



University of Kentucky
UKnowledge

Theses and Dissertations--Chemical and
Materials Engineering

Chemical and Materials Engineering

2018

UNDERSTANDING INHIBITION OF A BIODESULFURIZATION ENZYME TO IMPROVE SULFUR REMOVAL FROM PETROLEUM

Yue Yu

University of Kentucky, amirayuyue@gmail.com

Digital Object Identifier: <https://doi.org/10.13023/ETD.2018.024>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Yu, Yue, "UNDERSTANDING INHIBITION OF A BIODESULFURIZATION ENZYME TO IMPROVE SULFUR REMOVAL FROM PETROLEUM" (2018). *Theses and Dissertations--Chemical and Materials Engineering*. 81.

https://uknowledge.uky.edu/cme_etds/81

This Doctoral Dissertation is brought to you for free and open access by the Chemical and Materials Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Chemical and Materials Engineering by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Yue Yu, Student

Dr. Christina M. Payne, Major Professor

Dr. Thomas Dziubla, Director of Graduate Studies

UNDERSTANDING INHIBITION OF A BIODESULFURIZATION ENZYME
TO IMPROVE SULFUR REMOVAL FROM PETROLEUM

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering at the University of Kentucky

By

Yue Yu

Lexington, Kentucky

Co-directors: Dr. Christina M. Payne, Adjunct Associate Professor of Chemical and
Materials Engineering
and Dr. Barbara L. Knutson, Professor of Chemical and Materials Engineering

Lexington, Kentucky

Copyright © Yue Yu 2017

ABSTRACT OF DISSERTATION

UNDERSTANDING INHIBITION OF A BIODESULFURIZATION ENZYME TO IMPROVE SULFUR REMOVAL FROM PETROLEUM

The biodesulfurization 4S-pathway is a promising complementary enzymatic approach to remove sulfur from recalcitrant thiophenic derivatives in petroleum products that remain from conventional hydrodesulfurization method without diminishing the calorific value of oil. The final step of this pathway involves the carbon-sulfur bond cleavage from HBPS, and the production of the final products 2-hydroxybiphenyl (HBP) and sulfite, has been recognized as the rate-limiting step, partially as a result of product inhibition. However, the mechanisms and factors responsible for product inhibition in the last step have not been fully understood. In this work, we proposed a computational investigation using molecular dynamic simulations and free energy calculations on 2'-hydroxybiphenyl-2-sulfinate (HBPS) desulfinase (*DszB*) with different bound ligands as well as different solvent conditions to develop a fundamental understanding of the molecular-level mechanism responsible for product inhibition. Based on available crystal structures of *DszB* and biochemical characterization, we proposed a "gate" area close to substrate binding site of *DszB* is responsible for ligand egress and plays a role in product inhibition. We have conducted biphasic molecular dynamic simulations to evaluate the proposed gate area functionality. Non-bonded interaction energy analysis shows that hydrophobic residues around the gate area produce van der Waals interactions inhibiting translocation through the gate channel, and therefore, the molecules are easily trapped inside the binding site. Umbrella sampling molecular dynamics was performed to obtain the energy penalty associated with gate conformational change from open to close, which was 2.4 kcal/mol independent of solvent conditions as well as bound ligands. Free energy perturbation calculations were conducted for a group of six selected molecules bound to *DszB*. The selections were based on functional group representation and to calculate binding free energies that were directly comparable to experimental inhibition constants, K_I . Our work provides a fundamental molecular-level analysis on product inhibition for the biodesulfurization 4S-pathway.

KEYWORDS: Oil-refinery, Product Inhibition, Competitive Inhibition, Free Energy,
Molecular Dynamic

Yue Yu

Dec. 6th, 2017

UNDERSTANDING INHIBITION OF A BIODESULFURIZATION ENZYME
TO IMPROVE SULFUR REMOVAL FROM PETROLEUM

By

Yue Yu

Dr. Christina M. Payne

Co-Director of Dissertation

Dr. Barbara L. Knutson

Co-Director of Dissertation

Dr. Thomas Dziubla

Director of Graduate Studies

Dec. 6th, 2017

DEDICATION

To my beloved parents

ACKNOWLEDGMENTS

I would like to give my sincere thanks to my funding source, the American Chemical Society Petroleum Research Fund (53861-DNI4), as this doctoral research would not be possible without their support. In addition, my computational time for this research was provided by the National Science Foundation (NSF) Extreme Science and Engineering Discovery Environment (XSEDE) ACI-1053575 and the University of Kentucky.

Besides financial support, I would like to give my greatest gratitude to my advisor Dr. Christina M. Payne for her suggestions and endless help. She is always ready for every question and inquiry. Her inspiration and advice are the most important parts of my work.

I would like to extend the deepest thanks to my colleagues Cris, Jana, and Abhishek, for being so supportive. I was lucky that I could work with them, and I enjoyed the time we spent together solving the various puzzles related to our projects.

Finally, I want to thank my family and friends for giving me endless love and encouragement. Special thanks to Mingyang Sun, who gives me lots of support and motivation. Thank you again for supporting me and being on my side.

Table of Contents

ACKNOWLEDGMENTS	iii
Table of Contents	iv
List of Tables	viii
List of Figures	ix
Chapter 1 — Introduction	1
1.1 Motivation	1
1.2 Research Background	3
1.2.1 Biodesulfurization 4S pathway.....	3
1.2.2 Biochemical characterization of DszB	5
1.2.3 Structural characterization of DszB.....	8
1.2.4 Catalytic mechanism of desulfination by <i>DszB</i>	10
1.3 Outline of dissertation	14
1.3.1 Dynamic contributions to substrate specificity and inhibition (Chapter 3 , and Chapter 4).....	14
1.3.2 Quantify product and competitive inhibition effects with absolute binding free energy (Chapter 5)	15
Chapter 2 — Computational Methods	17
2.1 Introduction	17
2.2 Molecular Dynamics simulation	17
2.2.1 Initialization of MD simulation	17
2.2.2 The force calculation.....	18
2.2.3 Equation of motions	19

2.2.4	Energy function.....	20
2.2.5	Periodic boundary conditions.....	22
2.3	Free energy calculation methods.....	23
2.3.1	Basic approach to free energy calculation.....	23
2.3.2	Free energy perturbation with replica exchange molecular dynamics (FEP/ λ -REMD) 25	
2.3.3	Umbrella sampling.....	28
Chapter 3 — CHARMM force field parameters for 2'-hydroxybiphenyl-2-		
sulfinate, 2-hydroxybiphenyl, and related analogs.....		
3.1	Summary.....	32
3.2	Introduction.....	33
3.3	Methods.....	36
3.3.1	Parameterization.....	36
3.3.1.1	Initialize parameters by analogy with CGenFF.....	38
3.3.1.2	Geometry optimization.....	39
3.3.1.3	Charge optimization.....	39
3.3.1.4	Bond and angle optimization.....	41
3.3.1.5	Dihedral optimization.....	41
3.3.2	Validation of optimized parameters.....	43
3.3.2.1	Explicit solvent molecular dynamic simulations.....	43
3.3.2.2	Preparation of HBPS and experimental IR spectra.....	44
3.3.2.3	Calculated IR spectra.....	45
3.4	Results and discussions.....	45
3.4.1	Geometry optimization.....	45
3.4.2	Optimized parameters.....	47

3.4.3	Comparing QM and MM equilibrium bond and angle values	52
3.4.4	Comparison of calculated and experimental IR spectra.....	54
3.5	Conclusions	60
Chapter 4 — Conformational change of 2'-hydroxybiphenyl-2-sulfinate		
desulfinase upon ligand binding		
		62
4.1	Summary	62
4.2	Introduction.....	63
4.3	Methods and Materials.....	65
4.3.1	Molecular dynamic simulation	65
4.3.2	Principle component analysis (PCA).....	67
4.3.3	Clustering.....	68
4.3.4	Umbrella sampling.....	69
4.4	Results and Discussions.....	70
4.4.1	MD simulations of <i>DszB</i> -HBPS and <i>DszB</i> -HBP in aqueous solution.....	70
4.4.2	<i>DszB</i> -HBP in hexane-water biphasic solution	79
4.4.3	Umbrella Sampling	81
4.5	Conclusions	83
Chapter 5 — Thermodynamic relationships between 2'-hydroxybiphenyl-2-		
sulfinate desulfinase and putative inhibitors		
		85
5.1	Summary	85
5.2	Introduction.....	86
5.3	Method	88
5.3.1	Molecular dynamic simulations for putative inhibitors bound to <i>DszB</i>	88
5.3.2	Free energy calculation FEP/ λ -REMD	90

5.4	Results and discussions	93
5.4.1	Putative inhibitors binding inside <i>DszB</i>	93
5.4.2	Free energy calculations from FEP/ λ -REMD compared with experimentally determined parameters.....	97
5.4.3	Protonation state of R70	99
5.5	Conclusions	101
Chapter 6	— Conclusions and future directions.....	102
Appendix	108
References	139
Vita	149

List of Tables

Table 3.1 Dihedrals of each molecule included in the optimization process.....	42
Table 3.2 Optimized partial atomic charge parameters for each molecule.....	49
Table 3.3 Optimized bond and angle parameter for each molecule.....	50
Table 3.4 Optimized dihedral parameters for each molecule.....	51
Table 3.5 Experimental IR spectra frequency values compared with theoretical predictions.....	55
Table 5.1 Binding free energies calculated from FEP/ λ -REMD.....	98
Table 5.2 Binding free energy of deprotonated R70 for non-inhibitory molecules.....	101

List of Figures

Figure 1.1 Schematic of the 4S pathway for the biodesulfurization of dibenzothiophene (DBT).....	5
Figure 1.2 Representatives of aromatic analogs with different inhibitory behaviors on <i>DszB</i>	7
Figure 1.3 Known and putative substrates of <i>DszB</i>	8
Figure 1.4 Superimposition of bound color in red (with PDB code 2DE3) and unbound <i>DszB</i> color in yellow (with PDB code 2DE2) structures revealed a significant conformational change.	9
Figure 1.5 Active site models based on <i>DszB</i> -HBPS crystal structure with PDB code 2DE3, with C27S mutation being manually changed back to cysteine.	11
Figure 1.6 Proposed nucleophilic addition mechanism by Lee <i>et. al.</i> [39].	13
Figure 1.7 Proposed electrophilic aromatic substitution mechanism by Gray, <i>et. al.</i>	13
Figure 1.8 Possible reactant transition states, only polar hydrogen atoms are shown.	14
Figure 2.1 Illustration of periodic boundary conditions.	23
Figure 2.2 Thermodynamic cycle for calculating binding free energy ΔG_b° with FEP/ λ -REMD.	26
Figure 2.3 Schematic representation of parallel tempering simulation of FEP/ λ -REMD protocol.	28
Figure 3.1 Flow chart illustrating the ffTK parameterization process used to obtain MM force fields for the selected molecule.	37

Figure 3.2 Optimized geometry for each molecule obtained using Gaussian09 with the MP2/6-31++G** level of theory.....	47
Figure 3.3 Torsional profiles of HBP, HBPS, NTAM, BCA, BIPH, and NAPO, as labeled at the top of each plot.....	48
Figure 3.4 Comparison of the QM-optimized bonds and angles to the bond distances and angles sampled over the course of a 100-ns MD simulation.....	54
Figure 3.5 Experimental and theoretical IR spectra for HBP, HBPS, BCA, BIPH, NTAM, and NAPO.	60
Figure 4.1 (a) (b) PCA of <i>DszB</i> -HBPS and <i>DszB</i> -HBP MD trajectories, respectively. (c)(d) RMSD with time shown with the clustering analysis-suggested groupings along the corresponding time step.....	71
Figure 4.2 (a) (b) Superimposition of group structures 1 and 4 for <i>DszB</i> -HBPS and <i>DszB</i> -HBP in aqueous solution.	73
Figure 4.3 Secondary structure analysis along the MD trajectories. Structure analysis was performed using VMD.....	75
Figure 4.4 Non-bonded interaction energies of <i>DszB</i> with HBPS and HBP. Electrostatic interactions (blue) and VDW interactions (red) of the ligand and surrounding residues in the active site.....	78
Figure 4.5 (Left) Hydrogen bonds formed along the MD trajectory between the ligand and nearby protein residues inside binding pocket of <i>DszB</i> . (Right) All possible hydrogen bonds formed during MD simulation with the ligand and nearby residues.	79

Figure 4.6 (Left) Hexane distribution around the <i>DszB</i> protein surface. The y-axis represents the number of hexane molecules within a 9Å cutoff distance from the <i>DszB</i> surface at 200 ns of each MD simulations. (Right) MD simulation snapshots of HBP position change.....	81
Figure 4.7 Potential of Mean Force (PMF) plots from umbrella sampling MD simulations.....	83
Figure 5.1 Docking calculations predicted binding positions for selected inhibitors (BIPH, BCA, NTAM, and NAPO) inside binding pocket of <i>DszB</i>	90
Figure 5.2 Putative inhibitor binding positions inside <i>DszB</i>	94
Figure 5.3 Non-inhibitory molecule bound positions inside <i>DszB</i>	95
Figure 5.4 Root mean square fluctuations from MD simulation revealed the non-inhibitory molecules bound inside binding pocket resulted in different conformational dynamic changes of protein.	96
Figure 5.5 Multiple sequence alignment of <i>DszB</i> homologs.....	100
Figure 6.1 (A) Number of non-polar residues located at the gate regions.	105

Chapter 1 — Introduction

1.1 Motivation

Fossil fuels are the most widely used as a transportation fuel source. Even though the percentage of energy obtained from fossil fuels declined in recent years, over 20% of energy production in the United States is still from fossil fuels, despite diminishing world reserves of high-quality crude oil and coal [1]. Dwindling sweet crude oil supplies have increased reliance upon on impurity-rich (high-sulfur) North American crude oil to meet increasing demand for liquid fuels. The sulfur content of petroleum from different sources varies from 0.025%-5%, in the form of elemental sulfur, sulfate, sulfite, thiosulfate, and sulfide [2]. There are more than 200 organic sulfur-containing compounds that have been identified in crude oil, which include sulfides, thiols, thiophenes, substituted benzo- and dibenzothiophenes, and many more complex molecules. Crude oil from Texas has been reported to contain up to 70% organic sulfur, a majority of which is in the form of dibenzothiophene (DBT) [3].

