed in both the concentrated and diluted system;

and (4) whether the binding site was shared by both the 2-HBP and HBPS molecules.

This latter approach was taken since there is a probability that the 2-HBP and HBPS molecules bind the same binding pocket within the protein.

Common binding sites between protocol 1 and 2

As it is possible to observe in the previous section, not all residues that interacted with an inhibitor from protocol 1 interacted with its counterpart from protocol 2. Hence, a good approach to validate a given binding site is to establish whether an inhibitor was able to maintain interaction with the enzyme in both protocols. It was mentioned earlier that by allowing the protein to adapt to the inhibitor on its surface, the repulsive forces acting on that inhibitor should be lower. It was also verified that the same procedure allowed for

the protein motifs in which the inhibitor was initially docked to rotate and promoted a more conserved interaction with the inhibitor. Hence, whether by rotation of the protein motifs or by dispersion of the inhibitor in solution, most of the binding regions found in protocol 1 are different from those found in protocol 2. Hypothetically, the most promising binding sites are those that maintained interactions in both protocols. For some systems, the binding regions are the same even though the inhibitor molecules that bind are not. In Figure 24, the common residues found in protocols 1 and 2 are depicted.

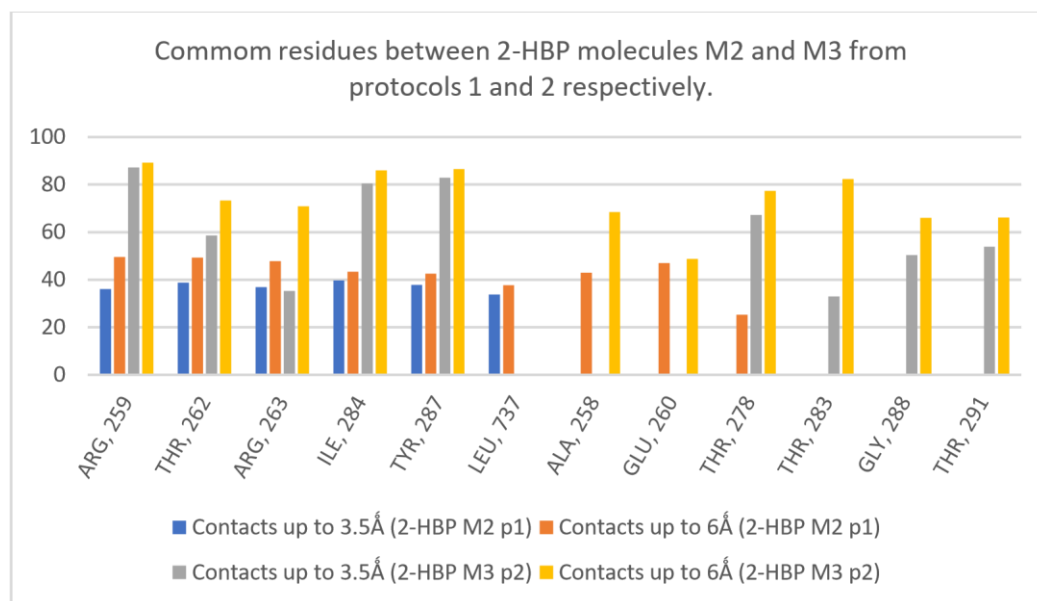


Figure 22 – Common residue interactions between concentrated system 2-HBP molecules M2 (blue and orange) and M3 (gray and yellow) from protocols 1 and 2 respectively. In blue and gray, the interactions up to 3.5 Å and 6.0 Å are represented in orange and yellow.

It is possible to see from Figure 22 that the cavity constituted by the residues Arg259, Thr262 Arg 263, Ile284, Tyr287, and Leu737 (primary structure Arg274, Thr277, Arg 278, Ile299, Tyr302, of chain A and Leu351 of chain F) binds 2-HBP molecules in either protocol 1 or protocol 2. However, the inhibitor molecule that bound the cavity is not the same, as can be observed by measuring the distance of each inhibitor molecule to the geometric center of the cavity (Figure 23). This is an interesting observation, as it speaks to the affinity of the inhibitor towards the cavity. Inhibitors in solution, after disrupting the interaction with the former binding site, manage to find and bind strongly a different binding site.

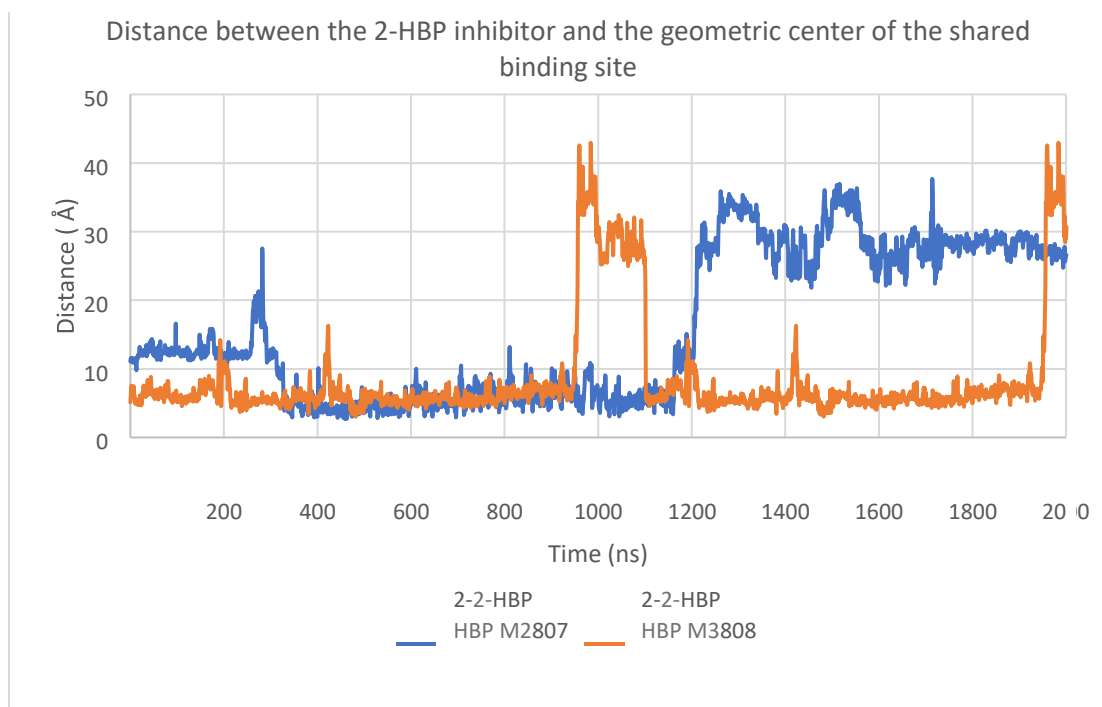


Figure 23 – Graphical representation of the distance between 2-HBP molecules M2 and M3 from concentrated system protocol 1 and 2 respectively and the geometric center of the shared binding site. In orange represented the 2-HBP molecule from the concentrated system submitted to protocol 2 and in blue the 2-HBP M2 of the same system submitted to protocol 1.

From Figure 23, it is possible to observe that the 2-HBP molecule M3 from protocol 2 was initially present in the cavity, and 2-HBP M2 from protocol 1 was initially bound in a nearby cavity, and then migrated to the same binding site as 2-HBP M3. It can also be seen that around 120 ns the 2-HBP molecule M2 migrated again and did not return to the binding site, which explains the lower contact percentage observed in Figure 24. Another cavity which bound 2-HBP inhibitors in both protocols is composed of residues Val524, Trp527, Ile573, Val624, Phe627, Ile628, Pro799, Gly800, and Phe801 (primary structure Val138, Trp141, Ile187, Phe241, Ile242, Pro413, Gly414, and Phe415 of chain F) (Figure 26). In this case, the 2-HBP inhibitors are the same, and both started in the binding site originally. This can be a good indicator that the binding site may interact with the 2-HBP inhibitor strongly. Finally, the last common binding site found for the system containing 2-HBP is made by residues Ala300, Lys329, Gly332, Val333, Leu336, and Asn339 (primary structure Ala315, Lys344, Gly347, Val348, Leu351, and Asn354 of chain A) (Figure 27). Similarly, the 2-HBP inhibitors also bound this binding site originally.

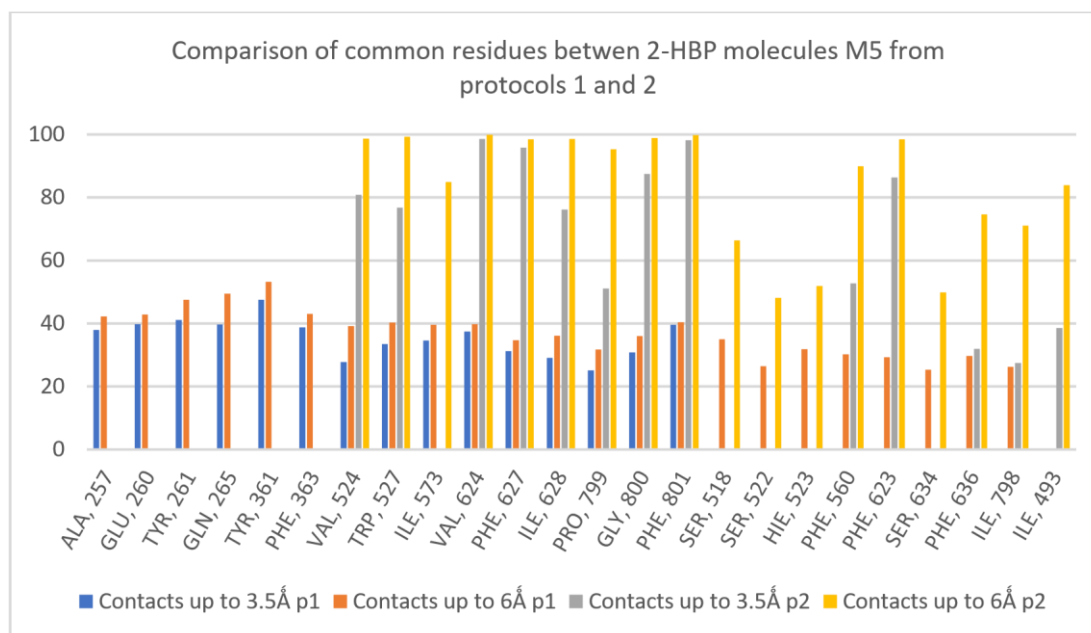


Figure 24 - Common residue interactions between concentrated system 2-HBP molecules M4 from protocol 1 (blue and orange) and 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6.0 Å are represented in orange and yellow.

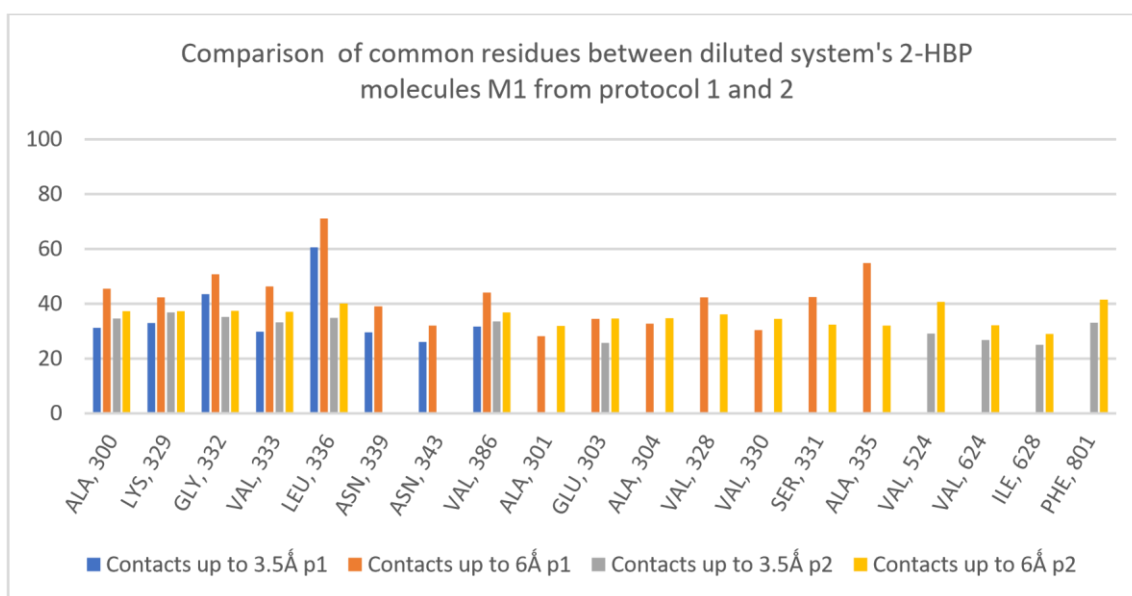


Figure 25 - Common residue interactions between concentrated system 2-HBP molecules M4 from protocol 1 (blue and orange) and 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å and 6.0 Å are represented in orange and yellow.

Regarding the systems containing HBPS inhibitor molecules, it was found that for the concentrated systems, the residues that interacted with the inhibitors from protocol 1 were different from the ones that interacted with their counterpart from protocol 2. In fact, in the systems containing HBPS inhibitors only one cavity was found to bind the inhibitor in both protocols (Figure 28).