Heavy reliance upon sulfur-rich crude oil poses both environmental and health risks, as desulfurization technology fails to completely remove sulfur from finished fuels. The primary objection to the combustion of sulfur-containing crude oil is the generation of sulfur oxides, which play a major role in environmental acid deposition [4]. Moreover, sulfur containing crude oils have also been implicated in pipeline corrosion, and the removal of recalcitrant sulfur can require considerable time and expense [2]. Combined with political pressures on decreasing dependence

on foreign oil and increasingly stringent fuel quality standards, the need to address effectiveness of desulfurization technology is stronger than ever [4-7].

Currently the most widely adopted desulfurization method is hydrodesulfurization, which effectively removes sulfur from light refinery fractions utilize molybdenum or other expensive metal catalysts in a sulfur linked network at high temperatures and pressures to bind organic sulfurs and cleave them from oil molecules. The sulfur is then released from the metal catalyst with hydrogen gas to form hydrogen sulfide. While hydrodesulfurization effectively removes simple sulfur compounds, including thiophene, more complex thiophenic molecules like dibenzothiophene (DBT) and its derivatives in the middle distillate and heavy oil fractions remain intact; these thiophenic molecules now account for 60% of sulfur emissions from liquid transportation fuels [8-10].

Biodesulfurization is a complementary approach to conventional hydrodesulfurization, given its capability to selectively remove organosulfurs that are recalcitrant to conventional methods [6, 11-13]. Unlike the energy and capital-intensive hydrodesulfurization process, biodesulfurization processes operate at ambient conditions and require few process modifications for implementation alongside existing operations [14]. Therefore, biodesulfurization potentially can be effective, when used in conjunction with existing hydrodesulfurization unit operations or oxidative desulfurization, in reducing sulfur content in heavy oil fractions in order to meet the regulatory standards [8, 15].

However, the slow overall reaction rate of biodesulfurization remains a primary technical challenge to economical implementation of biodesulfurization [16, 17].

The most widely studied biodesulfurization route comes from *Rhodococcus erythropolis*, which is a bacterium that uses a four-step reaction process by four different enzymes to remove sulfur from DBT. Within the oil fraction at the oil-water interface, after the four-step reaction, thiophenic sulfur degrades to the detected form of sulfite (SO_3^{2-}) [18]. The final step of the four-step reaction, is catalyzed by 2'-hydroxybiphenyl-2-sulfinate (HBPS) desulfinase (*DszB*) and has been recognized as having the slowest reaction rate in the pathway (the overall rate-limiting step), making it the most logical target for improvement [19, 20]. An additional hindrance to the overall conversion rate of *DszB* is product inhibition, where the final product, 2-hydroxybiphenyl (HBP), is responsible for the observed reduction in biocatalyst activity concomitant with HBP generation [20-22].

Currently, *DszB* activity enhancements are based on increased expression or isolation of new strains but overlook rational design as an avenue to enhanced performance. Therefore, the primary focus in our research is to address this technological challenge by elucidating the catalytic and substrate binding mechanisms of *DszB*, enabling rational design for enhanced performance [21, 23-26]. More specifically, we will address the hypothesis that *DszB* inhibition is affected by the chemical composition and dynamics of active site.

1.2 Research Background

1.2.1 Biodesulfurization 4S pathway

There are two primary pathways involving the microbial utilization of dibenzothiophene (DBT) from petroleum: ring-destructive and sulfur-specific. Among the two, the sulfur-specific “4S” pathway is the only commercially relevant

mechanism, as the aromatic ring remains intact after cleavage of carbon-sulfur bond so that the heating value of the fuel is unchanged (Figure 1.1) [18, 27]. It involves four sulfur (S)-oxidized chemical intermediates and requires 4 moles of nicotinamide adenine dinucleotide (NADH) [28]. This pathway has been observed in a wide range of bacterial species (the most extensively studied one is from *Rhodococcus erythropolis* IGTS8) and is controlled by three plasmid produced genes *dszABC* and one chromosomal gene *dszD* [29, 30]. Therefore, these genes code for four different enzymes that conduct the 4S biodesulfurization reaction. In the first three steps of the 4S pathway, *DszC* monooxygenase catalyzes two consecutive oxidation steps: converting DBT into DBT-sulfoxide (DBTO) and DBT-sulfone (DBTO₂). This is followed by *DszA* monooxygenase catalyzing the oxidative reaction and breaking the C-S bond in DBTO₂ to convert it into 2'-hydroxybiphenyl-2-sulfinate (HBPS); all three steps require the *DszD* oxidoreductase, which uses nicotinamide adenine dinucleotide (NADH to NAD⁺) to reduce flavin mononucleotide (FMN to FMNH₂) in support of *DszC* and *DszA*. Finally, in the last step, *DszB* desulfinate cleaves the C-S bond of the HBPS intermediate to produce 2-hydroxybiphenyl (HBP) and sulfite (SO₃²⁻) [18, 31, 32]. The HBP product is reincorporated in the oil fraction, and sulfite is removed in the aqueous fraction. The *DszB*-catalyzed desulfurization of HBPS to form HBP and sulfite is a very unusual reaction from a biochemical perspective, because unlike other enzymes that catalyze desulfurization, e.g., cysteine sulfinate desulfinate (CSD) [33] and L-aspartate β-decarboxylase [34], *DszB* is not assisted by pyridoxal 5'-phosphate or any other cofactors [19, 21]. Again, this final catalytic step has been identified as the

rate-limiting step, making *DszB* the most logical target for biodesulfurization rate improvements.

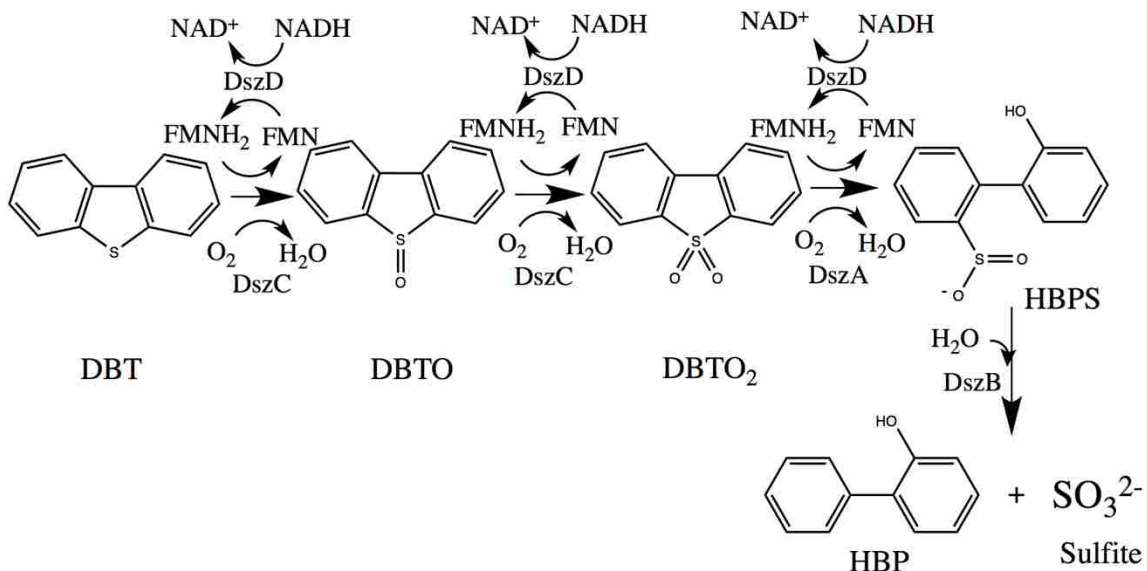


Figure 1.1 Schematic of the 4S pathway for the biodesulfurization of dibenzothiophene (DBT). This figure has been reproduced, with permission, from Yu, *et al.* [35], Copyright © 2017, Elsevier.

1.2.2 Biochemical characterization of *DszB*

DszB was first isolated and purified from the *Rhodococcus erythropolis* IGTS8 strain [21]. For implementation in industrial biodesulfurization processes, *Rhodococcus* strains are unique in that they are not plagued by the mass transport rate limitations seen from other bacterial strains [22, 36, 37]. The molecular mass for *DszB*, a monomer, is about 40 kDa, and it is a colorless protein as isolated.

Inhibition of *DszB* by its primary product, HBP, varies significantly based on environmental conditions. Watkins, *et al.* found that no inhibition by HBP was observed at low concentrations (less than 100 μM) in a buffer of HBPS, sodium

phosphate, and sodium chloride [20]; however, concentrations much above this behave to the contrary [22]. Caro, *et al.* found that inhibition by HBP was avoided up to 140 μM when using a biphasic buffer with oil/water media, although no inhibition constant was reported [22].

Inhibition by HBPS analog aromatic representatives commonly found in petroleum have also been investigated for inhibitory effects [20]. Figure 1.2 illustrates representatives from each functional group representatives, for a single functional group attached, such as 2-biphenyl carboxylic acid (BCA), were consistently found to be non-inhibitory at concentrations up to 100 μM . Within the two-functional group class, the most inhibitory of the studied analogs was 2,2-biphenol (BIPH) ($K_i = 17\mu\text{M}$). In the planar naphthenic class, the most strongly inhibiting compound was 1,8-naphthosultam (NTAM) ($K_i = 1.8 \mu\text{M}$). On the other hand, 1,8-naphthosultone (NAPO) was not inhibitory [20]. In general, biochemical analysis suggested that *DszB* preferentially binds bifunctional biphenyls, and free rotation between the two rings is not critical to formation of binding interactions [20, 21].

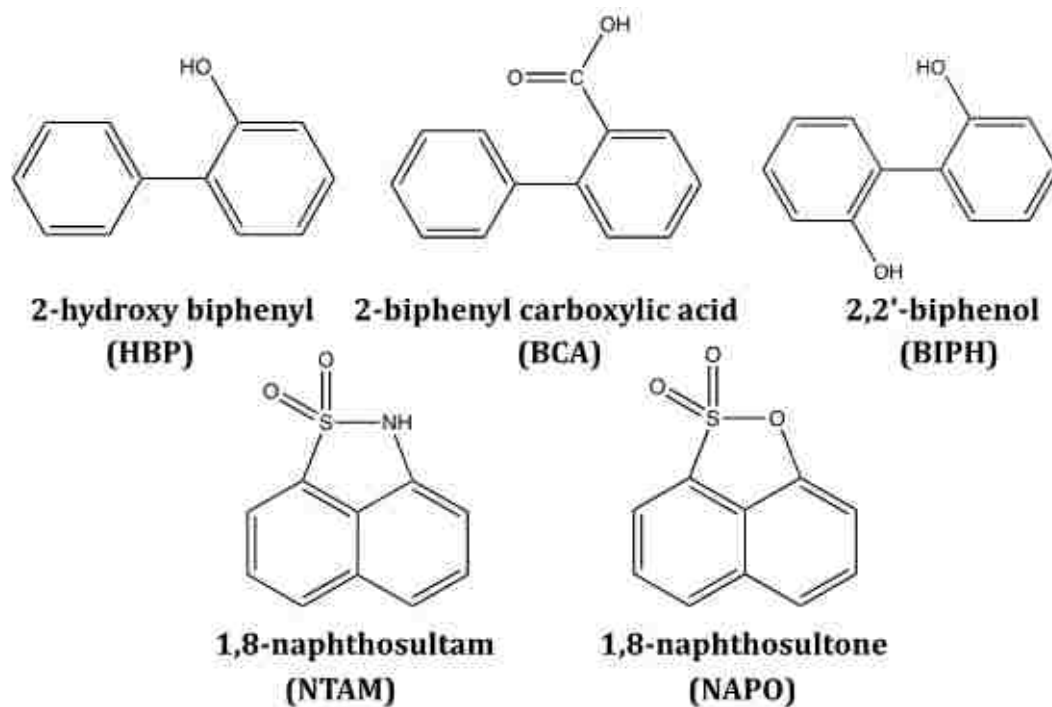


Figure 1.2 Representatives of aromatic analogs with different inhibitory behaviors on *DszB*. HBP is the reaction product and a K_I of 0.5 μM under various conditions [20]. BCA and NAPO are non-inhibitory, yet not productive. BIPH and NTAM inhibit *DszB* with K_I values of 17 μM and 1.8 μM , respectively [20].

Substrate specificity within the 4S pathway is invaluable from the perspective of specifically targeting thiophenic compounds that are recalcitrant to hydrodesulfurization. Activity assays suggest that *Rhodococcus* microbial strains are generally capable of acting on DBT and DBT-derivatives with alkyl- or aryl-substitutions to the monophenols [38]. However, within the 4S pathway, *DszB* has an extremely narrow specificity with no substrate other than HBPS having been demonstrated biochemically [20]. Therefore, understanding the active site interactions will be essential for rational design variants without adversely impacting bio-catalysis.

1.2.3 Structural characterization of DszB

Specificity of *DszB* lead to the substrates selections limitations [32, 39]. *DszB* can accept biphenyl-2-sulfinic acid (BPS) as a substrate, in addition to HBPS, and Lee, *et al.*, demonstrated that BPS is a putative alternate substrate through crystallization of the bound complex [21, 39]. Figure 1.3 illustrates the structural similarity of HBPS and BPS.

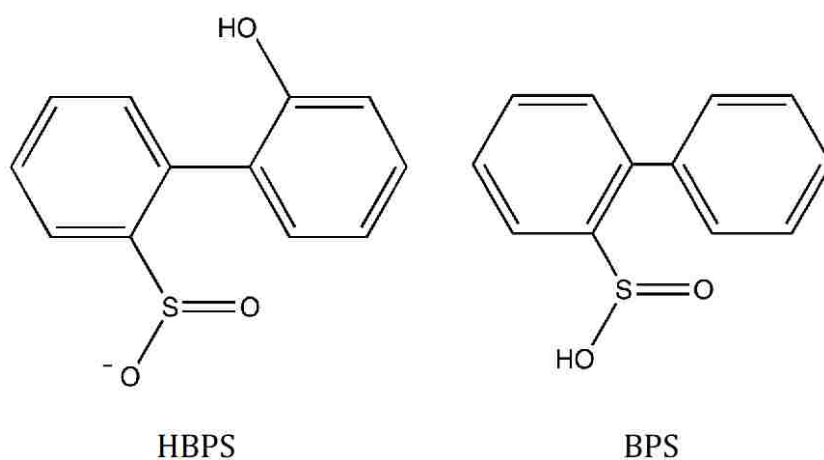


Figure 1.3 Known and putative substrates of *DszB*: 2-hydroxybiphenyl-2-sulfinate (HBPS) and biphenyl-2-sulfinic acid (BPS), respectively.

Substrate binding in *DszB* appears to involve a significant conformational change in the enzyme that may be important for positioning the catalytic residues in a productive conformation within active site [39]. Details of this conformational change, as well as the relationship to product inhibition, will be discussed in Chapter 4.

Lee *et al.*, solved several separate structures in their investigation: the unbound *DszB* structure with protein data bank (PDB) code 2DE2 and C27S, catalytically inactive-variant structure bound to both HBPS PDB code 2DE3 and BPS with PDB code 2DE4 [39]. Superimposition of the available crystal structures revealed that

the previously unstructured loop 1 and loop 3 gained α -helical structure upon substrate binding, and loop 2 moved upwards. This structure change upon binding also introduced H60 into the active site and exposed the mutation, S27, to the substrate (Figure 1.4). C27 was previously confirmed as the catalytic acid, and *DszB* is strongly inhibited by cysteine-modified reagents, and mutation of the catalytic cysteine to serine abolished the activity [21]. Therefore, we hypothesized that binding induced conformational change in the highlighted loop region forms a “gate” that is responsible for ligand egress from active site, and loops in this region may be vital to substrate specificity and catalysis.

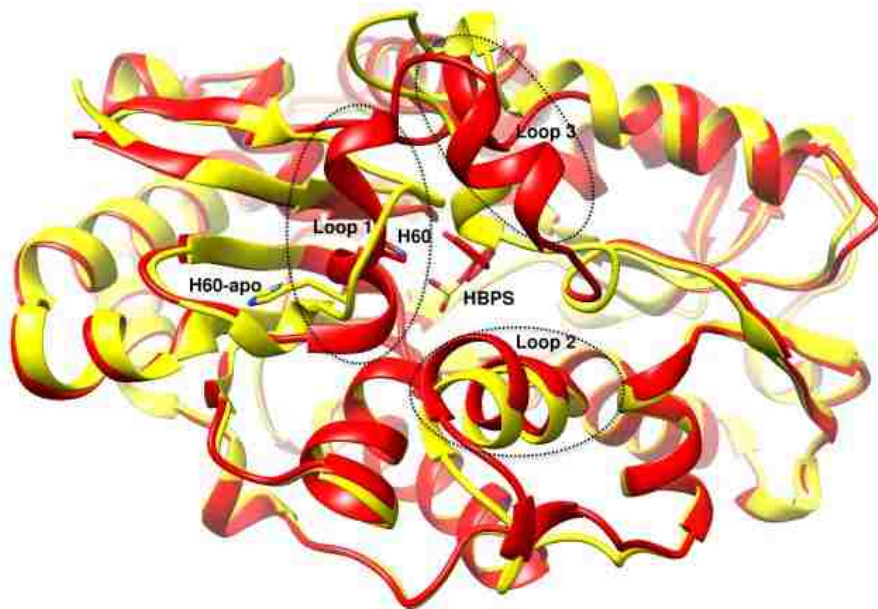


Figure 1.4 Superimposition of bound color in red (with PDB code 2DE3) and unbound *DszB* color in yellow (with PDB code 2DE2) structures revealed a significant conformational change, highlighted by three labeled loops. Upon substrate binding, loop 1 changes from a loop with no secondary structure elements to an α -helix, as does loop 3, and loop 2 moves slightly upwards toward loops 1 and

3. H60 is shown in stick to illustrate the significant change in position upon substrate binding.

1.2.4 Catalytic mechanism of desulfination by *DszB*

It was established that mutation of C27 to serine inactivated the *DszB*, thus, C27 is critical to activity [21]. Based on active site orientation showed in Figure 1.5, H60 and R70 are also thought to be involved in the reaction as they were hydrogen bonds with HBPS based on available *DszB*-HBPS structure with PDB code 2DE3. H60 moved to active site involved a big conformational change and potential gate formation, H60 mutation to glutamine does not fully abolish *DszB* activity but reduced it by ~17 fold [39]. In the same study, R70 mutation to isoleucine and lysine resulted in *DszB* losing detectable activity. Moreover, R70 is part of highly conserved motif, RXGG among *DszB* homologs, therefore, R70 combined with backbone of G73 are very critical for active site structure. As a result, all mentioned residues potentially participated in the desulfurization reaction.

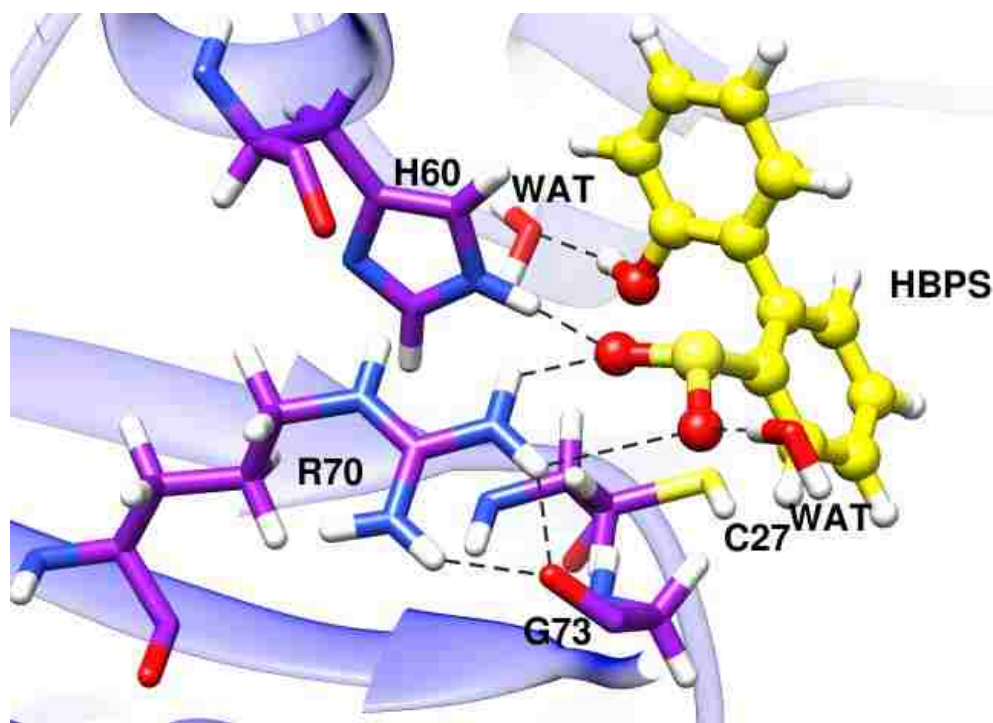


Figure 1.5 Active site models based on *DszB*-HBPS crystal structure with PDB code 2DE3, with C27S mutation being manually changed back to cysteine. Substrate HBPS was in stick representation colored in yellow; C27, H60, R70, and G73 were colored in purple with potential hydrogen-bonding with HBPS labeled with dashed lines; Crystallographic water molecules (labeled as WAT) were also included indicated the potential hydrogen-bonding with HBPS

There are two proposed reaction mechanisms for desulfination of HBPS by *DszB*: (1) Lee, *et. al.* suggested a mechanism based on *DszB* bound with HBPS crystal structure (PDB code 2DE3), and it involved nucleophilic attack on the sulfinate sulfur of HBPS by C27 to break the C-S bond and form a thiosulfonate-like intermediate as a plausible first step showed in Figure 1.6 [39]. Due to the large distance ($\sim 17\text{\AA}$) between H60 and C27 at the absence of substrate, it was unlikely for ion pair formation between the two residues, therefore, sulfinate group from

HBPS would serve as a general base similar to the role of histidine in the cysteine-histidine ion pair of cysteine protease [40]. Following with hydrolysis reaction to subsequently release the final product bisulfite; (2) Before structural resolution of *DszB*, Gray, *et. al* also proposed a mechanism (Figure 1.7) based on tyrosine phenolase model [41], which involved electrophilic substitution of the sulfinate group by C27 proton [32]. The released SO₂ reacts with solvent H₂O to form HSO₃⁻ and H⁺. The latter mechanism has been suggested by Geronimo, *et. al.* to be the most feasible pathway for desulfination of HBPS by *DszB* through molecular dynamic (MD) simulation and density function theory (DFT) calculations, which indicated the direct release of SO₂ without formation of an arenium ion (σ -complex) [42]. The reaction was suggested endothermic by 7.0 kcal/mol, with reaction free energy ΔG_r equals to 2.8 kcal/mol [42].

The transition state for electrophilic substitution mechanism was determined to have a distance between C27-H and HBPS-C of 1.3 Å with energy equal to 26 kcal/mol. Moreover, the transition state directly led to SO₂ release without formation of an arenium ion, as determined through intrinsic reaction coordinate calculations [42]. The proposed mechanism follows a one-step, concerted pathway that has been reported for electrophilic aromatic substitution reactions similar with halogenation with Cl₂ [32, 42, 43].

Other residues besides C27, including H60, R70, and G73, also play important roles in stabilizing transition states during the reaction (Figure 1.8). Geronimo, *et al.*, suggested that H60 plays the most important role in lowering the activation enthalpy to desulfination by withdrawing negative charge from C27 [42]. On the

other hand, R70 and G73 increase the activation enthalpy by shifting the transition state more towards to the product-like character [42].