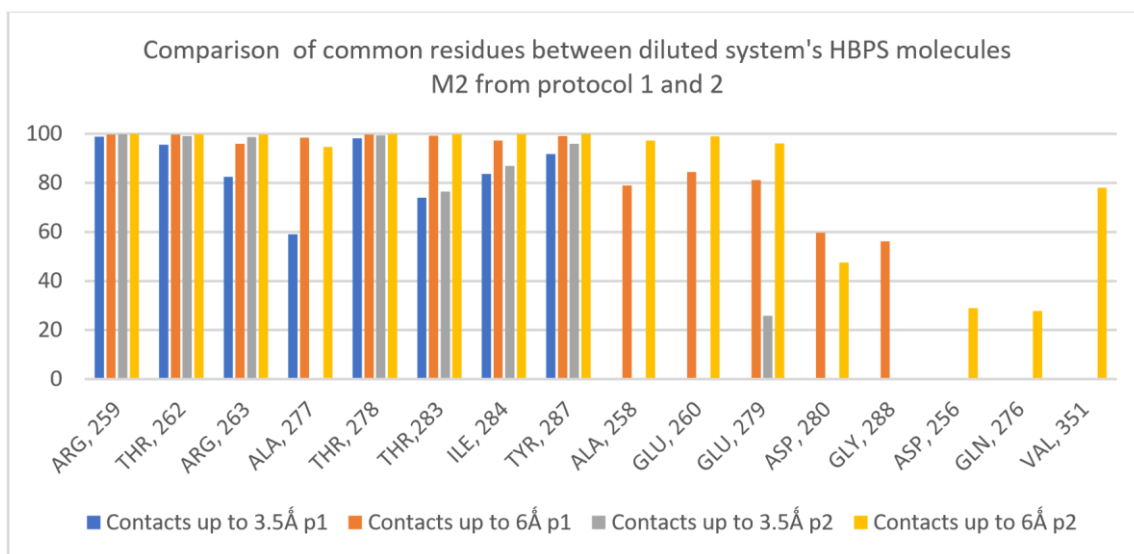


Figure 26 - Common residue interactions between diluted system's HBPS molecules M2 from protocol 1 (blue and orange) and 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å; in orange and yellow the interaction up to 6.0 Å.

One particularity regarding this cavity is that in protocol 1 the HBPS M2 molecule migrated from a region comprising chain A residues Arg259, Thr262, Arg263, Ala277, Thr283 (primary structure Arg274, Thr277, Arg288, Ala292, and Thr298) to those equivalents in chain F (Figure 20 (C)). Note that to make the graphical representation the residues Arg 660, Thr 663, Arg 664 of chain F (primary structure Arg274, Thr277, Arg278) were represented as Arg 259, Thr 262, Arg 263, the equivalent residues of chain A. The fact that the inhibitor molecule was able to dissociate from the region to bind the equivalent binding region in the opposite chain is not only an indication that the interactions it makes with the binding site are strong, but also enforces that the extra step taken in protocol 2 may be essential for the discovery of appropriate binding cavities. It is likely that the inhibitor was driven from the original binding site due to the repulsive forces mentioned previously; however, when it found the equivalent binding site in the opposite monomer, and the enzyme was able to interact with the inhibitor much more strongly (contact percentage close to 100% on the equivalent binding site).

Common binding sites between concentrated and diluted systems

When the systems were first designed, the diluted systems were created with the poses that were common between HBPS and 2-HBP inhibitors. The reason for that was that we believed that those binding sites (*i.e.*, with a higher number of poses docked in the same binding site) deserved closer attention (Figure 27).

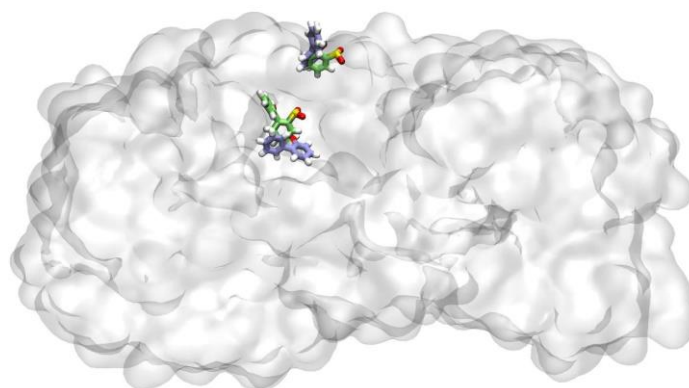


Figure 27 – Relative positions of the inhibitors 2-HBP and HBPS regarding each other and the enzyme. In surface, the DsxC is represented, and in sticks, the 2-HBP (cyan) and HBPS (green).

Since several of the initial binding sites are different among protocols 1 and 2, it becomes relevant to see if any of the poses present in the concentrated and diluted system migrated to the same or equivalent binding sites. Hence, similarly to the analysis that has been done, a search for interactions with similar/equivalent residues was made between concentrated and diluted systems.

Regarding the systems containing HBPS inhibitors, no relevant information was found upon comparing data from concentrated and diluted systems. Only one binding site (Arg259, Thr262, Arg263, Ala277, Thr283) (primary structure Arg274, Thr277, Arg278, Ala292, and Thr298) was found to be present in both the concentrated and diluted system; however, it presented contact percentages below 20% for one system and therefore was not considered relevant. When observing the results from systems including 2-HBP, some binding sites appear in both the diluted and concentrated systems. As seen for the systems containing HBPS, one binding site (Arg259, Thr262, Arg263, Ala277, Thr283) (primary structure Arg274, Thr277, Arg278, Ala292, and Thr298) binds 2-HBP in both the diluted and concentrated systems; however its interactions are not systematically prevalent (low contact percentage). Additionally, only one other binding site was found between the concentrated and diluted system (Figure 28).

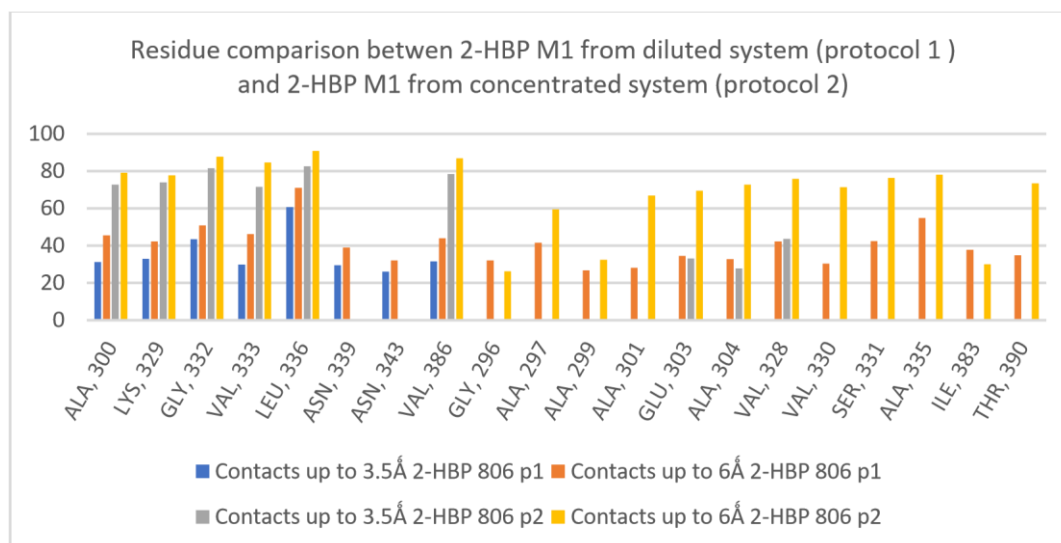


Figure 28 - Common residue interactions between diluted system's 2-HBP molecule M1 from protocol 1 (blue and orange) and concentrated system's 2-HBP molecule M1 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å; in orange and yellow the interaction up to 6.0 Å.

Analysis of the results thus far has already provided relevant information regarding the binding of the 2-HBP and HBPS inhibitors. From the evaluation of the graphical representations, it is possible to see that several binding sites have a greater contact percentage with 2-HBP when compared to HBPS. However, it is known from literature that HBPS has higher inhibitory power than 2-HBP (depicted in Figure 6),⁶⁵ as DszC inhibition by HBPS has a lower IC₅₀ (15 µM) than by 2-HBP (50 µM). Hence, this information can be useful in selecting the most probable binding site.

Common binding sites between 2-HBP and HBPS inhibitors

As mentioned before, it is known that both 2-HBP and HBPS molecules inhibit DszC, and therefore it is possible that both inhibitors bind the same binding site. Hence, an analysis was made to ascertain whether there were any binding sites that bound both inhibitors simultaneously and whether it interacted with them strongly.

One interesting binding site found was constituted by residues Arg259, Thr262, Arg263, Ala277, Thr283 (primary structure Arg274, Thr277, Arg278, Ala292, and Thr298). This binding site was already discussed above because of its strong and persistent contacts with either 2-HBP and HBPS. Additionally, we found that its equivalent binding site on the opposite monomer also bonded both inhibitors. Moreover, analysis of Figure 31 and Figure 32 also show that both cavities bind simultaneously 2-HBP and HBPS. Hence, we believe that this binding site may be one of the most promising, in terms of being responsible for DszC inhibition by 2-HBP and HBPS.

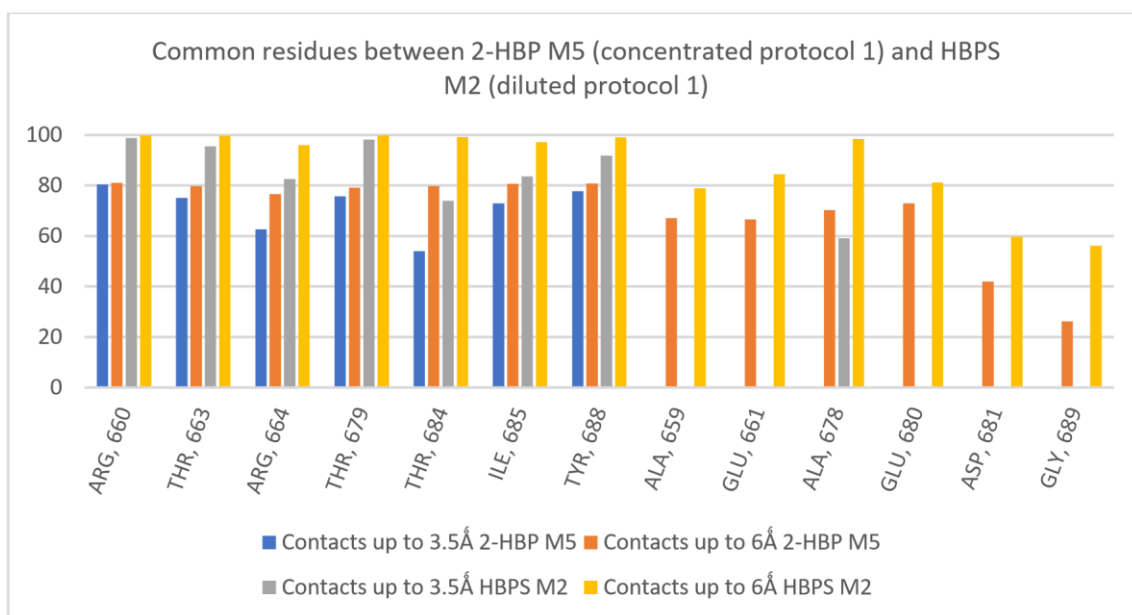


Figure 29 - Common residue interactions between concentrated system's 2-HBP molecule M5 from protocol 1 (blue and orange) and diluted system's HBPS molecule M2 from protocol 1 (gray and yellow). In blue and gray, the interactions up to 3.5 Å; in orange and yellow the interaction up to 6.0 Å.

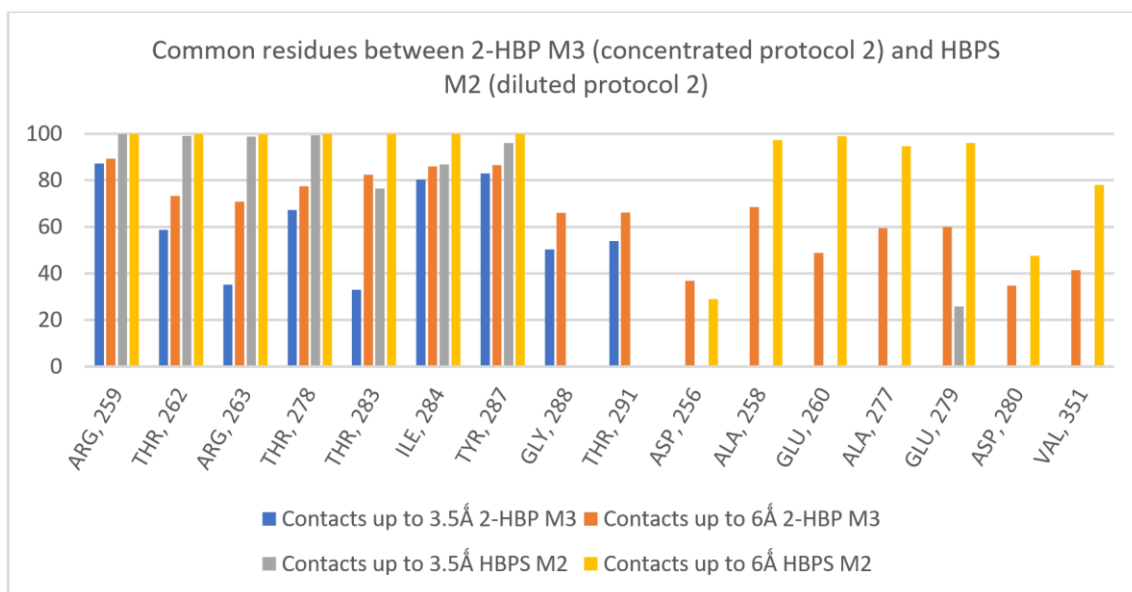


Figure 30 - Common residue interactions between concentrated system's 2-HBP molecule M3 from protocol 2 (blue and orange) and diluted system's HBPS molecule M2 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å; in orange and yellow the interaction up to 6.0 Å.