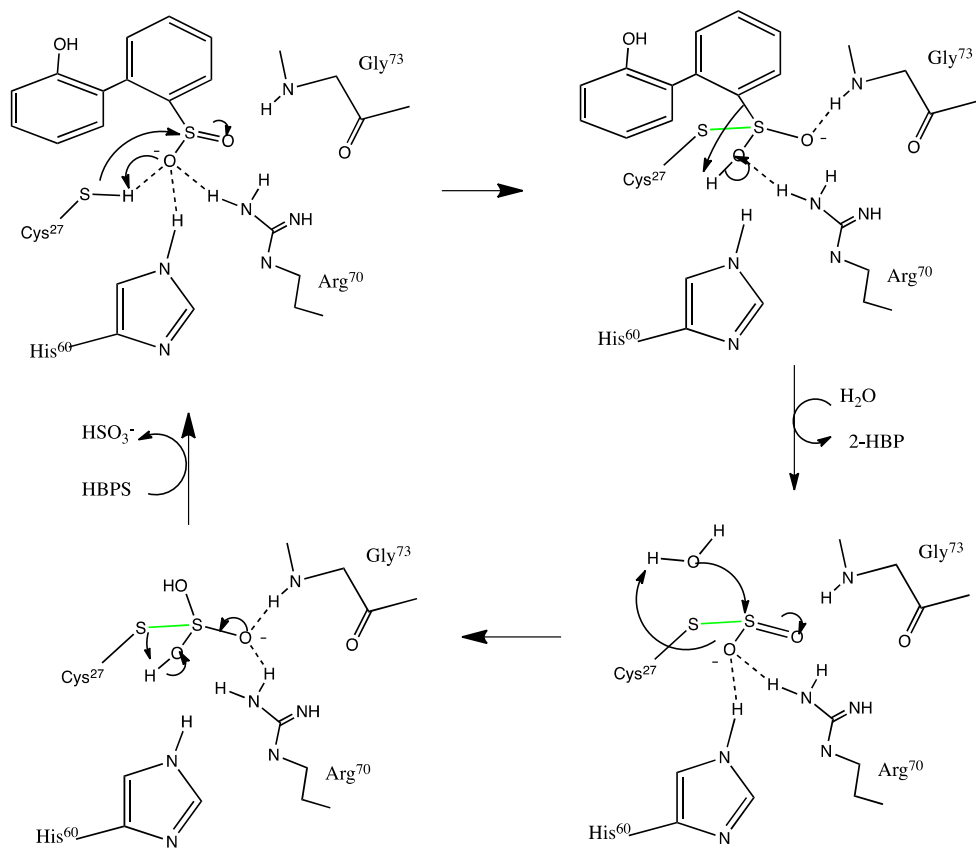


Figure 1.6 Proposed nucleophilic addition mechanism by Lee *et. al.* [39].

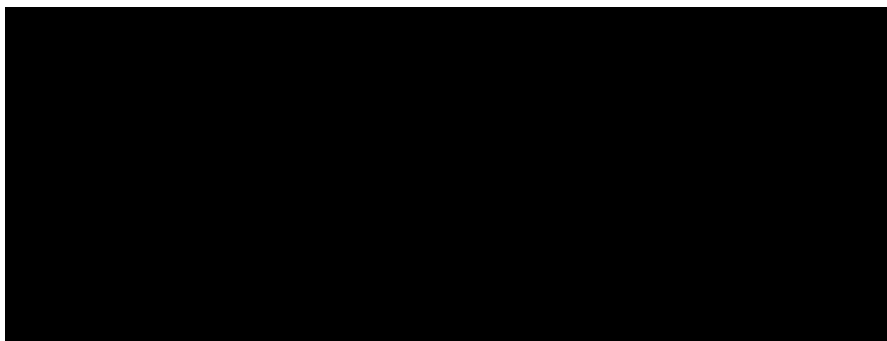


Figure 1.7 Proposed electrophilic aromatic substitution mechanism by Gray, *et. al.* [32]. Residues H60, R70, G73 do not directly participate in the reaction, rather, stabilize the transition states during the reaction [42].

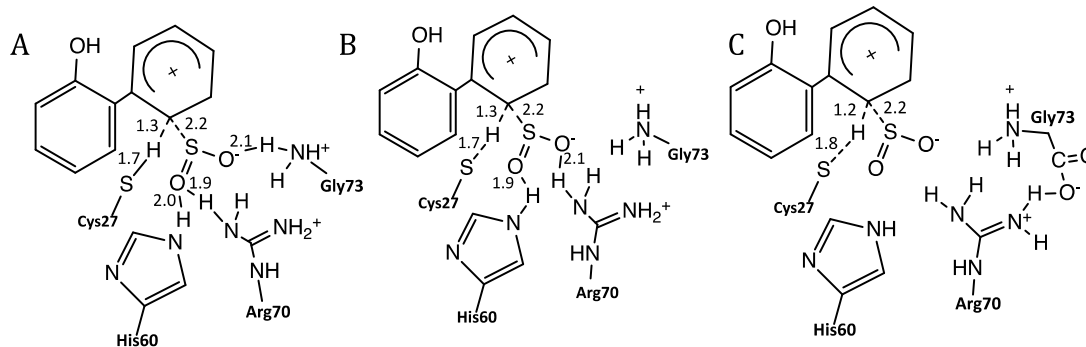


Figure 1.8 Possible reactant transition states, only polar hydrogen atoms are shown [42]. The labeled distances are represented in Å. This figure is modified from Geronimo, *et. al.*[42].

1.3 Outline of dissertation

To develop a fundamental understanding of the molecular-level mechanisms underlying substrate recognition and specificity and product inhibition, we will develop molecular models and perform free energy calculations of *DszB* to identify active interactions with a variety of selected ligands and potential inhibitors. By identifying characteristic behaviors within the active site relative to the bound molecules, we expect to uncover chemical and dynamical relationships contributing to specificity and inhibitions. These observations will serve as foundations for future rational design to improve *DszB* performances.

1.3.1 Dynamic contributions to substrate specificity and inhibition (Chapter 3 , and Chapter 4)

Fully atomistic classical MD simulations of *DszB* bound with different ligands will be performed to understand the relationship between protein dynamics and inhibition and specificity. In Chapter 3, I describe force field parameter development

for the selected six molecules, including HBPS (Figure 1.2), that was necessary for MD simulations. The optimized parameters were tested in both explicit and implicit solutions to assess the abilities of optimized parameter to recapitulate optimized geometries. Calculated infrared spectra (IR) were obtained and compared with experimental IR spectra for validation of the optimized parameters. In Chapter 4, we use the optimized parameters to perform fully atomistic classical MD simulations of *DszB* in the unbound state and bound to substrates (HBPS) and product (HBP) allowing us to gain insight into the dynamical differences resulting from the small substitutions between the bound molecules. Results from MD simulations address the role a given residue plays in ligand interaction, such as steric constraint of the substrate, electrostatic stabilization, substrate stacking interactions [44, 45]. Umbrella sampling is used to identify mechanistic relationships of the protein with the substrate that contribute to the unique conformational change upon binding (Figure 1.4), which we anticipate arise as a result of gate formation responsible for ligand ingress/egress. Moreover, we propose that interactions between gate residues with inhibitors are essential factors responsible for product inhibition.

1.3.2 Quantify product and competitive inhibition effects with absolute binding free energy (Chapter 5)

To determine the role that individual residues plays in binding, we are interested in quantifying overall active site contributions to specificity and inhibition by computing the absolute binding free energy. This approach addresses topological contributions to binding free energy and is directly comparable to measured K_I values. This approach has been proven useful in other enzyme systems

[46]. We anticipate more favorable binding free energy correlates with increased inhibition as observed experimentally (i.e., NAPO or BCA will have the lowest binding free energy, while NTAM and BIPH will have the highest binding free energy). In order to test the hypothesis, we compute binding free energy using free energy perturbation with the Hamiltonian replica-exchange protocol (FEP/ λ -REMD) in NAMD [47], as outlined by Jiang *et al.* [48]. The expected outcome is a ranking of the binding affinities of the selected potential inhibitors to *DszB*, thus quantifying molecular-level contributions to *DszB* inhibition. The binding affinity of *DszB* is also the thermodynamic contribution to substrate specificity, which combined with the mechanistic free energy from classical MD simulation, provides a complete description of substrate specificity.

Chapter 2 — Computational Methods

2.1 Introduction

We performed molecular dynamic simulations and free energy calculations to identify the roles individual residues play in determining how *DszB* interacts with different bound ligands and the mechanisms behind substrate specificity; these types of calculations provide dynamic and thermodynamic information regarding protein-ligand complex systems. Descriptions for each method used are provided below.

2.2 Molecular Dynamics simulation

Molecular Dynamics (MD) simulations is a technique used to compute the equilibrium and transport properties of a classical many-body system; therefore, MD simulation can be used to study the time-dependent dynamical behavior of biomolecules [49-51]. First, one must construct a model system consisting of N particles. Newton's equation of motion is applied to this system until the bulk properties of the system no longer change with time (approximating equilibration). After equilibration, data collection occurs. To measure an observable quantity in an MD simulation, the particles in the system should be able to express this observable as a function of both positions and momenta.

2.2.1 Initialization of MD simulation

To start the simulation, all particles in the system are assigned initial Cartesian coordinates and velocities in such a way that the positions do not overlap with the atomic or molecular cores. Subsequently, the particle velocities are shifted so that the total momentum is zero and resulting velocities adjust the mean kinetic energy

to the desired value. In thermal equilibrium, the following relation (Eq.2.1), should hold,

$$\langle V_{\alpha}^2 \rangle = k_B T / m \quad 2.1$$

where V_{α} is the α component of the given particle velocity, k_B is the Boltzmann constant, T is temperature, and m is mass of the particle. This equation can be used to define the instantaneous temperature at time t , Eq. 2.2,

$$k_B T(t) = \sum_{i=1}^N \frac{m V_{\alpha,i}^2(t)}{N_f} \quad 2.2$$

where N_f equal to the system the numbers of degree of freedom, so that we can adjust the instantaneous temperature $T(t)$ to match the desired set point temperature by scaling all velocities with a factor $(\frac{T}{T(t)})^{1/2}$. To predict the particle's next position in time, the positions of all particle's present positions (x) and previous positions (x_m) is combined with the force acting on the particles to approximate position for a particle, Eq. 2.3,

$$x_m(i) = x(i) - v(i)dt \quad 2.3$$

where $v(i)$ is the particle i present velocity.

2.2.2 The force calculation

Force calculation is the most time-consuming part of an MD simulation, because all possible force contributions from all neighbors have to be considered for a given particle i . For instance, if one only considers the interaction between a particle and the nearest neighbor particle, for a system of N particles, $N \times (N - 1)/2$ pair distances have to be evaluated, which means the time needed for evaluation scales as N^2 . If we use the Cartesian coordinates and compute the current distance in the x ,

y, and z direction between each pair of particles, these distances are indicated as x_r . Therefore, for a given pair of particles close enough to interact, the force $f_x(r)$ between these particles are calculated using an interatomic potential energy function, U , for N atoms in a system as shown in Eq. 2.4 [52],

$$f_x(r) = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} = -\nabla_{\vec{r}_i} U(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N), \quad i = 1, 2, 3, \dots, N \quad 2.4$$

Where m_i is the mass of atom i , \vec{a}_i represents the acceleration of atom i , and \vec{r}_i represents the position.

2.2.3 Equation of motions

After forces are computed, we can integrate Newton's equations of motion to predict the next position. In MD simulation, there are several different algorithms to do this, but the Velocity Verlet is the simplest and the most widely used.

To determine the next position, we start with a Taylor expansion of the coordinate of a particle around time t ,

$$r(t + \Delta t) = r(t) + V(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\ddot{r} + O(\Delta t^4) \quad 2.5$$

similarly,

$$r(t - \Delta t) = r(t) - V(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\ddot{r} + O(\Delta t^4) \quad 2.6$$

Summing these two equations,

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 \quad 2.7$$

The estimate of the next position contains an error of order Δt^4 , where Δt is the time step in MD simulation. Now that we have the new position, the old particle position at $(t - \Delta t)$ can be discarded; the current position becomes the old

positions, and the new positions become the current positions. After each time step, the current temperature, current potential energy calculated in the force loop, and the total energy, which should be conserved, could all be obtained. The resulting output of the simulation is a trajectory, which is a chronological list of atom positions and velocities as time progresses. There are several commercial software packages available to perform MD simulation, such as CHARMM [47], NAMD [53], AMBER [54], and GROMACS [55-59]. We used CHARMM and NAMD in this study but expect any of the other packages would yield similar results.

2.2.4 Energy function

The relationship between structure and energy is an essential element of many computational studies. The potential energy function, by custom also called a force field, is used to calculate the potential energy and its derivatives from the coordinates corresponding to the structure or conformation. Force fields are approximations of the exact potential energy; therefore, in Chapter 3, we will describe how we optimized and validated the force field parameters of the small molecules not contained in existing force field parameter sets.