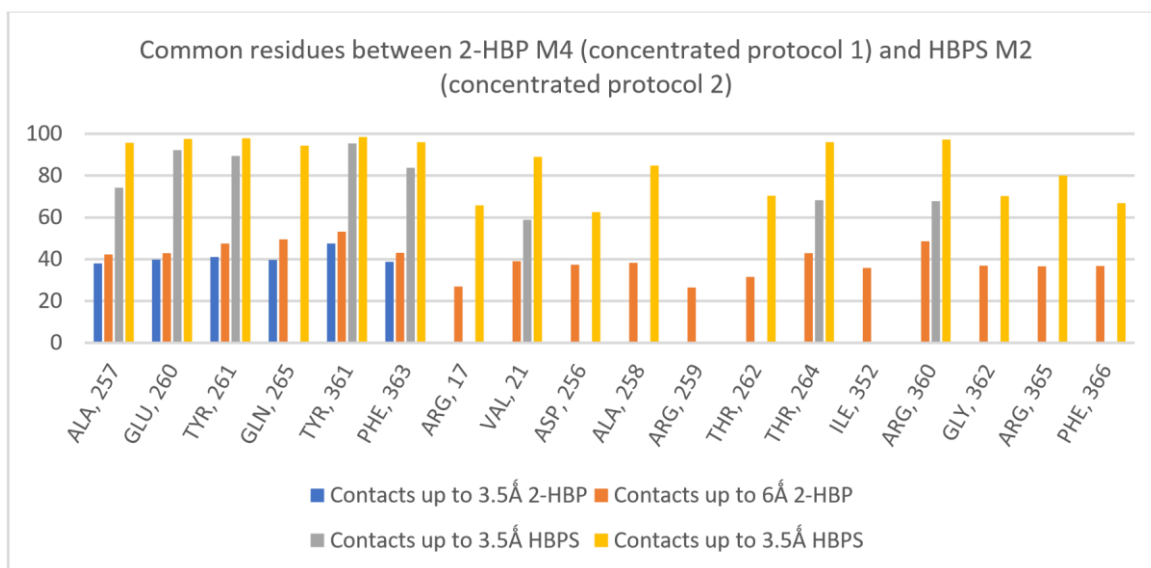


Figure 31 - Common residue interactions between concentrated system's 2-HBP molecule M4 from protocol 1 (blue and orange) and concentrated system's HBPS molecule M2 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å; in orange and yellow the interaction up to 6.0 Å.

Analysis of Figure 31 shows that the cavity is composed of residues Ala257, Glu260, Tyr261, Gln265, Tyr361, and Phe363 (primary structure Ala272, Glu275, Tyr276, Gln280, Tyr376, and Phe378 of chain A) binds both 2-HBP and HBPS. However, its binding affinity towards HBPS inhibitors is much greater than its affinity towards 2-HBP, which is not surprising given its polar nature.

Lastly, the final common binding site found is composed of residues Ile493, Val524, Val624, Phe627, Ile628, Pro799, Gly800, and Phe 801 (primary structure Ile107, Val138, Val238, Phe241, Ile242, Pro413, Gly414, and Phe415 of chain F) (Figure 32). This binding site also shows great promise due to the very high contact percentages between both inhibitors and the residues that make up the binding site.

It is important to mention that additional binding sites, common to 2-HBP and HBPS were found; however, due to low contacts or few common interacting residues, they are not shown here.

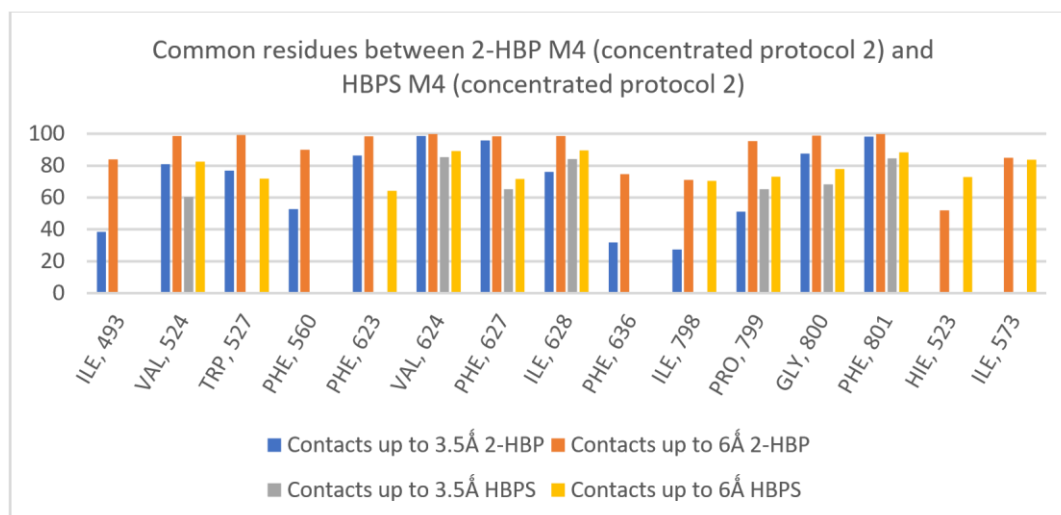


Figure 32 - Common residue interactions between concentrated system's 2-HBP molecule M4 from protocol 2 (blue and orange) and concentrated system's HBPS molecule M4 from protocol 2 (gray and yellow). In blue and gray, the interactions up to 3.5 Å; in orange and yellow the interaction up to 6.0 Å.

Most promising binding sites

Docking procedures are not the most adequate to mimic the interaction of ligands and proteins. This is because most docking software is designed to be a fast method to screen large libraries of compounds, generally used in drug development. Additionally, the most accurate methods, such as flexible ligand and receptor go hand-in-hand with high computational power. Applying molecular mechanics simulations to docked complexes is a good way to validate docking results, as those can be viewed as fast and accurate descriptors of molecular and enzymatic motions. Simulation of the different systems allowed the exploration of the docked poses and the application of the four descriptors – (1) whether or not the contacts existed in both protocol 1 and 2; (2) the abundance/contact percentage higher than 80%; (3) whether the contacts existed in both the concentrated and diluted system; and (4) whether the binding site was shared by both the 2-HBP and HBPS molecules – resulted in the selection of the two most promising poses that may be responsible for inhibiting DszC upon binding of 2-HBP and HBPS.

The results of the simulations show that residues Arg259, Thr262, Arg263, Thr278, Thr283, Ile284, Tyr287, Gly288, and Thr291 (primary structure Arg274, Thr277, Arg278, Thr293, Thr298, Ile299, Tyr302, Gly303, and Thr306 of chain A), appear very often in both the diluted and concentrated systems, as well as in both protocols. Additionally, the systems in which the inhibitors do not bind those residues, often bind the equivalent

binding site in chain F. In some cases, inhibitor molecules that dissociate from their original binding site also ended up binding one of those two binding sites. Furthermore, this binding site is consistent across all four parameters defined for assessment of possible binding sites, that is, not only it can be found in both the concentrated and diluted systems in protocol 1 and 2, but it also interacts with both inhibitors with an abundance of at least 80%. For that reason, we believe that this cavity is highly promising and may be involved in the inhibition of DszC by 2-HBP and HBPS. Another cavity that shows promise is constituted by Ile493, Val524, Trp527, Phe623, Val624, Phe627, Ile628, Pro799, Gly800, and Phe801 (primary structure Ile107, Val138, Trp141, Phe237, Val238, Phe241, Ile242, Pro413, Gly414, and Phe415 of chain F). This second binding site does not appear as often; however, it does exist in both concentrated and diluted systems and in protocols 1 and 2, even though its contacts with the inhibitor are not as abundant in all systems. Nonetheless, from Figure 32 we can observe that it does also interact with both inhibitors strongly.

In summary, the binding sites that appear to have a more relevant role in DszC inhibition, are composed of:

Binding Site 1. Arg 259, Thr 262, Arg 263, Thr 278, Thr 283, Ile 284, Tyr 287, Gly 288, and Thr 291 (primary structure Arg274, Thr277, Arg278, Thr293, Thr298, Ile299, Tyr302, Gly303 and Thr306 of chain A);

Binding Site 2. Ile 493, Val 524, Trp 527, Phe 623, Val 624, Phe 627, Ile 628, Pro 799, Gly 800, and Phe 801 (primary structure Ile107, Val138, Trp141, Phe237, Val238, Phe241, Ile242, Pro413, Gly414, and Phe415 of chain F);

In the section below, the interaction between the inhibitors and the residues composing each binding site will be described in detail.

Characterization of the interactions between residues within binding sites and 2-HBP/HBPS inhibitors

Once the most promising binding sites were discovered, a cluster analysis of the binding sites and the inhibitor poses was made. Such analysis revealed the binding poses of the inhibitors as well as the position of the protein residues, allowing for a better understanding of the interaction between the two. The clustering of the structures was performed by RMSd of enzyme backbone and ligand, with a cutoff distance of 1 Å. That is, if the selection differed 1 Å from itself from frame to frame, its pose was placed in a

different cluster. Analysis of Figure 35 revealed the predominant binding pose for either 2-HBP and HBPS. That pose is represented by the structures found in the first cluster since it possesses most of the poses. Visual analysis of the results for the HBPS inhibitor shows that there is no visual difference between the clustered structures 2, 3, 4, and 5.

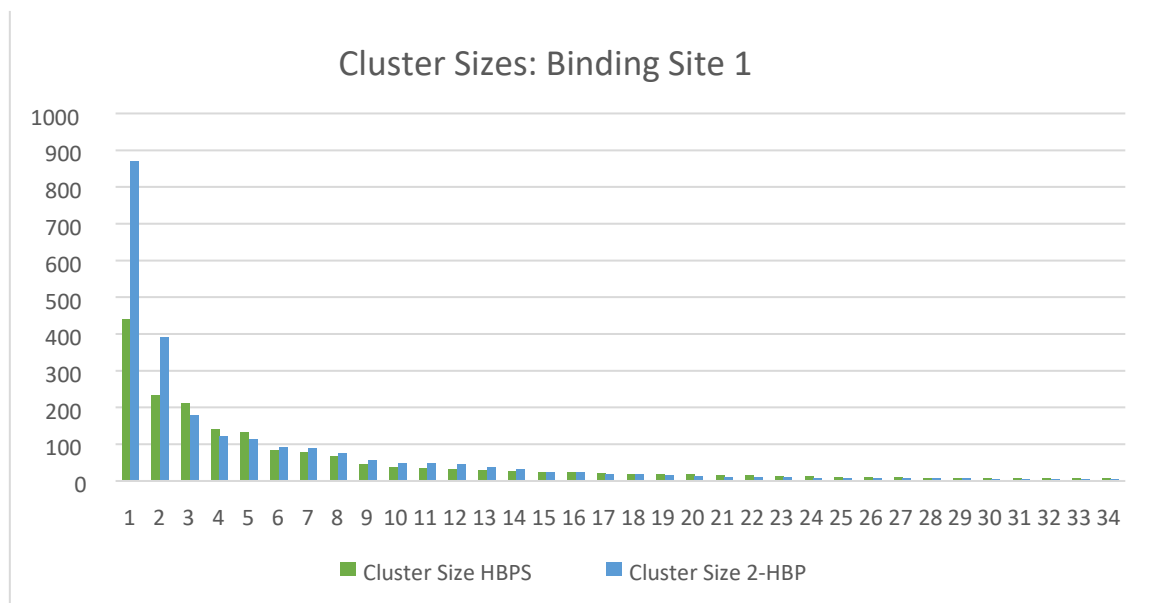


Figure 33 – Cluster sizes clustering analysis on the inhibitors binding on binding site 1. On the xx axis represented the cluster id for cluster of 2-HBP (blue) and HBPS (orange), in the yy axis the number of poses found per cluster.

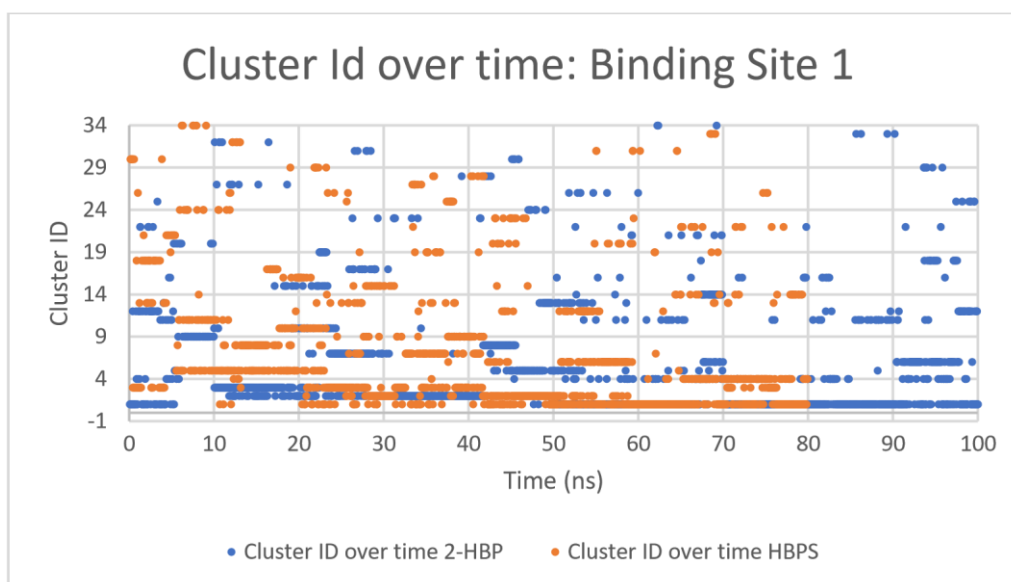


Figure 36 – Graphical representation of the clustered structures over time for binding site 1.

Regarding the interaction of HBPS and binding site 1, it appears that the interaction follows the same pattern as observed for 2-HBP. The main interactions seem to be with hydrophobic residues. The polar charged group of HBPS faces the solvent for most of the simulated time. Nonetheless, there are times where it can interact with the residues by salt bridges (Arg 259 and Arg274) (Figure 38). This interaction only happens for a short period of time and therefore does not appear to be relevant to the affinity of the HBPS inhibitor towards the site. However, the interaction does occur before the packing of the inhibitor inside the binding site and therefore it may play a role in the initial binding of the HBPS inhibitor.

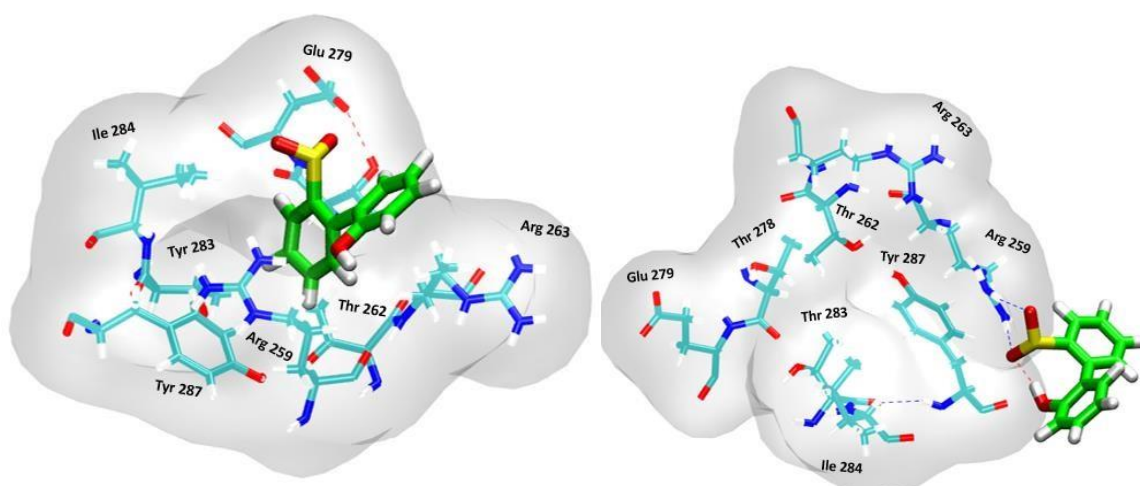


Figure 36 - Representation of binding site 1 (residues Arg 259, Thr 262, Arg 263, Thr 278, Thr 283, Ile 284, Tyr 287, Gly 288, and Thr 291) and the most representatives binding poses for HBPS. In sticks are represented the residues (cyan) and the HBPS inhibitor (green).

Additionally, a particularity of this binding site is that the cavity is only formed upon approximation of either the 2-HBP or HBPS inhibitors. When neither is present, the cavity is shown to be closed, and the enzyme residues interact with each other (Figure 37). Because the binding site is in the interface of the dimers, this interaction may have an effect on the enzyme dimerization, which is crucial for DszC activity because the catalytic site is formed by residues of both monomers (Figure 40).

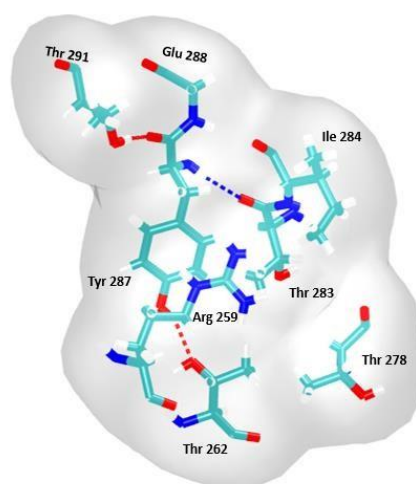


Figure 37 – Representation of binding site 1 in the absence of 2-HBP or HBPS inhibitors. When no inhibitors are in proximity, the binding site is closed.

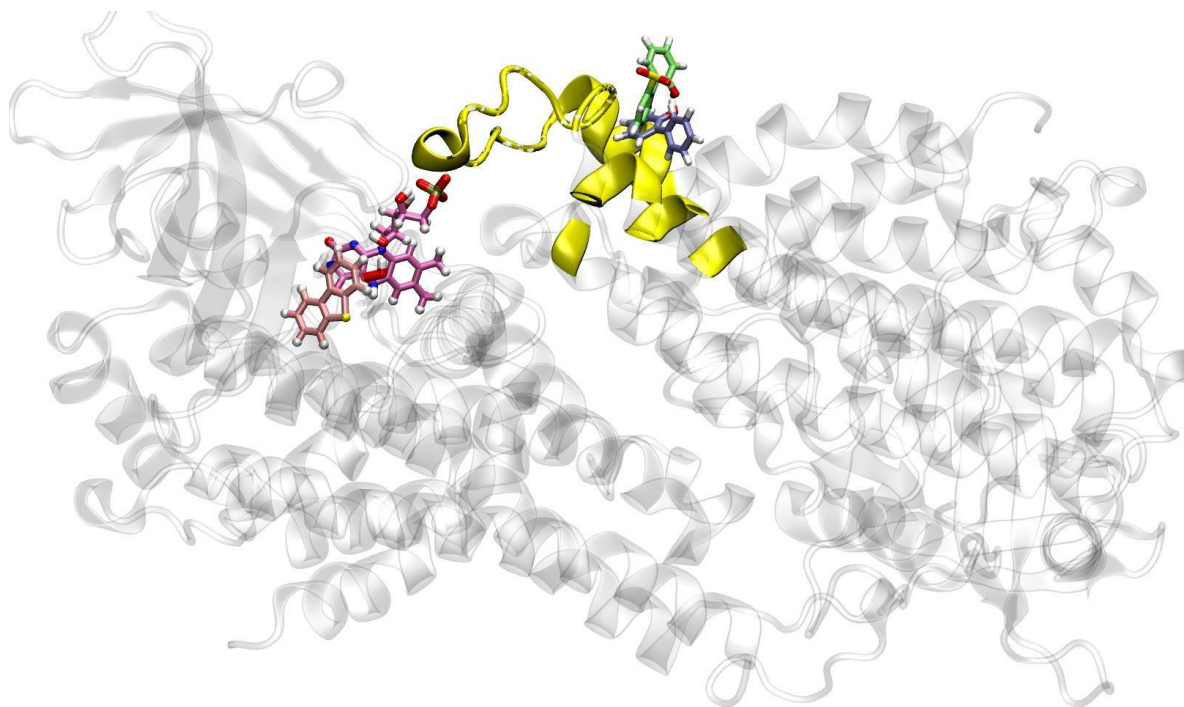


Figure 38 – Representation of the entire DszC complex. In cartoon, a representation of DszC. In sticks, the 2-HBP inhibitor (cyan) and HBPS (green), the C^{4a}OOH intermediate (purple), DBT substrate (pink), and the FMNH₂ cofactor (brown). Highlighted in yellow, the loop portion that interacts with the catalytic sites, bound to the same motive as binding site 1.

A similar cluster analysis was made for binding site 2 (Figure 39).

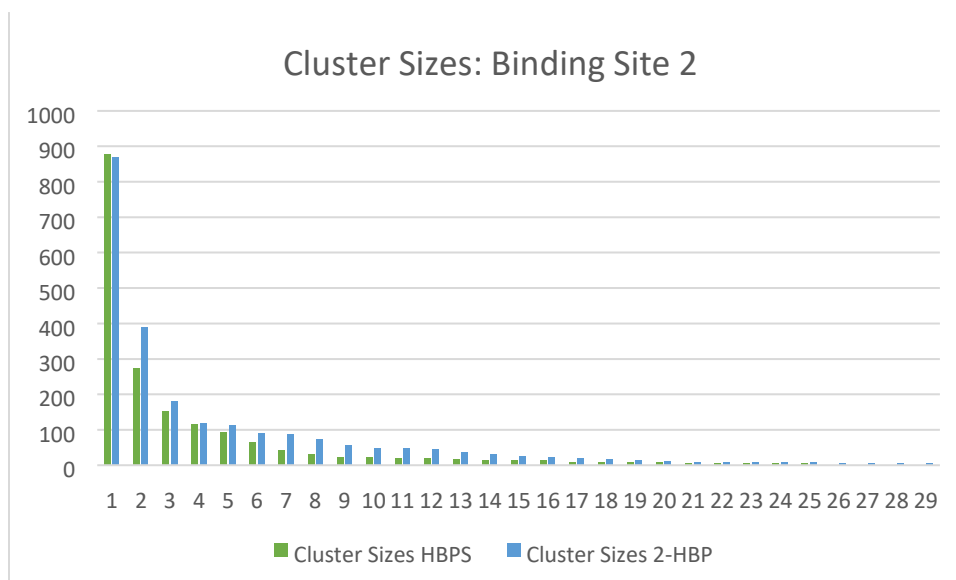


Figure 39 - Cluster sizes clustering analysis on the inhibitors binding on binding site 2. On the xx axis represented the cluster id for cluster of 2-HBP (blue) and HBPS (orange), in the yy axis the number of poses found per cluster.

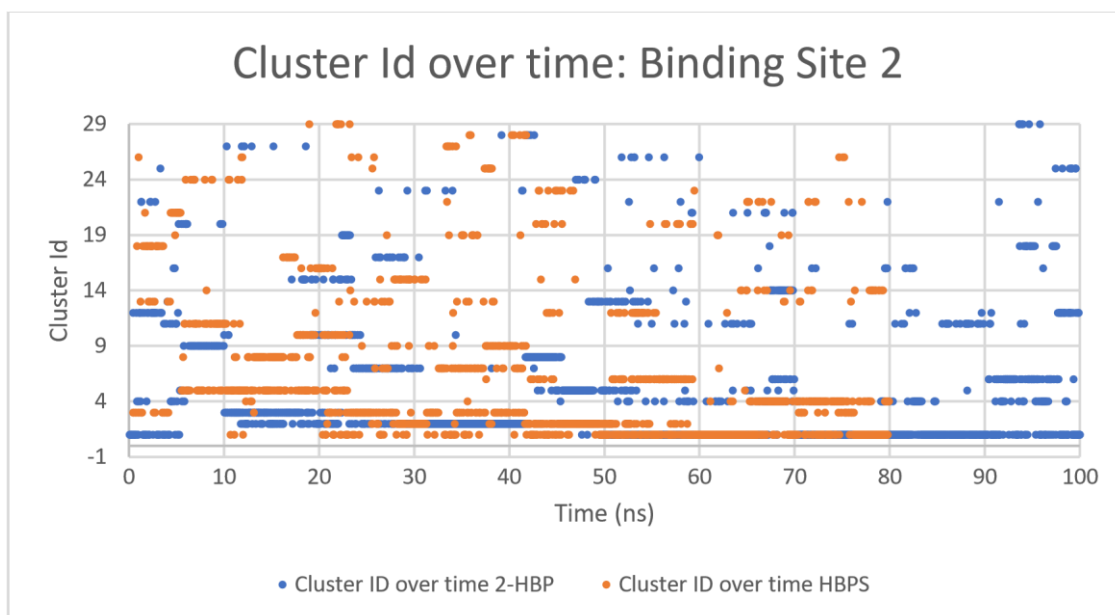


Figure 40 - Graphical representation of the clustered structures over time for binding site 2.

As opposed to the result found for binding site 1, clustering analysis of binding site 2 suggests only one binding pose, either for 2-HBP and HBPS. The first cluster includes most of the structures (~36% 2-HBP and ~45% HBPS). Additionally, those have a large distribution over time, as seen in Figure 42. Once again, the main force behind the binding of 2-HBP and HBPS appears to be hydrophobic interactions with the enzyme residues (Figure 41).