The general form of the potential energy function most commonly used in CHARMM for macromolecules has two major components, including bonded and non-bonded contributions, that define the interatomic interactions shown below in Eq. 2.8 [60-62],

$$\begin{aligned}
U(\vec{r}) &= U_{\text{bonded}} + U_{\text{non-bonded}} \\
&= \left(\sum_{\text{bond}} k_b (b - b_0)^2 + \sum_{\text{angle}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedral}} k_\varphi (1 \right. \\
&\quad \left. + \cos(n\varphi - \delta)) \right. \\
&\quad \left. + \sum_{\text{improper}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_{UB} (S - S_0)^2 \right) \quad 2.8 \\
&\quad + \sum_{\text{non-bonded}} \frac{q_i q_j}{4\mu D r_{ij}} + \epsilon_{ij} \left[\left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right]
\end{aligned}$$

The bonded interactions include: bond stretching (b), angle bending (θ), dihedral angle rotation (φ) with a phase shift (δ), improper angle bending (ω), and an Urey-Bradley vibration term. The parameters k_b , k_θ , k_φ , k_ω , and k_{UB} are the respective force constants, and the variables with the subscript 0 are the respective equilibrium values. The non-bonded terms include van der Waals and electrostatic interactions. Among them, the van der Waals interaction is represented by 12-6 Lennard-Jones potential, which is used for the treatment of the core-core repulsion and the attractive van der Waals dispersion interaction. Where $\epsilon_{i,j}^{\text{min}}$ represents the depth of the potential well, $R_{i,j}^{\text{min}}$ represents the distance at which the Lennard-Jones potential reaches the minimum, and $r_{i,j}$ is the interatomic distance between atoms i , and j . The electrostatic contribution is defined based on Coulombic interaction, where q_i and q_j are the partial charges for atoms i , and j , and ϵ is the relative dielectric constant, ϵ_0 is the permittivity of vacuum. Non-bonded interactions are calculated

between all atom pairs within a user-defined cutoff distance, except for covalently bonded pairs.

2.2.5 Periodic boundary conditions

The behavior between finite systems compared to bulk systems is very different. In finite systems, the fraction of the surface atoms is much more significant, and the behavior would be dominated by surface effects. Periodic boundary conditions (PBC) can effectively overcome both finite and surface effect issues. It implies that particles are enclosed in a box which is infinitely replicated in all three Cartesian directions to completely fill space. If the provided potential range is not too long, the minimum image convention can let each atom interact with the nearest atom or image in the periodic array. This means that if one uses a potential with a finite range, interactions between two particles whose distance exceeds the defined cut-off value can be ignored. If an atom leaves the primary simulation box, attention can switch to the incoming image (Figure 2.1), so that the number of particles from the simulation region is always conserved. Therefore, the surface effects can be virtually eliminated, and the position of the box boundaries is not essential.

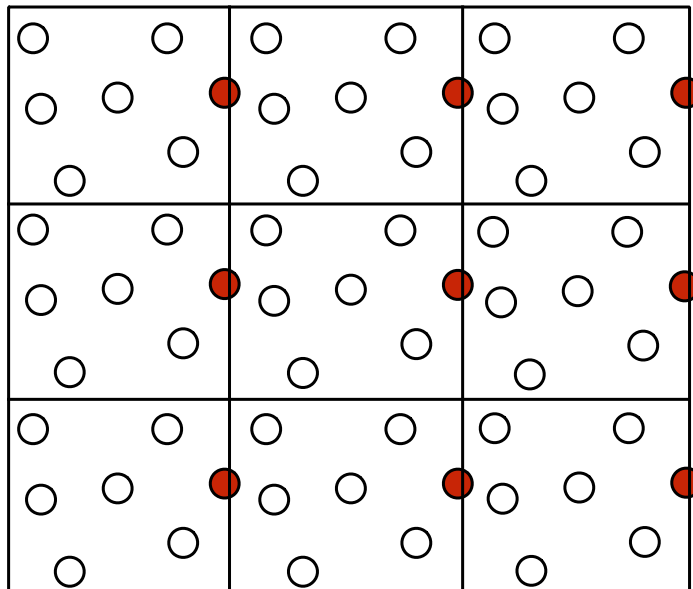


Figure 2.1 Illustration of periodic boundary conditions. When a particle moves out of the simulation box, an image particle moves in to replace it. Both real and image neighbors are involved in calculating the interactions within the cut-off distance.

2.3 Free energy calculation methods

To understand the chemical processes of protein-ligand binding and conformational change, it is necessary to examine the underlying free energy behavior. The determination of free energy changes using numerical simulation based on the fundamental statistical mechanics is applicable. We have implemented several free energy calculation methods including free energy perturbation, and umbrella sampling, which I will briefly describe below.

2.3.1 Basic approach to free energy calculation

The Helmholtz free energy, A , which is the thermodynamic potential of MD simulation canonical ensemble, can be expressed in terms of the partition function, Q , shown in Eq.2.9:

$$A = -\beta^{-1}\ln Q(N, V, T) \quad 2.9$$

where $\beta = (k_B T)^{-1}$, and Q is the partition function with arguments N , V , and T representing the number of degrees of freedom, volume of the system, and the absolute temperature, respectively. For most condensed phase applications, including those we consider here, ΔA and Gibbs free energy, ΔG , are numerically quite similar. Therefore, this equation provides a connection between thermodynamic and statistical mechanics, implying that calculating A is equivalent to calculating the value Q . Moreover, in both experimental and theoretical approaches, we are interested in estimating the free energy difference (ΔA) of two system states represented by 0 and 1, see Eq.2.10.

$$\Delta A = -\beta^{-1}\ln Q_1/Q_0 \quad 2.10$$

If masses are conserved for two systems, Eq. 2.10 can be rewritten as:

$$\Delta A = -\beta^{-1}\ln Z_1/Z_0 \quad 2.11$$

where Z represents the configurational integrals. Therefore, the focus on calculating the energy difference between two systems, ΔA , can be determined by the ratio of Q_1/Q_0 or equivalently Z_1/Z_0 , so Eq.2.11 can be transformed into Eq. 2.12:

$$\begin{aligned} \Delta A &= -\beta^{-1}\ln \frac{\int \exp[-\beta U_1(x)] dx}{\int \exp[-\beta U_0(x)] dx} \\ &= -\beta^{-1}\ln(\exp\{-\beta[U_1(x) - U_0(x)]\})P_0(x) \\ &= -\beta^{-1}\ln\langle \exp\{-\beta[U_1(x) - U_0(x)]\} \rangle_0 \end{aligned} \quad 2.12$$

where systems 0 and 1 can be described by their potential functions, $U_0(x)$ and $U_1(x)$, respectively, and P_0 is the probability density function of finding system 0 in the microstate defined by positions x of the particles.

$$P_0(x) = \frac{\exp[-\beta_0 U_0(x)]}{Z_0} \quad 2.13$$

According to Eq.2.12 and 2.13, the free energy difference of two systems, ΔA , can be estimated solely from simulation of system 0. Therefore, using one system as the reference and focusing on energy difference is the basic concept of the free energy perturbation method.

2.3.2 Free energy perturbation with replica exchange molecular dynamics (FEP/ λ -REMD)

Free energy perturbation with replica exchange molecular dynamics (FEP/ λ -REMD) is a powerful method designated to calculate free energy of solvation and binding of small molecules, which is directly comparable with experimentally measured binding affinity from Isothermal Titration Calorimetry (ITC). This method is implemented based on free energy perturbation calculations from MD simulations, and it developed through a step-by-step decomposition of the total reversible work [63-66]. It breaks the free energy calculation into several independent MD simulations. With REMD, the coordinates from each independent MD simulation generated from different Hamiltonians can be swapped to enhance the rate of the configuration exploration [67-73]. The distributed Replica (REPDSTR) implemented in CHARMM, by Hodoscek and co-workers [74, 75], allows for performing MD simulation on similar systems simultaneously in a high efficient parallel mode. According to the basic free energy calculation method (section 2.3.1), free energy difference of two systems 0 and 1 can be calculated using the FEP protocol using replica-exchange MD simulations with λ -swap moves, where λ

represents the number of sub-process along the two states. It has been shown that this method improved the Boltzmann sampling of kinetically trapped conformations [64, 76, 77].

Free energy calculation of this type involves two thermodynamic steps: ligand decoupling from protein-ligand complex in solution and ligand decoupling from a solvated ligand system. The difference of the two gives absolute ligand binding free energy (ΔG_b°) of the enzyme-ligand complex (Figure 2.2).

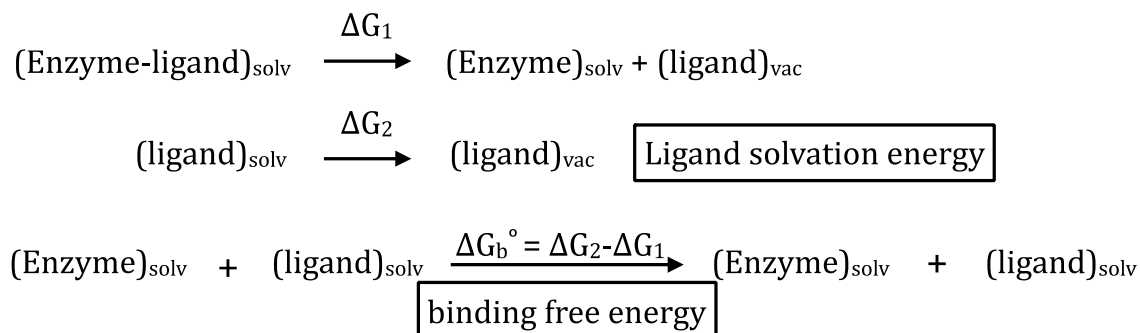


Figure 2.2 Thermodynamic cycle for calculating binding free energy ΔG_b° with FEP/ λ -REMD. “solv” and “vac” refer to solvated and vacuum, respectively.

The insertion of ligand in the binding pocket occurs over three steps with the help of three thermodynamic coupling parameters, repulsive, dispersive, and electrostatic represented by, λ_{rep} , λ_{dis} , and λ_{elec} , to control non-bonded interactions. An additional parameter restraint, λ_{rstr} , controls translational and orientational restraints. Therefore, the potential energy can be represented by four coupling parameters shown in Eq. 2.14

$$\begin{aligned}
 U(\lambda_{rep}, \lambda_{dis}, \lambda_{elec}, \lambda_{rstr}) \\
 = U_0 + U_{rep}(\lambda_{rep}) + U_{dis}(\lambda_{dis}) + U_{elec}(\lambda_{elec}) + U_{rstr}(\lambda_{rstr})
 \end{aligned} \tag{2.14}$$

where U_0 is the potential of the system without interacting with the ligand, U_{rep} and U_{disp} are the shifted Weeks-Chandler-Anderson (WCA) repulsive and dispersive components of the Lennard-Jones potential, U_{elec} is the electrostatic contribution, and U_{rstr} is the restraining potential that improves phase space sampling [76]. The repulsive contribution, ΔG_{rep} , can be represented by Eq.2.15

$$\begin{aligned}
 U(\lambda_{rep} = 0, \lambda_{dis} = 0, \lambda_{elec} = 0, \lambda_{rstr} = 1) \\
 \rightarrow U(\lambda_{rep} = 1, \lambda_{dis} = 0, \lambda_{elec} = 0, \lambda_{rstr} = 1)
 \end{aligned}
 \tag{2.15}$$

the dispersive contribution ΔG_{dis} corresponds to Eq. 2.16

$$\begin{aligned}
 U(\lambda_{rep} = 1, \lambda_{dis} = 0, \lambda_{elec} = 0, \lambda_{rstr} = 1) \\
 \rightarrow U(\lambda_{rep} = 1, \lambda_{dis} = 1, \lambda_{elec} = 0, \lambda_{rstr} = 1)
 \end{aligned}
 \tag{2.16}$$

the electrostatic contribution ΔG_{elec} corresponds to Eq.2.17

$$\begin{aligned}
 U(\lambda_{rep} = 1, \lambda_{dis} = 1, \lambda_{elec} = 0, \lambda_{rstr} = 1) \\
 \rightarrow U(\lambda_{rep} = 1, \lambda_{dis} = 1, \lambda_{elec} = 1, \lambda_{rstr} = 1)
 \end{aligned}
 \tag{2.17}$$

and restraint contribution ΔG_{rstr} can be represented by Eq.2.18

$$\begin{aligned}
 U(\lambda_{rep} = 1, \lambda_{dis} = 1, \lambda_{elec} = 1, \lambda_{rstr} = 1) \\
 \rightarrow U(\lambda_{rep} = 1, \lambda_{dis} = 1, \lambda_{elec} = 1, \lambda_{rstr} = 0)
 \end{aligned}
 \tag{2.18}$$

Each λ -staging FEP window is treated as a replica, and according to Eq. 2.15-2.17, free energy calculation is separated into four different types corresponding to repulsive, dispersive, electrostatic, and restraint. The replica exchange algorithm follows the conventional Metropolis Monte Carlo exchange probability with λ -swap moves as shown in Eq.2.19.

$$\begin{aligned}
P(\lambda_{type}^i \leftrightarrow \lambda_{type}^j) & \\
&= \min \left\{ 1, e^{-[U(\lambda_{type}^i x^i) + U(\lambda_{type}^j x^j) - U(\lambda_{type}^i x^j) - U(\lambda_{type}^j x^i)] / k_B T} \right\}
\end{aligned}
\tag{2.19}$$

where U is the total potential energy of the replica, and $\lambda_{type}^i, \lambda_{type}^j$ represent the staging parameters. Figure 2.3 shows the how these replica pairs interact during the exchange mode.