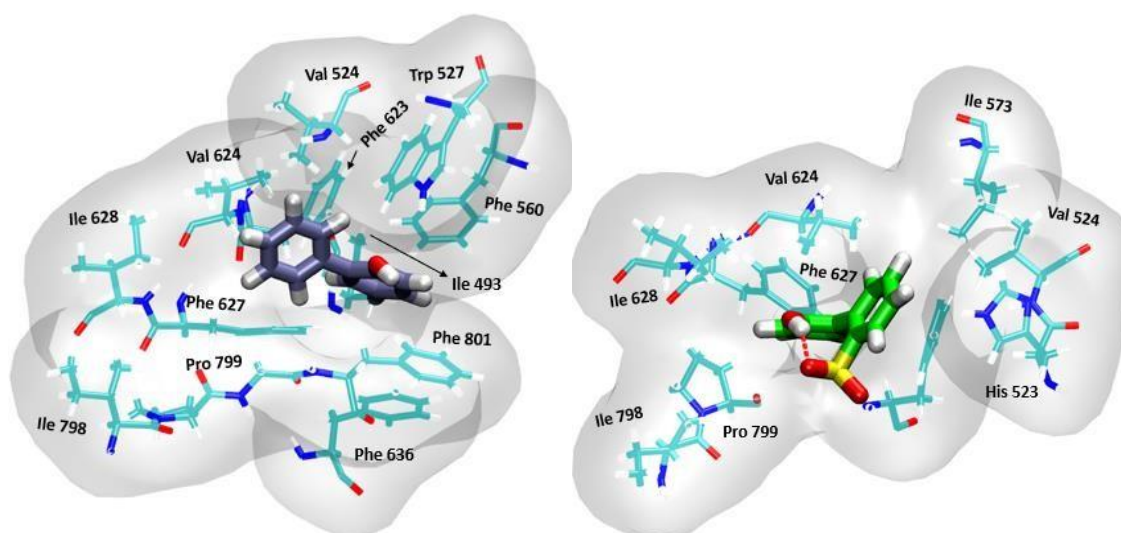


Figure 41 - Representation of binding site 2. In lines are represented the DszC's residues (blue); in sticks the 2- HBP (cyan) and HBPS (green).

As opposed to binding site 1, binding site 2 does not seem to make any representative hydrogen bonds with either inhibitor. Additionally, it is possible to see that 2-HBP is much more buried than HBPS, which explains the contact differences observed in Figure 34. Regarding inhibition, binding site 2 is located near the catalytic site and therefore may interfere with the reaction process (Figure 44).

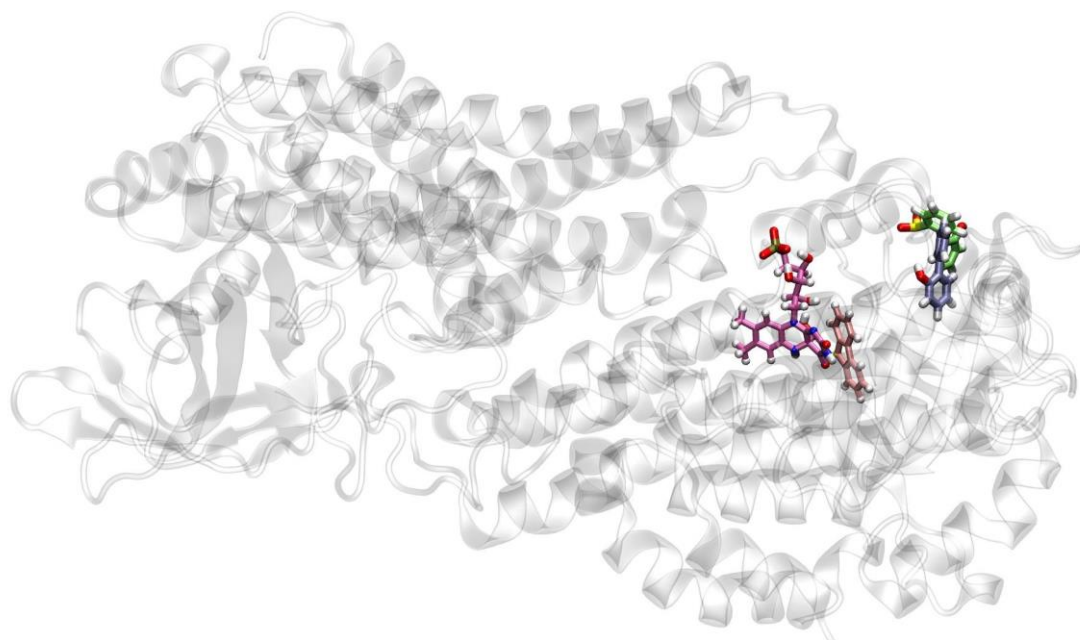


Figure 42 - Representation of the entire DszC complex. In cartoon a representation of DszC chain's A (blue) and F (orange); in blue lines the residues that make up binding site 1; in sticks the 2-HBP inhibitor (cyan) and HBPS (green), the C^{4a}OOH intermediate (purple), DBT substrate (pink) and the FMNH₂ cofactor (brown).

Note that for binding site 1, plenty of inhibitor molecules bound the equivalent cavity of the opposite monomer; however, binding site 2 only bound inhibitors in monomer F, where the DBT and C^{4a}OOH had been placed. This may be an indication that the binding of 2-HBP and HBPS in binding site 2 may be formed upon the entering of the DBT substrate inside the catalytic site. Another hypothesis is that the transformation of reduced FMN to C^{4a}OOH, may change the chemical environment and favor the binding of the inhibitor molecules.

Mechanism of Inhibition

The mechanisms through which inhibition may occur can vary severely from enzyme to enzyme. Knowing that the inhibition is non-competitive in the DszC enzyme, means that the inhibitor does not enter the catalytic site. That is because non-competitive inhibition means that the binding of the inhibitor does not affect the binding of the substrate. In fact,

as we have seen, the catalytic site of DszC does not have enough space to holster both the substrate and the inhibitor, and therefore, when one binds the other cannot.

There are several other mechanisms through which inhibition may occur in a noncompetitive inhibition. Those can be by disrupting the chemical environment of the catalytic site, due to conformational changes that occur upon the binding of the substrate elsewhere. Additionally, because the reaction is dependent on dimerization (which helps stabilizing the cofactor), any changes in the dimerization of DszC may cause inhibition. As demonstrated below, 2-HBP and HBPS show high affinity towards a site located in the dimerization interface of DszC and to a site located in close proximity to the catalytic site. Consequently, any of the previous mechanisms are still in play. Hence, to understand the mechanism of inhibition of DszC by either 2-HBP and HBPS and to try to validate one of the two possible binding cavities discovered, a series of analyses were made to seek structural changes occurring in the protein.

Interactions at the catalytic site

The chemical environment is of utmost importance for the reaction mechanism of any enzyme, and, sometimes, even the smallest modification can prevent the reaction from occurring. The reaction mechanism was already proposed by *Barbosa et. al.*, and helped to decipher the main residues involved in catalysis and key distances for the reactions taking place in the active site, which are depicted in Figure 43. Based on their work, we measured the distances of each residue to the C^{4a}OOH intermediate in the systems that contained the most promising binding sites and compared them to the distances found in the control system (without the presence of either inhibitor). Only the contacts with C^{4a}OOH in the concentrated systems are shown (Figure 44) because no relevant binding sites were found for the diluted system except for the HBPS diluted system from protocol 2, which has HBPS molecule M2 that bind binding site 1, that does not show significant differences from the control system.

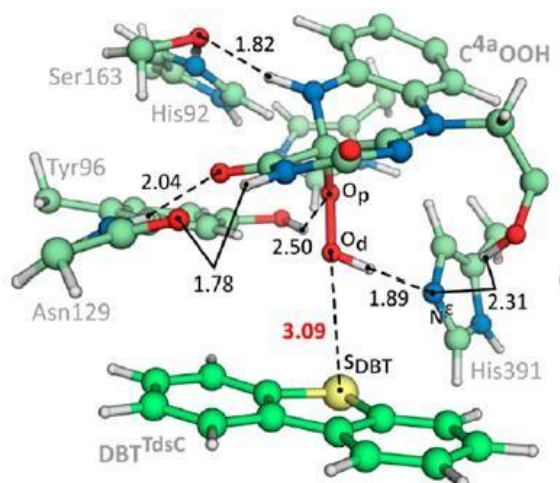


Figure 43 – Representation of the catalytic site of DszC as well as the most important distances for the reaction mechanism. Retrieved from Barbosa et al.

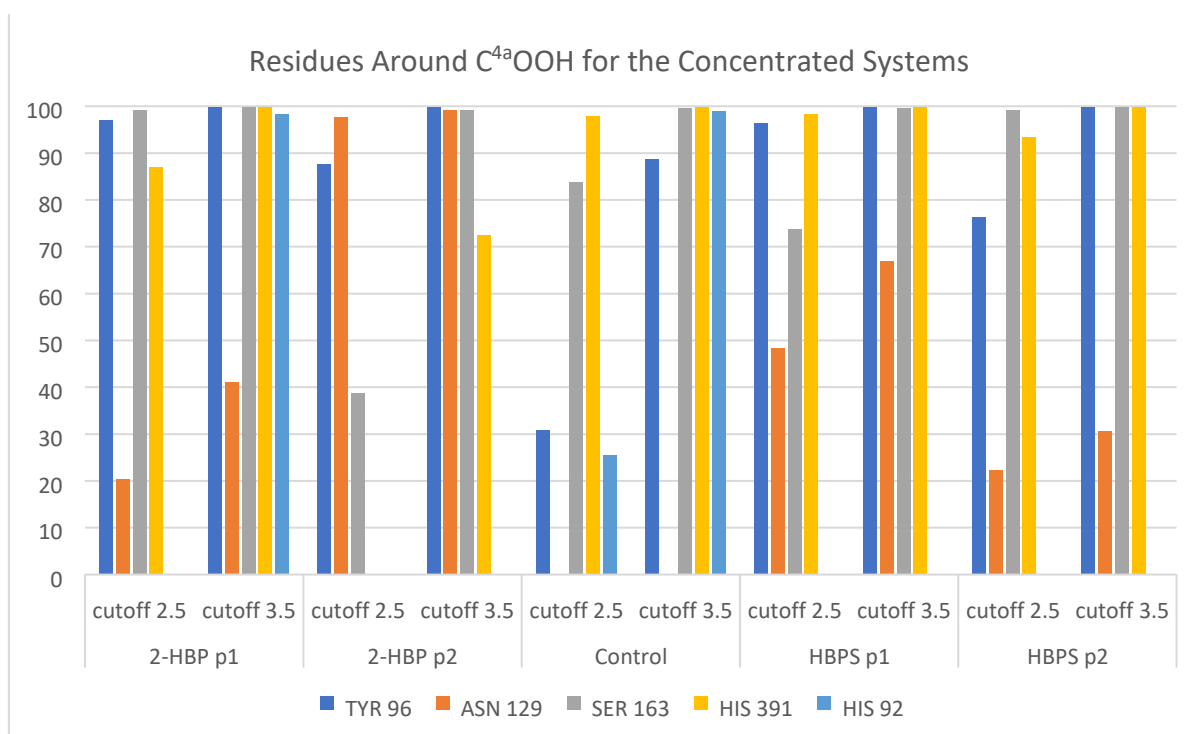


Figure 44 – Residue interaction for C^{4a}OOH depicted for the different concentrated systems.

As shown in Figure 44 there are significant changes amongst the different systems and the control system. The higher differences occur in the concentrated 2-HBP system from protocol 2, where the interactions with His92 and His391 were completely lost. The system in question holds 2-HBP molecules M3 and M4 that bind binding sites 1 and 2, respectively. For the 2-HBP system from protocol 1, the differences are not as evident; however. Nonetheless, the C^{4a}OOH lost interaction with His92. For HBPS systems, the

differences between protocol 1 and 2 are subtle; however, the system from protocol 2 shows, overall, lower contacts

To further validate the results shown, a cluster analysis of the most important residues for the reaction mechanism was made for the systems represented above. Additionally, since the 2-HBP concentrated system showed the greatest differences only the clustered structures resulting from that system are represented (Figure 45). As depicted, it is possible to see that His391 has completely distanced from the C^{4a}OOH intermediate. Additionally, His92 is also not present around the intermediate in the system simulated with 2-HBP.

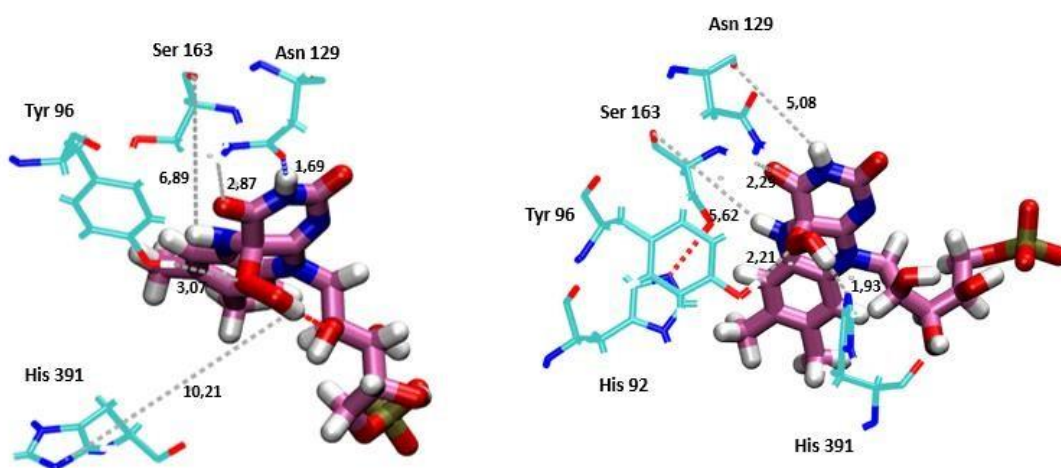


Figure 45 – Representation of the catalytic site of DszC. To the left, the catalytic site after 200 ns simulation of 2-HBP concentrated system from protocol 2; to the right, the catalytic site after 200 ns simulation of the control system. In lines, the catalytic residues, and in sticks, the C^{4a}OOH intermediate. Dotted lines in red and blue represent hydrogen bonds, and dotted lines in gray represent the distance between the most important residues and the C^{4a}OOH.

Showing that the binding of 2-HBP might have caused modifications to residue interactions within the catalytic site is not enough to help select one binding site, because either binding sites 1 or 2 may lead to modifications at the catalytic site. Binding site 1 is located near a loop region that interacts with C^{4a}OOH's tail and therefore leads to destabilization of the intermediate, thus changing its interactions within the binding site. Binding site 2 is located near the catalytic site and may cause modifications directly. Additionally, the remaining systems do not show similar changes like the ones observed for the 2-HBP concentrated system from protocol 2. Hence further investigation will help to determine whether the observable differences are in fact due to 2-HBP binding.

Moreover, because enzyme motions such as large conformational changes occur in the millisecond time scale, it is possible that the simulated time might have not been enough to see any changes upon binding of 2-HBP or HBPS.

Structural changes in DszC upon binding of 2-HBP/HBPS

Even though complete motion of large enzyme motives such as α -helixes or β -sheets occurs in the millisecond time scale, smaller events that lead to those conformational changes may occur first, in smaller time scales. Such events can be large residue fluctuation or changes in binding energies of residues within the protein. So, we analyzed the residue fluctuations (RMSf) of the systems that contained poses that bound the most promising binding sites and the binding energies of the residues at the DszC dimerization interface.

Analysis of residue fluctuation by itself does not bring much information, because residues, especially surface residues, usually have large fluctuations values. To have a better grasp of the results, the residue fluctuation of the control system was subtracted from the residue fluctuation of the systems containing inhibitors. That way, large fluctuation values in those systems could be due to the binding of 2-HBP or HBPS.

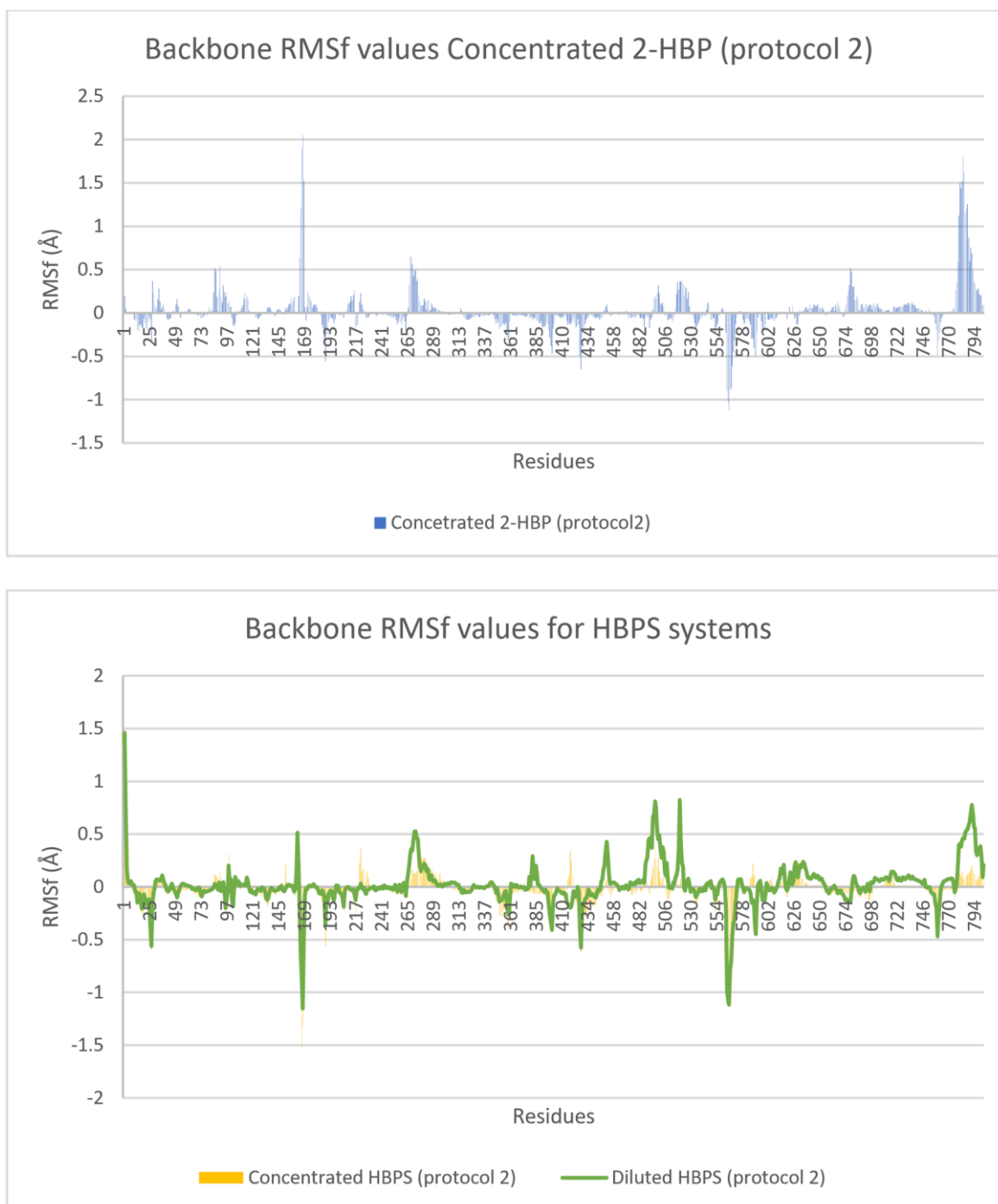


Figure 46 – Normalized backbone RMSf values for concentrated 2-HBP system from protocol 2 (top) and concentrated and diluted HBPS systems from protocol 2 (bottom).

From the analysis of Figure 46, it is possible to see that most varying regions encompass residues 166-169, 567-570, and 780-789 (primary structure 181-184 of chain A and 181-184 and 394-403 of chain F). The first two regions (residues 166-169 and 567-570) correspond to equivalent loop regions (same loop on chains A and F). However, the region composed of residues 780-789 (394-403 of chain F) is an α -helix, which should

not be as flexible. Additionally, it is in proximity to both binding sites 1 and 2. Figure 46 (bottom) shows a similar tendency for loop regions 166-169 and 567-570 but presents different behavior for the α -helix 780-789. In the system where HBPS inhibitor binds binding site 1 (concentrated HBPS protocol 2), there is no relevant movement of the α -helix region; however, when HBPS binds binding site 2 (diluted HBPS protocol 1), it is possible to see an increase in the RMSf of that α -helix. Hence, it is possible that the binding of the inhibitors to binding site 2 may cause conformational changes that lead to the displacement of that enzyme motif. Additionally, the fact that the displacement is not as substantial in the system containing HBPS may explain why the interactions within the catalytic site differ less from the control system when compared to the system containing 2-HBP, where the α -helix displacement is more discernible.

Regarding the dimerization of DszC upon binding of 2-HBP inhibitors, an MM-PBSA analysis was made for the interface residues (within 5 Å of the opposable monomer). Each residue was mutated by an Ala residue and changes in binding energy were calculated.

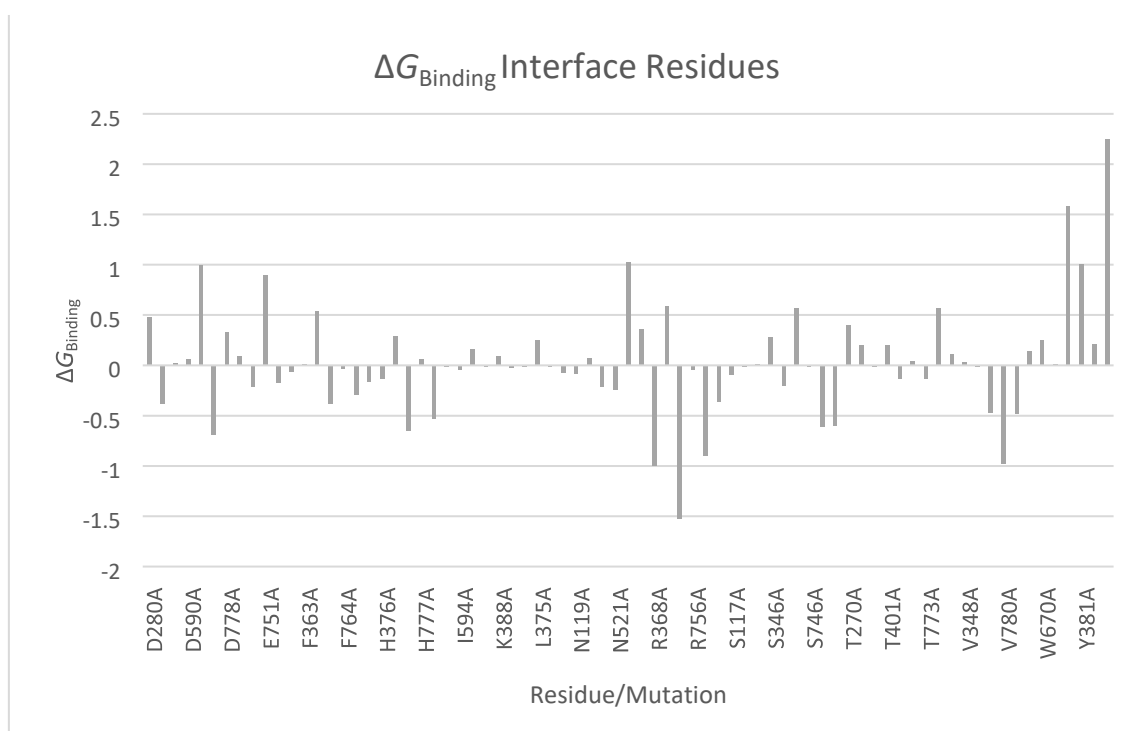


Figure 47 – Binding energy of the interface residues for the 2-HBP concentrated system from protocol 2 upon subtraction of the binding energy of the control system.

As shown in Figure 47, there are two major “warm-spots” ($\Delta G_{\text{binding}} > 1.5$ kcal) at residues Tyr 297 (chain A) and Tyr 396 (chain F). Therefore, those should have a higher contribution upon dimerization. Nonetheless, there are no major differences at the

interface residues upon binding of the 2-HBP inhibitor. Hence, one can assume that the binding of DszC may not be affected upon the binding of the inhibitor.

Conclusion

Fossil fuels play a crucial role in nowadays society, but their use is associated with the emission of greenhouse gases like CO₂ as well as toxic gases like SO₂ and NO_x. To alleviate the effects that fossil fuel consumption has on the environment and Human health, several laws and directives have been emitted that limit the amount of sulfur in crude oil derivatives. The main sulfur-containing molecule in crude oil, DBT, is inert to several desulfurization techniques, and therefore, removing the sulfur from its composition becomes extremely difficult and expensive. Alternatively, biodesulfurization is a greener and cheaper compared to chemical desulfurization techniques. This method uses the 4S pathway where four enzymes catalyze the sequential oxidation of DBT to 2-HBP. However, it is currently slower than its chemical counterparts. Additionally, most enzymes in the 4S pathway suffer from non-competitive inhibition by the pathway end-product, 2-HBP. One, in particular, DszC, is especially affected by inhibition by not only 2-HBP but also by HBPS. Throughout this work, we have studied DszC inhibition by both molecules, *via* docking and molecular dynamics and proposed the most probable binding sites, responsible for DszC inhibition. We started by studying the contacts between the inhibitors and the protein residues through a 200 ns simulation. To those contacts we applied four parameters: (1) whether or not the contacts existed in both protocol 1 and 2; (2) the abundance/contact percentage higher than 80%; (3) whether the contacts existed in both the concentrated and diluted system; and (4) whether the binding site was shared by both the 2-HBP and HBPS molecules); which allowed the identification of 2 most promising binding sites – (1) Arg259, Thr262, Arg263, Thr278, Thr283, Ile284, Tyr287, Gly288 and Thr291 (primary structure Arg274, Thr277, Arg278, Thr293, Thr298, Ile299, Tyr302, Gly303 and Thr306 of chain A); (2) Ile493, Val524, Trp527, Phe623, Val624, Phe627, Ile628, Pro799, Gly800, and Phe801 (primary structure Ile107, Val138, Trp141, Phe237, Val238, Phe241, Ile242, Pro413, Gly414, and Phe415 of chain F) –, that may play a crucial role in DszC inhibition. Upon further analysis of the enzyme catalytic site, we were able to observe some structural differences that may impair its activity, namely the displacement of the catalytic residues (His 92, His 391, Tyr 96). The binding of the inhibitor may cause such structural changes directly or by disrupting the dimerization of DszC which is crucial for its activity. To ascertain whether the dimerization of the enzyme was affected upon binding of the 2HBP, an MM/PBSA analysis was performed. The results suggest that the dimerization was not affected by the binding of 2-HBP, since there were no relevant differences in the $\Delta G_{\text{binding}}$ of the residues at the DszC dimerization interface. RMSf analysis of residue's backbone shows that the systems where 2-HBP or

HBPS bound binding site 2 present a displacement of an α -helix (residues 780 to 790) (394 to 404 of chain F), which may be involved in the structural changes observed in the catalytic site. The results obtained thus far do not allow the selection of one binding site responsible for the inhibition of DszC upon binding of 2-HBP or HBPS. To do that further analysis need to be performed. As an example, systems with 2-HBP or HBPS bound at the discovered binding sites could be simulated with accelerated molecular dynamics in order to determine if large enzyme motions, such as conformational changes like the ones that probably occur upon binding of an inhibitor in an allosteric site, occur upon binding of 2-HBP or HBPS. Such conformational changes take place in the millisecond time scale. Hence, more simulation time should provide further information to confirm the data found so far. Moreover, exploring the binding affinities of the inhibitors towards the binding sites through MMPB(GB)SA should also provide relevant information regarding the binding of the inhibitors as well as the most important residues in each binding site. Such information will play a valuable role in impairing DszC inhibition by 2-HBP and HBPS.