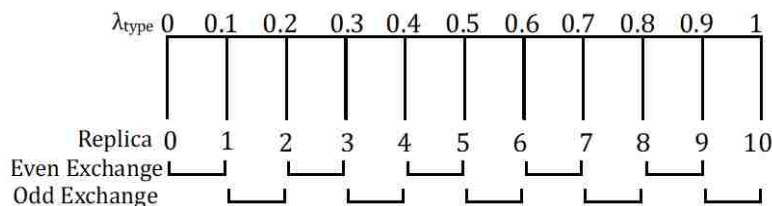


Figure 2.3 Schematic representation of parallel tempering simulation of FEP/ λ -REMD protocol. λ_{type} is the staging parameter. Each replica represents a single MD simulation with its own input and output under the specific staging parameter. The exchange between 0 and 1, 2 and 3, ..., and 8 and 9 are denoted as possible Hamiltonians exchanged between the even number of MD simulation runs. Similarly, odd exchange represents the possible Hamiltonians exchanged between odd MD simulation runs.

Finally, we use multistate Bennett acceptance ratio (MBAR) to calculate the free energies and the statistical uncertainty from each repulsive, dispersive, electrostatic, and restraining energies contributions [75]. The ligand solvation energy was calculated in the similar manner except without the restraining contribution. We implemented FEP/ λ -REMD in NAMD to complete the calculations for ligand binding free energy ΔG_b° [47, 53].

2.3.3 Umbrella sampling

Chemical potential, and accordingly free energy difference is the driving force for equilibrium in chemical processes. Umbrella sampling provides the free energy along a reaction coordinate in a chemical process [78, 79]. In general, a biased potential is applied to the system to enhance the sampling along the whole reaction coordinate from one system state to another. This can apply to one or different simulation windows, the distribution of which should overlap. This biased potential, represented by symbol w_i of window i , is an additional energy term shown in Eq. 2.20 [80],

$$U^b(\mathbf{r}) = U^u(\mathbf{r}) + w_i(\xi) \quad 2.20$$

where the superscripts b and u represent the biased and unbiased quantities, respectively, and the symbol ξ is identified as geometric grounds, such as distance, torsion, or root mean square deviation (RMSD), from the two reference states. Often the biased potential, as it is in our study, is a simple harmonic bias, where $w_i(\xi) = \frac{K}{2}(\xi - \xi_0^{ref})^2$, and K is the spring constant. The MD simulation of the biased system provides the biased density function along the reaction coordinate, P_i^b in Eq. 2.21

$$P_i^b(\xi) = \frac{\int \exp\{-\beta[U(\mathbf{r}) + w_i(\xi'(\mathbf{r}))]\} \delta[\xi'(\mathbf{r}) - \xi] d^N \mathbf{r}}{\int \exp\{-\beta[U(\mathbf{r}) + w_i(\xi'(\mathbf{r}))]\} d^N \mathbf{r}} \quad 2.21$$

where β equals $1/(K_B T)$, K_B is the Boltzmann constant, and T is the absolute temperature. Similarly, the unbiased probability function P_i^u , can be represented by Eq. 2.22.

$$P_i^u(\xi) = \frac{\int \exp\{-\beta[U(\mathbf{r})]\} \delta[\xi'(\mathbf{r}) - \xi] d^N \mathbf{r}}{\int \exp[-\beta U(\mathbf{r})] d^N \mathbf{r}} \quad 2.22$$

According to Eq. 2.21 and 2.22, the unbiased potential can be directly retrieved from the biased potential, see Eq. 2.23.

$$\begin{aligned}
 P_i^u(\xi) &= P_i^b(\xi) \exp[\beta w_i(\xi)] \times \frac{\int \exp\{-\beta[U(r) + w_i(\xi'(r))]\} d^N r}{\int \exp\{-\beta U(r)\} d^N r} \\
 &= P_i^b(\xi) \exp[\beta w_i(\xi)] \langle \exp[-\beta w_i(\xi)] \rangle
 \end{aligned}
 \tag{2.23}$$

Therefore, the free energy, $A_i(\xi)$, between two reference states can be evaluated directly from $P_i^b(\xi)$,

$$A_i(\xi) = -\beta^{-1} \ln P_i^b(\xi) - w_i(\xi) + F_i
 \tag{2.24}$$

where $F_i = -\beta^{-1} \ln \langle \exp[-\beta w_i(\xi)] \rangle$, which depends only on ξ . In general, if the sampling distribution simultaneously covers the regions of configuration space relevant to the physical system, Eq. 2.24 is sufficient to unbiased the simulation and obtain the free energy difference along the selected reaction coordinate.

Here, we used weighted histogram analysis method (WHAM) to analyze the umbrella sampling MD data to determine a potential of mean force (PMF) [81, 82]. This method is a numerical method to calculate F_i from Eq.2.24 and aiming to minimize the statistical error of $P^U(\xi)$ to evaluate the unbiased PMF by removing the restraint biasing via a self-consistent iteration method (Eq. 2.25-2.27) [83-85]. The global distributions of each individual window can be represented by Eq.2.25.

$$P^u(\xi) = \sum_i^{window} p_i(\xi) P_i^U(\xi)
 \tag{2.25}$$

The weights, p_i , are chosen based on minimizing statistical error of P^U , where

$$\frac{\partial \sigma^2(P_i^U)}{\partial p_i} = 0 \text{ and } \sum p_i = 1. \text{ Eq. 2.26 denotes the } p_i \text{ representation,}$$

$$p_i = \frac{a_i}{\sum_j a_j}, a_i(\xi) = N_i \exp[-\beta w_i(\xi) + \beta F_i] \quad 2.26$$

with N_i equal to the total number of steps sampled among window i . The $\exp(-\beta F_i)$ can be calculated through Eq. 2.27.

$$\exp(-\beta F_i) = \int P^U(\xi) \exp[-\beta w_i(\xi)] d\xi \quad 2.27$$

Since P^U can be represented by F_i through Eq.2.27, therefore, combined with Eq. 2.25, which complete the iteration process.

Chapter 3 — CHARMM force field parameters for 2'-hydroxybiphenyl-2-sulfinate, 2-hydroxybiphenyl, and related analogs

This chapter has been reprinted from Yu, *et al.* [35], Copyright © 2017 Elsevier with permission. The author of this dissertation performed the force field parameterization for the selected small molecules that are necessary for the MD simulations. This chapter will describe how to optimize the CHARMM force field parameters including charges, bond distances, angles, and dihedrals using the Force Field Toolkit (ffTK) in VMD. The optimized parameters were tested to reproduce Infrared spectra (IR) and compared with experimental results for validation. The experimental results used for validation of the force fields were obtained by Ishan Fursule and Landon C. Mills, University of Kentucky. Chapter 3 was a collaborative effort, applying both experimental and computational approaches, aimed at finding optimized force field parameters for the potential inhibitors of *DszB*.

3.1 Summary

DszB catalyzes the cleavage of the carbon-sulfur bond from HBPS in the final step of microbial 4S pathway desulfurization reactions. *DszB* is notable for its substrate specificity and exhibits product inhibition, both of which hinder the overall 4S pathway turnover rate. To understand the molecular-level contributions to substrate and inhibitor binding to *DszB*, we plan to perform molecular dynamic simulations bound to an array of naphthenic molecules and biphenyl analogues of HBPS. However, many of the small molecules we are interested in are not included in standard force field packages, and thus, we must first produce accurate molecular mechanics force fields. Here, we develop and validate CHARMM-compatible force

field parameters for the HBPS substrate, the 2-hydroxybiphenyl product, and potential inhibitors including: 2,2'-biphenol, 2-biphenyl carboxylic acid, 1,8-naphthosultam, and 1,8-naphthosultone. The selected molecules represent biphenyl compounds having both a single and double functional group and the planar naphthenic molecule class, all likely present in the oil-rich environment surrounding *DszB*-producing microorganisms. The Force Field Toolkit (ffTK) in VMD was used to optimize charge, bond distance, angle, and dihedral parameters. Optimized geometries were determined from quantum mechanical calculations. Molecular simulations of the molecules in explicit and implicit water solutions were conducted to assess the abilities of optimized parameters to recapitulate optimized geometries. Calculated infrared (IR) spectra were obtained and compared with experimental IR spectra for validation of the optimized MM parameters.

3.2 Introduction

Currently, a majority of the world's transportation fuels are derived from sulfur-containing crude oil. The concentration of sulfur largely depends on the region of origin. Sulfur remaining in finished fuels, post-refining, is converted to sulfur oxides (SO_x) upon combustion, which are then emitted as a toxic pollutant. Accordingly, sulfur concentration in refined petroleum products is strictly regulated at extremely low levels to minimize SO_x emissions [4]. Hydrodesulfurization (HDS) is the traditional industrial approach to liberating the majority of sulfur containing compounds from crude oil; this process occurs in the presence of metallic catalysts and hydrogen gas at high pressures and temperatures. Unfortunately, recalcitrant sulfur heterocycles, such as dibenzothiophene (DBT), often remain intact after HDS

treatment [6-9, 11, 86]. Biodesulfurization is an enzymatic process by which organosulfur compounds recalcitrant to HDS can be freed of sulfur at ambient conditions. Naturally, this is a promising complementary process for implementation downstream of traditional HDS units [6, 11, 12]. However, improving the overall rate of enzymes that desulfurize thiophenic molecules remains a primary technical challenge preventing economical implementation of biodesulfurization [16, 17].

The most suitable biodesulfurization method for industrial application is the 4S pathway (Figure 1.1), two monooxygenases (*DszA* and *DszC*), a desulfinase (*DszB*), and an oxidoreductase (*DszD*) convert DBT to 2-hydroxybiphenyl (HBP). The HBP product is reincorporated in the oil fraction, and sulfite is removed in the aqueous fraction. This final catalytic step has been widely identified as rate-limiting, making *DszB* the most logical target for activity improvements [12, 20]. Product inhibition also affects *DszB* conversion rates, wherein increasing concentrations of product, and possibly HBP derivatives, inhibit turnover [21, 22]. Despite this knowledge, we have relatively little mechanistic understanding of *DszB* activity, substrate binding, or product inhibition, making rational protein design efforts challenging.

We anticipate molecular simulation is capable of addressing many questions underlying poor substrate turnover and product inhibition. Prior to that, however, we must first develop force field parameters for a subset of chemically relevant aromatic molecules with which to complete these studies. The molecules we parameterize (Figure 1.2) were selected based on Watkin *et al.* [20] experimental study, where author established *DszB* rate and inhibition constant for several

functionalized biphenyl compounds and planar naphthenic molecules that exist in petroleum. In addition to the HBPS substrate and the HBP product, we examine 2,2'-biphenol (BIPH), 2-biphenyl carboxylic acid (BCA), 1,8-naphthosultam (NTAM), and 1,8-naphthosultone (NAPO). These molecules represent a variety of functional groups that are anticipated to mediate different molecular-level interactions. With accurate force-field parameters for these molecules, we will be capable of developing molecular models to describe substrate binding and inhibition phenomena in *DszB*.

The CHARMM all-atom additive force field is a widely accepted parameter set for biomolecular simulation [47, 87-89]; it is also easily expanded upon given the availability of the compatible CHARMM General Force Field (CGenFF) parameter set for organic, “drug-like” molecules and force field fitting tools [89]. Accordingly, we build upon the CHARMM parameter set, ideal for modeling *DszB*, by using CGenFF nomenclature and parameter analogies as a starting point for biphenyl and planar naphthenic analog parameter generation. As a result, all the conventions and recommendations for usage of the biomolecular CHARMM force field and CGenFF apply to the newly generated parameter set.

Both CHARMM and CGenFF use the Class I potential energy function, which is the sum of bonded and non-bonded contributions Eq. 2.8. Here, we report parameters for the six molecules shown in Figure 1.2 that are compatible with CHARMM and CGenFF potential energy functions. Initial parameters were obtained by analogy with molecules available in CGenFF with force field version 3.0.1. Using Visual Molecular Dynamics (VMD) version 1.9.2 and the Force Field Tool Kit (ffTK)

plugin [90, 91], we have optimized the missing CGenFF parameters, iterating until molecular mechanics (MM) simulations were capable of reproducing quantum mechanical (QM)-derived target data. We have optimized bond, angle, and dihedral parameters for each of the molecules, ffTK does not provide support for optimization of improper dihedrals or the Urey-Bradley 1,3-term [91], but for aromatic compounds, these terms are not likely to significantly contribute to the potential energy given their relatively stiff torsion potential and cyclic constraints. The optimized parameters were validated by comparing calculated MM geometries to the optimized molecular geometries and through recapitulation of experimental infrared (IR) spectra.