Future Perspectives

The work developed so far has shed some light into two possible binding sites responsible for DszC inhibition by 2-HBP and HBPS and explored potential mechanisms through which inhibition occurs. However, to eventually impair the feedback inhibition, it is still necessary to investigate the residues that play a more relevant role in the interaction with 2-HBP and HBPS. To uncover the binding affinities of the inhibitors and the binding site residues, MM/PB(GB)SA analysis could be performed for each cavity found. With that information, rational enzyme mutations could be made to impair the binding of the inhibitor, followed by an MD simulation of the mutated complex to provide information regarding the quality of the mutation. Nonetheless, computational methods provide only an approximation to *in vivo* processes and, therefore, results found through computational methods should be further verified by experimental assays. Additionally, certain mutations could have effects on the folding of the enzyme. Such effects would be unlikely seen in MD simulations and therefore, experimental analysis is always a requirement.

Bibliography

1. Ali, U. The history of the oil and gas industry from 347 AD to today. <https://www.offshore-technology.com/comment/history-oil-gas/>.
2. Abhishek Shah. Advantages and Disadvantages of Oil – Cons Disregarded by Powerful Lobbies. <https://www.greenworldinvestor.com/2011/07/07/advantages-anddisadvantages-of-oil-cons-disregarded-by-powerful-lobbies/>.
3. Editorial Debt. No Title. *What is Crude Oil? A Detailed Explanation on this Essential Fossil Fuel* <https://oilprice.com/Energy/Crude-Oil/What-Is-Crude-Oil-A-Detailed-ExplanationOn-This-Essential-Fossil-Fuel.html>.
4. European Parliament. Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018. *Off. J. Eur. Union* **328**, 1–77 (2018).
5. Zhong, X. Z., Wang, G. C., Wang, Y., Zhang, X. Q. & Ye, W. C. Monomeric indole alkaloids from the aerial parts of *Catharanthus roseus*. *Yaoxue Xuebao* **45**, 471–474 (2010).
6. IRENA. *Renewable Energy Statistics 2019*. International Renewable Energy Agency vol. 1 (2019).
7. Data & Statistics. No Title. *Total Energy Supply by source* <https://www.iea.org/data-andstatistics?country=WORLD&fuel=Energy supply&indicator=TPESbySource> (2019).
8. Our World in Data. CO₂ Emissions by Fuel Type. <https://ourworldindata.org/grapher/co2-by-source?time=earliest..latest> (2019).
9. Statista. Daily demand for crude oil worldwide from 2006 to 2021. <https://www.statista.com/statistics/271823/daily-global-crude-oil-demand-since-2006/> (2020).
10. The International Council On Clean Transportation. An Introduction To Petroleum Refining and the Production of Ultra Low Sulfur Gasoline. *Energy Econ. Appl. Optim.* 1–38 (2011).
11. Wright, A. A. & of Marine Engineers, I. *Exhaust Emissions from Combustion Machinery*. (Institute of Marine Engineers, 2000).
12. Duissenov, D. Production and processing of sour crude and natural gas - challenges due to increasing stringent regulations. 101 (2013).
13. Szklo, A. & Schaeffer, R. Fuel specification, energy consumption and CO₂ emission in oil refineries. *Energy* **32**, 1075–1092 (2007).
14. ICARUS-4. Sector study for the refineries.

15. Gerlach, S. When the price is right. *Econ. (United Kingdom)* **413**, 1–6 (2017).
16. Babich, I. V. & Moulijn, J. A. Science and technology of novel processes for deep desulfurization of oil refinery streams: A review. *Fuel* **82**, 607–631 (2003).
17. Gerden, T. The adoption of the kyoto protocol of the united nations framework convention on climate change. *Prisp. za Novejšo Zgodovino* **58**, (2018).
18. Mochida, I. & Choi, K. H. An overview of hydrodesulfurization and hydrodenitrogenation. *J. Japan Pet. Inst.* **47**, 145–163 (2004).
19. Hocking, M. B. Sulfur and Sulfuric Acid. *Handb. Chem. Technol. Pollut. Control* 253–287 (2005) doi:10.1016/b978-012088796-5/50012-0.
20. Chandra Srivastava, V. An evaluation of desulfurization technologies for sulfur removal from liquid fuels. *RSC Adv.* **2**, 759–783 (2012).
21. Speight, J. A review of: “Hydrotreating Catalysis Science and Technology” Topsoe, H., Clausen, B.S., and Massoth, F.E., Springer-Verlag New York, 1996. ISBN No. 3-540-603808 Price not available at time of review. *Fuel Sci. Technol. Int.* **14**, 1465–1465 (1996).
22. Abro, R. *et al.* A review of extractive desulfurization of fuel oils using ionic liquids. *RSC Adv.* **4**, 35302–35317 (2014).
23. Gao, J. *et al.* Performance and mechanism for extractive desulfurization of fuel oil using modified polyethylene glycol. *Fuel* **233**, 704–713 (2018).
24. Kuwahara, H. Desulfurization of heavy oil. *Chem. Econ. Engng Rev.* **5**, 35–40 (1973).
25. Seeberger, A. & Jess, A. Desulfurization of diesel oil by selective oxidation and extraction of sulfur compounds by ionic liquids—a contribution to a competitive process design. *Green Chem.* **12**, 602–60 (2010).
26. Dehghan, R. & Anbia, M. Zeolites for adsorptive desulfurization from fuels: A review. *Fuel Process. Technol.* **167**, 99–116 (2017).
27. Mjalli, F. S., Ahmed, O. U., Al-Wahaibi, T., Al-Wahaibi, Y. & AlNashef, I. M. Deep oxidative desulfurization of liquid fuels. *Rev. Chem. Eng.* **30**, 337–378 (2014).
28. Lee, K. X. & Valla, J. A. Adsorptive desulfurization of liquid hydrocarbons using zeolitebased sorbents: A comprehensive review. *React. Chem. Eng.* **4**, 1357–1386 (2019).
29. Silva, J. M. P. F. *et al.* Influence of the chemical composition of silica-alumina adsorbents in sulfur and nitrogen compounds removal from hydrotreated diesel. *Ind. Eng. Chem. Res.* **53**, 16000–16014 (2014).
30. Gawande, P. & Kaware, J. Review on Research for Desulphurization of Diesel By Adsorption. *Int. Res. J. Eng. Technol.* 337–340 (2016).
31. Li, Y. X. *et al.* What matters to the adsorptive desulfurization performance of metal - Organic frameworks? *J. Phys. Chem. C* **119**, 21969–21977 (2015).
32. Attar, A. & Corcoran, W. H. Desulfurization of Organic Sulfur Compounds by Selective Oxidation. 1. Regenerable and Nonregenerable Oxygen Carriers. *Ind. Eng. Chem. Prod. Res. Dev.* **17**, 102–109 (2002).

33. Collins, F. M., Lucy, A. R. & Sharp, C. Oxidative desulphurisation of oils via hydrogen peroxide and heteropolyanion catalysis. *J. Mol. Catal. A Chem.* **117**, 397–403 (1997).
34. Guth. Method for removing sulfur and nitrogen in petroleum oils. (1973).
35. Betiha, M. A., Rabie, A. M., Ahmed, H. S., Abdelrahman, A. A. & El-Shahat, M. F. Oxidative desulfurization using graphene and its composites for fuel containing thiophene and its derivatives: An update review. *Egypt. J. Pet.* **27**, 715–730 (2018).
36. Ismagilov, Z. *et al.* Oxidative Desulfurization of Hydrocarbon Fuels. <http://dx.doi.org/10.1080/01614940.2011.596426> **53**, 199–255 (2011).
37. Houda, S., Lancelot, C., Blanchard, P., Poinel, L. & Lamonier, C. Oxidative desulfurization of heavy oils with high sulfur content: A review. *Catalysts* **8**, (2018).
38. Hossain, M. N., Park, H. C. & Choi, H. S. A comprehensive review on catalytic oxidative desulfurization of liquid fuel oil. *Catalysts* **9**, 1–12 (2019).
39. Morawska, L., Ristovski, Z., Jayaratne, E. R., Keogh, D. U. & Ling, X. Ambient nano and ultrafine particles from motor vehicle emissions: Characteristics, ambient processing and implications on human exposure. *Atmos. Environ.* **42**, 8113–8138 (2008).
40. Diversity, L. Handbook of Hydrocarbon and Lipid Microbiology. *Handb. Hydrocarb. Lipid Microbiol.* (2010) doi:10.1007/978-3-540-77587-4.
41. Soleimani, M., Bassi, A. & Margaritis, A. Biodesulfurization of refractory organic sulfur compounds in fossil fuels. *Biotechnol. Adv.* **25**, 570–596 (2007).
42. Ismail, W. A., Van Hamme, J. D., Kilbane, J. J. & Gu, J. D. Editorial: Petroleum microbial biotechnology: Challenges and prospects. *Front. Microbiol.* **8**, 8–11 (2017).
43. Lee, D., Lowe, D. & Grant, P. Microbiology in the oil patch: A review. *Annu. Tech. Meet. ATM 1996* (1996) doi:10.2118/96-109.
44. Lin, J. *et al.* A study on the microbial community structure in oil reservoirs developed by water flooding. *J. Pet. Sci. Eng.* **122**, 354–359 (2014).
45. Sadare, O. O., Obazu, F. & Daramola, M. O. Biodesulfurization of petroleum distillates—current status, opportunities and future challenges. *Environ. - MDPI* **4**, 1–20 (2017).
46. Leòn, V. & Kumar, M. Biological upgrading of heavy crude oil. *Biotechnol. Bioprocess Eng.* **10**, 471–481 (2005).
47. KODAMA, K. *et al.* Microbial Conversion of Petro-sulfur Compounds. *Agric. Biol. Chem.* **34**, 1320–1324 (1970).
48. Gallagher, J. R., Olson, E. S. & Stanley, D. C. Microbial desulfurization of dibenzothiophene: A sulfur-specific pathway. *FEMS Microbiol. Lett.* **107**, 31–35 (1993).
49. Bahrami, A., Shojaosadati, S. A. & Mohebali, G. Biodegradation of dibenzothiophene by thermophilic bacteria. *Biotechnol. Lett.* **23**, 899–901 (2001).
50. Sohrabi, M., Kamyab, H., Janalizadeh, N. & Huyop, F. Z. Bacterial desulfurization of organic sulfur compounds exist in fossil fuels. *J. Pure Appl. Microbiol.* **6**, 717–729 (2012).

51. Ramírez, C. A. O. The Role of Glycerol in the Degradation of Dibenzothiophene by *Burkholderia* sp. C3. (2017).
52. Alfieri, A. *et al.* Structure of the monooxygenase component of a two-component flavoprotein monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1177–1182 (2007).
53. Barbosa, A. C. C., Neves, R. P. P., Sousa, S. F., Ramos, M. J. & Fernandes, P. A. Mechanistic Studies of a Flavin Monooxygenase: Sulfur Oxidation of Dibenzothiophenes by DszC. *ACS Catal.* **8**, 9298–9311 (2018).
54. Okai, M. *et al.* Crystal structure of dibenzothiophene sulfone monooxygenase BdsA from *Bacillus subtilis* WU-S2B. *Proteins: Structure, Function and Bioinformatics* vol. 85 1171–1177 (2017).
55. Pimviriyakul, P. & Chaiyen, P. Chapter One - Overview of flavin-dependent enzymes. in *Flavin-Dependent Enzymes: Mechanisms, Structures and Applications* (eds. Chaiyen, P. & Tamanoi, F.) vol. 47 1–36 (Academic Press, 2020).
56. Geronimo, I., Nigam, S. R. & Payne, C. M. Desulfination by 2'-hydroxybiphenyl-2-sulfinate desulfinase proceeds: Via electrophilic aromatic substitution by the cysteine-27 proton. *Chem. Sci.* **8**, 5078–5086 (2017).
57. Woo, C. L., Ohshiro, T., Matsubara, T., Izumi, Y. & Tanokura, M. Crystal structure and desulfurization mechanism of 2'-hydroxybiphenyl-2-sulfinic acid desulfinase. *J. Biol. Chem.* **281**, 32534–32539 (2006).
58. Lee, W. C., Ohshiro, T., Matsubara, T., Izumi, Y. & Tanokura, M. Crystal Structure and Desulfurization Mechanism of 2'-Hydroxybiphenyl-2-sulfinic Acid Desulfinase *. *J. Biol. Chem.* **281**, 32534–32539 (2006).
59. Sousa, J. P. M., Neves, R. P. P., Sousa, S. F., Ramos, M. J. & Fernandes, P. A. Reaction Mechanism and Determinants for Efficient Catalysis by DszB, a Key Enzyme for Crude Oil Bio-desulfurization. *ACS Catal.* **10**, 9545–9554 (2020).
60. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. General Performance of Density Functionals. *J. Phys. Chem. A* **111**, 10439–10452 (2007).
61. Gray, K. A., Mrachko, G. T. & Squires, C. H. Biodesulfurization of fossil fuels. *Curr. Opin. Microbiol.* **6**, 229–235 (2003).
62. Kamali, N., Tavallaie, M., Bambai, B., Asghar Karkhane, A. & Miri, M. Site-directed mutagenesis enhances the activity of NADH-FMN oxidoreductase (DszD) activity of *Rhodococcus erythropolis*. doi:10.1007/s10529-010-0254-4.
63. Abin-Fuentes, A. *et al.* Exploring the Mechanism of Biocatalyst Inhibition in Microbial Desulfurization. **79**, 7807–7817 (2013).
64. JJ, K. Microbial biocatalyst developments to upgrade fossil fuels. *Curr. Opin. Biotechnol.* **17**, 305–314 (2006).
65. Abin-Fuentes, A., Mohamed, M. E. S., Wang, D. I. C. & Prather, K. L. J. Exploring the mechanism of biocatalyst inhibition in microbial desulfurization. *Appl. Environ. Microbiol.* **79**, 7807–7817 (2013).

66. Woo, C. L., Ohshiro, T., Matsubara, T., Izumi, Y. & Tanokura, M. Crystal structure and desulfurization mechanism of 2'-hydroxybiphenyl-2-sulfinic acid desulfinase. *J. Biol. Chem.* **281**, 32534–32539 (2006).
67. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. Current computer-aided drug design. *Curr. Comput. Aided Drug Des.* **7**, 146–157 (2011).
68. Neudert, G. & Klebe, G. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein–Ligand Complexes. *J. Chem. Inf. Model.* **51**, 2731–2745 (2011).
69. Levitt, D. G. & Banaszak, L. J. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graph.* **10**, 229–234 (1992).
70. Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.* **13**, 307–308, 323–330 (1995).
71. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–857 (2002).
72. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A powerful approach for structure-based drug discovery.
73. Pagadala, N. S., Syed, K. & Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* doi:10.1007/s12551-016-0247-1.
74. Sci-Hub | Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics*, 12(40), 12899 | 10.1039/c0cp00151a. <https://sci-hub.se/10.1039/C0CP00151A>.
75. Yadava, U. Search algorithms and scoring methods in protein-ligand docking. *Endocrinol. Int. J.* **6**, (2018).
76. Guedes, I. A., Pereira, F. S. S. & Dardenne, L. E. Empirical Scoring Functions for StructureBased Virtual Screening: Applications, Critical Aspects, and Challenges. *Front. Pharmacol.* **0**, 1089 (2018).
77. Shen, Q. *et al.* Knowledge-Based Scoring Functions in Drug Design: 2. Can the Knowledge Base Be Enriched? *J. Chem. Inf. Model.* **51**, 386–397 (2010).
78. Muegge, I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Virtual Screen. An Altern. or Complement to High Throughput Screening?* **20**, 99–114 (2000).
79. De Revisión, A., Prieto-Martínez, F. D., Arciniega, M. & Medina-Franco, J. L. Molecular docking: current advances and challenges-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). TIP Revista Especializada en Ciencias. *TIP Rev.Esp.Cienc.Quím.Biol* **21**, 65–87 (2018).
80. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
81. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

82. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **11**, 3499–3509 (2015).
83. Walter R. P. Scott *et al.* The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **103**, 3596–3607 (1999).
84. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. doi:10.1016/j.neuron.2018.08.011.
85. González, M. A. Force fields and molecular dynamics simulations. *Collect. SFN* **12**, 169–200 (2011).
86. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Advances and Applications in Bioinformatics and Chemistry Dovepress Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinforma. Chem.* **8**, 37–47 (2015).
87. Lazim, R., Suh, D. & Choi, S. Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. *Int. J. Mol. Sci.* **21**, 1–20 (2020).
88. and, D. J. & Praprotnik, M. Molecular Dynamics Integration Time Step Dependence of the Split Integration Symplectic Method on System Density. *J. Chem. Inf. Comput. Sci.* **43**, 1922–1927 (2003).
89. Lustig, R. Statistical thermodynamics in the classical molecular dynamics ensemble. I. Fundamentals. *J. Chem. Phys.* **100**, 3048 (1998).
90. Sharma, S. *et al.* Overview of BIOVIA materials studio, LAMMPS, and GROMACS. *Mol. Dyn. Simul. Nanocomposites using BIOVIA Mater. Stud. Lammmps Gromacs* 39–100 (2019) doi:10.1016/B978-0-12-816954-4.00002-4.
91. S, H. & PA, K. Calculation of the relative binding free energy of 2'GMP and 2'AMP to ribonuclease T1 using molecular dynamics/free energy perturbation approaches. *J. Mol. Biol.* **212**, 197–209 (1990).
92. Clark, A. J. *et al.* Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. *J. Mol. Biol.* **429**, 930 (2017).
93. CX, W., YY, S., F, Z. & L, W. Thermodynamic integration calculations of binding free energy difference for Gly-169 mutation in subtilisin BPN'. *Proteins* **15**, 5–9 (1993).
94. M, L., R, B., Y, W. & JA, M. Independent-Trajectory Thermodynamic Integration: a practical guide to protein-drug binding free energy calculations using distributed computing. *Methods Mol. Biol.* **819**, 469–486 (2012).
95. Sci-Hub | End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chemical Reviews | 10.1021/acs.chemrev.9b00055.
<https://scihub.se/https://doi.org/10.1021/acs.chemrev.9b00055>.
96. Homeyer, N. & Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Mol. Inform.* **31**, 114–122 (2012).
97. T, H., J, W., Y, L. & W, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area

- methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **32**, 866–877 (2011).
98. Wang, J., Hou, T. & Xu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput. Aided-Drug Des.* **2**, 287–306 (2006).
 99. Straatsma, T. P., Berendsen, H. J. C. & Postma, J. P. M. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *J. Chem. Phys.* **85**, 6720–6727 (1986).
 100. Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science*, 1995, 268, 1144-1149. <http://www.sciepub.com/reference/75759>.
 101. Guan, L.-J. *et al.* Crystal structures of apo-DszC and FMN-bound DszC from *Rhodococcus erythropolis* D-1. *FEBS J.* **282**, 3126–3135 (2015).
 102. Li, H., Robertson, A. D. & Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pK_a Values. (2005) doi:10.1002/prot.20660.
 103. Davies, M. N., Toseland, C. P., Moss, D. S. & Flower, D. R. Benchmarking pK_a prediction. (2006) doi:10.1186/1471-2091-7-18.
 104. Schrödinger, LLC. *The {PyMOL} Molecular Graphics System, Version~1.8.* (2015).
 105. Schrödinger, LLC. *The {JyMOL} Molecular Graphics Development Component, Version~1.8.* (2015).
 106. Schrödinger, LLC. *The {AxPyMOL} Molecular Graphics Plugin for {Microsoft PowerPoint}, Version~1.8.* (2015).
 107. Hassan, N. M., Alhossary, A. A., Mu, Y. & Kwoh, C.-K. Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration OPEN. doi:10.1038/s41598-017-15571-7.
 108. Zhang, L. *et al.* Structural insights into the stabilization of active, tetrameric DszC by its Cterminus. *Proteins Struct. Funct. Bioinforma.* **82**, 2733–2743 (2014).
 109. Strelow, J. *et al.* Mechanism of Action Assays for Enzymes. *Assay Guid. Man.* (2012).

Annexes

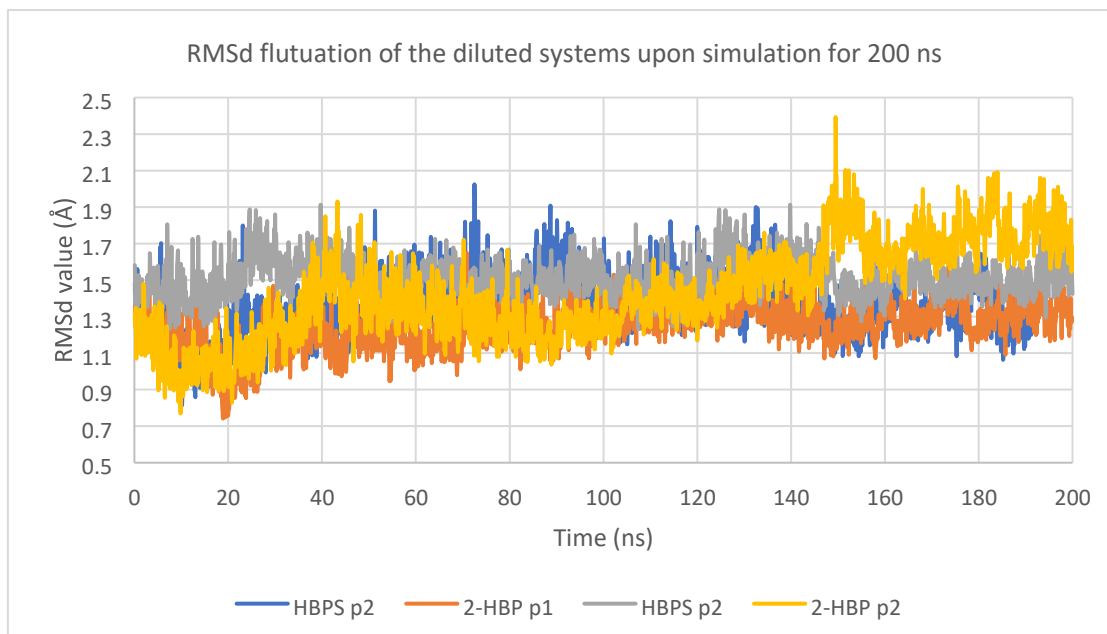


Figure 48 - Graphical representation of the backbone RMSd fluctuation of the diluted systems upon 200 ns simulation. In the legend, "p1" and "p2" refer to systems submitted to protocol 1 and protocol 2, respectively.

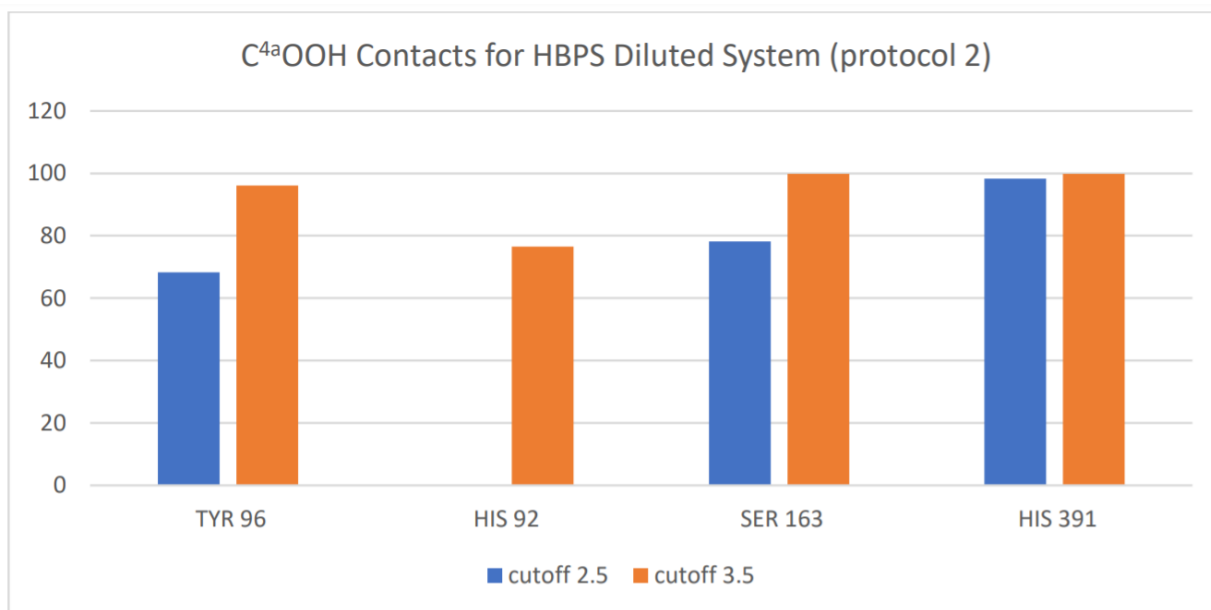


Figure 49 - C^{4a}OOH contacts with the most relevant residues for the reaction mechanism..