3.3 Methods

3.3.1 Parameterization

The parameterization process used in this study follows the general ffTK procedure outlined by Mayne *et al.* [91]. We will discuss each step, including the file types generated and use of external software and web servers, in the sections that follow. The inherently iterative force field optimization procedure is illustrated as a flow chart in Figure 3.1 below. Briefly, the procedure entails (1) developing an initial parameter set based on analogy with existing CGenFF parameters, (2) optimizing the geometry of the molecule, (3) optimizing partial atomic charges against QM-derived water interaction profiles, (4) optimizing bonds and angles against the QM-derived Hessian matrix, and (5) optimizing the dihedrals against QM-derived potential energy surfaces (PES).

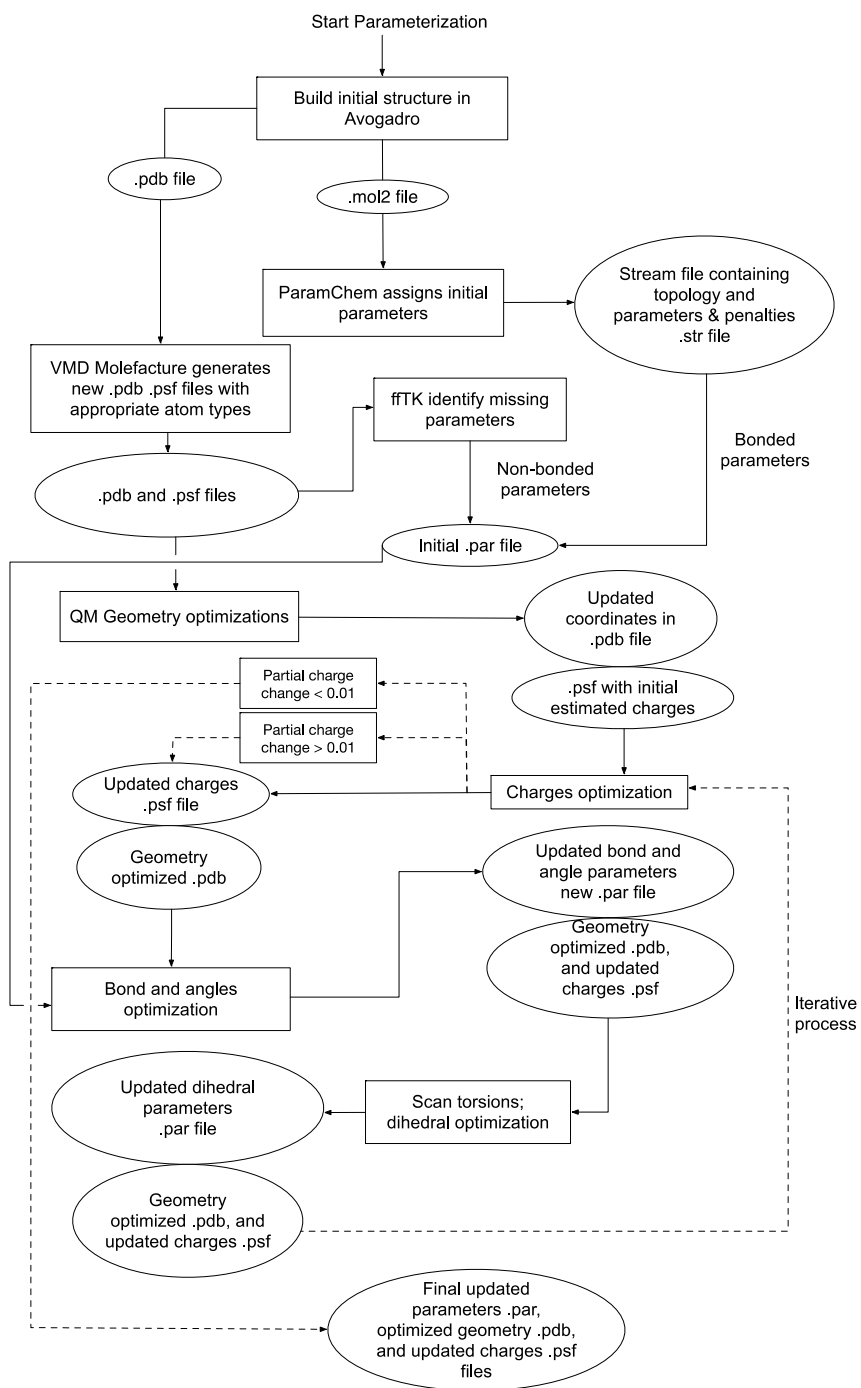


Figure 3.1 Flow chart illustrating the ffTK parameterization process used to obtain MM force fields for the selected molecule shows in Figure 1.2. Rectangular boxes represent a step in the workflow, Ovals are used to describe the set of files generated from the preceding step. Text below or next to the lines between steps

provides additional contextual information. Dashed lined represent the iterative decision point in the workflow, where convergence of the partial charges results in a final set of MM parameters.

This procedure is repeated, returning the charge optimization step, as many times as necessary to obtain convergence. We defined our convergence here to the tolerance as partial atomic charge differences of less than 0.01 between the previous parameter set and the current set. As mentioned above, we have not optimized the Urey-Bradley term or the improper dihedral term, as ffTK does not currently support optimization of these terms. Improper dihedral terms are not employed in the selected molecules with the exception of BCA (C13-C1-O2-O1), the parameter of which are well represented by initial estimates. Additionally, chemical intuition suggests the cyclic and planar nature of the selected molecules will result in only minor contributions from these terms to the bonded potential energy term. Non-bonded parameters were obtained from the CGenFF parameters set and were not optimized, per standard protocol.

3.3.1.1 Initialize parameters by analogy with CGenFF

The initial molecular structure was generated using the Avogadro molecular editor (V.1.1.1) [92], resulting in .mol2 and .pdb formatted files containing coordinates and bond orders, in the case of the former. The .mol2 formatted file was then used as input to the ParamChem server (<https://cgenff.paramchem.org>), which assigned atom types, charges, and parameters by analogy to existing molecules in the CGenFF parameter set [89, 93, 94]. From ParamChem, we obtained a stream file (.str) containing the CHARMM/CGenFF-compatible topology and bonded

parameters for the input molecule. We have provided the initial .str files obtained from ParamChem in .str format files. ParamChem assigned a “penalty value” to each generated parameter, where parameters with penalties should be validated.

Starting from the .pdb from Avogadro, VMD Molefactory was used to assign CGenFF atom types in an updated .pdb file. The initial .str file from ParamChem was used, in Molefactory, to initialize charges and generate .psf file that required for ffTK optimization. Finally, non-bonded parameters were assigned by analogy with CGenFF using ffTK and were combined with the initial bonded parameters from the .str file in a .par file. The .par file, with both non-bonded and bonded parameters, is used in bond, angle, and dihedral optimization and is updated with new parameters following each step.

3.3.1.2 Geometry optimization

The optimized geometry for each molecule in Figure 1.2 was determined from quantum mechanical calculation using Gaussian09 [95]. The geometry optimization started with a moderate-size basis set, MP2/6-31G*, with polarization functions applied to only heavy atoms; this initial optimization served as an input for a more extensive optimization using the MP2/6-31++G** basis set, adding polarization functions and diffuse functions to all atoms. This dual minimization approach was necessary to reach a global minimum in the case of sulfur-containing molecules. For consistency, the approach was applied in all geometry optimizations.

3.3.1.3 Charge optimization

Partial atomic charges were optimized in ffTK on the basis of calculated water-interaction profiles, consistent with existing CHARMM force field parameters [87,

89]. For each hydrogen bond donor or acceptor atom, a water molecule was inserted within hydrogen bonding distance. The ffTK algorithm optimized the placement of water molecules such that they were automatically oriented for hydrogen bonding and steric repulsion between water molecules was minimized [91]. The generated water-interaction Gaussian input files were visually inspected prior to QM calculation to prevent over constraining orientation of the water molecules. In cases where molecules contained aromatic rings (i.e., where aromatic carbons are *sp*² hybridized), aromatic carbons were assigned as both hydrogen bond donors and acceptors. The distance between water molecules and the target atoms and the rotation angle of water molecules were optimized quantum mechanically at the HF/6-31G* level of theory. HF/6-31G* was used in this step to maintain consistency with CHARMM force field [91].

The output from QM calculations was imported into ffTK for MM force field fitting. For every water-atom interaction, a MM interaction energy as a function of distance was calculated using the provided parameters. Partial atomic charges were varied as the deviation between the QM and MM minimum distances between the atoms, interaction energies, and dipole moments were minimized [91]. Constraints for the upper and lower charge boundaries were manually adjusted according to the assigned ParamChem penalties, where smaller penalties were more strictly constrained. In each charge optimization step, the default ffTK advanced setting parameters were applied for all molecules except for HBPS, which is a negatively charged molecule. For HBPS, the interaction energy scaling factor was manually changed from 1.16 to 1 better approximate the bulk phase [89]. The convergence

tolerance of each calculation was set to 0.005. An updated .psf file was generated at this point.

3.3.1.4 Bond and angle optimization

When available, experimental vibrational spectra should be used for bond and angle parameter optimization. However, this information was not available for any of the compounds in Figure 1.2; thus, bond and angle parameters were optimized against the QM Hessian matrix calculated at the MP2/6-31G* level of theory. From this matrix of second derivatives of energy with respect to coordinates, a QM PES was determined for comparison with the MM-derived PES, using the input parameter set. The equilibrium bond and angle values and the force constants, k_b and k_θ , changed as the deviations in the QM and MM energies were minimized. A new .par file, with optimized bond and angle parameters, was generated following optimization.

3.3.1.5 Dihedral optimization

Dihedral parameters were the last to be optimized, as they critically depend on charges and bond and angle parameters. Only dihedrals with ParamChem penalties were optimized (Table 3.1). Planar naphthenic ring structures were not well represented among existing CGenFF molecules; therefore, many more dihedrals were optimized in these two molecules than the other four. A MM PES surface was calculated using the updated parameter set, and dihedral parameters were optimized against a QM calculated PES, allowing the force constant, K_ϕ , of each dihedral to change. The periodicity, n , and phase shift angle, δ , were set according to initial parameters from ParamChem and were not varied in the optimization [91].

The QM PES were generated in Gaussian09. Each dihedral angle was scanned bidirectionally from equilibrium value in 5° increments at the MP2/6-31G* level of theory. The resulting QM log files, containing data from the equilibrium value to the final scanned dihedral, were used as input to ffTK for the optimization.

Table 3.1 Dihedrals of each molecule included in the optimization process. The dihedrals were scanned 90° in either direction about the equilibrium value in 5° increments with the exception of HBPS; for HBPS the dihedral angles marked with an asterisk were scanned 40° about the equilibrium value. Atom numbers correspond to the convention illustrated in Figure 3.2

Scan Torsions			
HBP	HBPS	BCA	BIPH
O C1 C2 C3 O C1 C2 C7 C2 C1 O H	O1 C8 C7 C12 O1 C8 C7 C2 S C3 C2 C1* S C3 C2 C7* C7 C8 O1 H* C2 C3 S O2*	O2 C13 C1 C2 O1 C13 C1 C2 C13 C1 C2 C3 C13 C1 C2 C7	O1 C8 C7 C12 O1 C8 C7 C2 C7 C8 O1 H
NTAM		NAPO	
C1 C2 C3 S C2 C3 C4 C5 C3 C4 C5 N C6 C5 N S C2 C3 S N C4 C3 S N C5 N S C3 H N S C3	H C2 C3 S C9 C4 C3 S C5 C4 C3 S C4 C5 N S C2 C3 S O2 C4 C3 S O2 C5 N S O2 H N S O2	C3 C2 C1 S C2 C1 C10 C9 C1 C10 C9 O3 C8 C9 O3 S C2 C1 S O2 C10 C1 S O2 C9 O3 S C1	H C2 C1 S C5 C10 C1 S C9 C10 C1 S C10 C9 O3 S C2 C1 S O3 C10 C1 S O3 C9 O3 S O2

The QM and MM PES results were visually inspected using a VMD plotting utility to determine if further refinement was necessary based on the Root Mean Square Error (RMSE) value and general fit of the curves. All optimizations started with the simulated annealing protocol [96]. Once the general shape of the MM PES matched that of the QM PES, the optimization method was switched to the downhill

simplex method and the PES cutoff was decreased from 10 kcal/mol to 5 kcal/mol, further enhancing convergence [97, 98]. The optimization of the dihedral parameters was considered converged when the RMSE was less than 0.5 and the minimum energy difference of the QM and MM PES basins was less than 0.2 kcal/mol. After the optimization converged, the .par parameter file was again updated.

All the parameters including partial charges, bonded, and nonbonded parameters influence MM behavior. If the optimization is incomplete, changing any bonded parameter will influence partial charges and vice versa. Thus, the parameterization process is inherently iterative. Following dihedral optimization, we returned to the charge optimization step (Figure 3.1), using the new .par file to optimize the partial charges. When the partial charges did not change by more than 0.01, the parameterization process was considered complete. At completion, the parameterization process yielded a .par file with the optimized parameters, a .pdb file with the optimized geometry, and a .psf file with the molecular topology and charges.

3.3.2 Validation of optimized parameters

3.3.2.1 Explicit solvent molecular dynamic simulations

Molecular dynamics simulations of the explicitly-solvated molecules were conducted using the optimized parameter sets for comparison to the optimized geometries. Each molecule was solvated in a cubic periodic cell of TIP3P water using the Multiscale Modeling Tools for Structural Biology (MMTSB) Tool Set [99]; the buffer between the molecule and the edge of the periodic cell was 13 Å. For HBPS, a

charged molecule, one sodium ion was included to neutralize the overall system charges for application of long-range electrostatic methods. In CHARMM, the system was minimized for 2000 steps using the steepest descent algorithm followed by 5000 steps using the Adopted-basis Newton-Raphson algorithm. The minimized system was then heated from 90 K to 300 K in 25 K increments. This was followed by isothermal-isobaric (NPT) simulation for 100 ps at 300 K and 1.01325 bar to equilibrate the system density. The density-equilibrated systems were then simulated for 100 ns at 300 K in the NVT ensemble in NAMD.

3.3.2.2 Preparation of HBPS and experimental IR spectra

Except for HBPS, IR spectra for all molecules in Figure 1.2 are available in the literature. The experimental HBP, BIPH, and BCA spectra were obtained from the NIST Chemistry WebBook [100]. NTAM and NAPO spectra were from BIORAD (BIORAD: Bio-Rad/Sadtler IR Data Collection obtained from Bio-Rad Laboratories, Philadelphia, PA (US)). HBPS IR spectrum was experimentally determined. HBPS was synthesized by suspending 20 mM biphenosultine in 44 mM NaOH at room temperature overnight on a Fisher Scientific Reliable Tube Rotator. The biphenosultine was a gift from the Texas A&M University LINCHPIN Laboratory. The resulting aqueous HBPS was lyophilized at -35°C and 60 mT in a Benchtop Freezedryer (SP Scientific VirTis Advantage Plus XL 85). Lyophilized HBPS was analyzed via Attenuated Total Reflectance (ATR) FTIR using a Varian Inc. 7000e Spectrometer. The lyophilized sample was placed on the diamond ATR crystal, and the spectrum was collected for wavelengths between 700 and 4000 cm^{-1} and averaged over 32 scans.

3.3.2.3 Calculated IR spectra

The optimized parameters were validated by comparing IR spectra calculated from simulation with experimental spectra. The IR Spectra Density Calculator in VMD was used to compute spectra from simulation. Implicit solvent MD simulations were used as input to the IR calculations, as to reduce statistical error that may arise as a result of incomplete sampling in the presence of explicit solvent over relatively short simulations [101]. As with explicit solvent MD, the implicit solvent MD simulations were conducted in NAMD for each molecule. The Generalized Born implicit solvent (GBIS) model was used to obtain 20 ps of data following a 2000 steps minimization. The hydrogen bonds were not constrained to obtain vibrational frequency data. The simulation time step was 1 fs and the Born radius cutoff was set at 13 Å.

The VMD IR spectra density calculator computes the spectra from simulation trajectories by determining the dipole moment of all atoms in the molecule and calculating the auto-correlated Fourier transform. The default parameters in the IR Spectra Density Calculator were used with the exception of the time step and the maximum frequency, which were set to 1 fs and 4000 cm^{-1} , respectively.

3.4 Results and discussions

3.4.1 Geometry optimization

The optimized geometries of each of the six molecules is shown in Figure 3.2. The molecules can be divided into two groups: biphenyl derivatives (HBPS, HBP, BCA, and BIPH) and naphthalene derivatives (NTAM and NAPO). The aromatic rings of unsubstituted biphenyl in the solution and vapor phase are not coplanar and

exhibit a characteristic twist about the central dihedral (e.g., dihedral C3-C2-C7-C8 of Figure 3.2), which experiment suggests a $44.4(1.2)^\circ$ [102]. Substitution, particularly at the *ortho*- position, introduces steric effects that lead to an increase in this angle. As expected, QM geometry optimization gives a dihedral angle of 53.5° , 59.2° and 59.4° for HBP, HBPS, and BCA, respectively. For HBP, experimental X-ray data suggests the value should be $54.3(8)^\circ$ [103]. Unfortunately, experimental structures of HBPS and BCA are not available for comparison but the proximity to the dihedral angle of *ortho*-substituted 2-hydroxybiphenyl suggests the calculated values are reasonable. The larger functional groups of HBPS and BCA introduce more steric effects on the biphenyl rings slightly increasing the dihedral angle this is also observed in the BIPH molecule. BIPH is a symmetric molecule, with hydroxyl groups in the *trans* position in the hydrated form. The X-ray structure of the 2,2'-biphenol monohydrate reveals the dihedral angle to be $67.6(1)^\circ$ [104]. QM geometry optimization at the MP2/6-31++G** level of theory resulted in a dihedral angle of 70.7° , in relative agreement with experiment.

As with several biphenyl derivatives, structural information of the two naphthalene derivatives is not available in the literature. However, the base naphthalene structure has been shown to be planar, and thus, we expect that calculations will yield relatively planar geometry overall [105]. Some strain in the aromatic rings, as a result of the *peri*-interactions, is expected to manifest in slight variations in bond lengths compared to unsubstituted naphthalene [106]. We did, in fact, observe this phenomenon in our geometry optimization where bond distance between C3 and C5 atoms were slightly shortened by 0.17 \AA compared to bond

distance between C8 and C10 in NTAM; similarly the bond distance between C1 and C9 was slightly shortened by 0.19 Å compared to C4 and C6 in NAPO. Overall, the naphthenic molecules maintained planar geometries, with slight protrusion of the *peri*-substitutions from the plane. The optimized geometry file (in .pdb format) have been provided in .pdb formatted files.

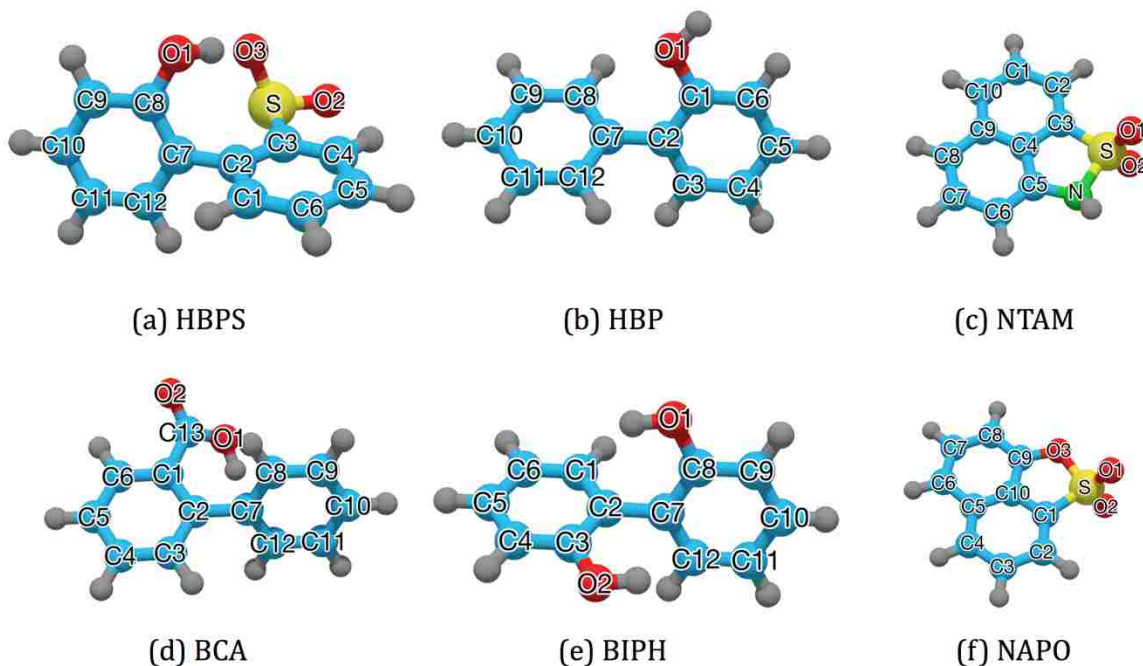


Figure 3.2 Optimized geometry for each molecule obtained using Gaussian09 with the MP2/6-31++G** level of theory. The atom labeling convention used in the parameterization process for each molecule is shown in black letter over the atoms. Note, the atom names/number do not necessary correspond to IUPAC naming conventions.

3.4.2 Optimized parameters

Charge, bond, angle, and dihedral optimization, conducted as described above, resulted in a unique set of parameters for each molecule. The optimized charges (Table 3.2) and bond and angle parameters (Table 3.3) were manually inspected to

ensure optimization did not result in values or force constants enforcing physically unrealistic conformations. The dihedral parameters (Table 3.4) were visually inspected as described above. For the optimized dihedrals, i.e., those with unreasonable penalties assigned by ParamChem, the MM-derived PES was fit to the QM PES (Figure 3.3). Good agreement with QM was achieved in all six cases. The MM-energy minima positional aligned with the QM PES minima and were within 0.2 kcal/mol. In the sections that follow, we will describe how these parameters correspond to optimized geometries and the ability of the force constants to maintain these bond distances and angles.

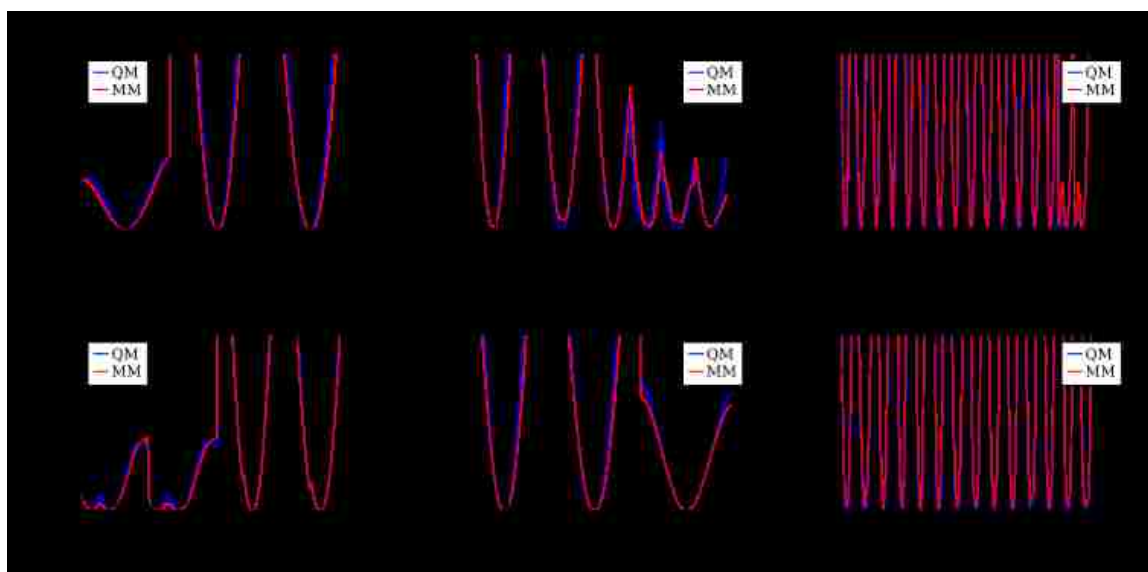


Figure 3.3 Torsional profiles of HBP, HBPS, NTAM, BCA, BIPH, and NAPO, as labeled at the top of each plot. The QM PES (blue) and the MM PES (red) are shown for each of the scanned dihedrals (Table 3.1). As aromatic rings are relatively well described by existing molecules in CGenFF, the dihedrals scanned largely represent those of the functional groups, the connections of the functional group with the aromatic ring, and bridging aromatic carbons.

Table 3.2 Optimized partial atomic charge parameters for each molecule. Charges have been optimized according to the described procedure.

HBP			HBPS			NTAM		
Atoms	Atom type	Charges	Atoms	Atom type	Charges	Atoms	Atom type	Charges
			C1	CG2R61	-0.627	C1	CG261	-0.278
			C2	CG2R67	-0.0123	C2	CG2R61	-0.102
			C3	CG2R61	0.321	C3	CG2RC0	0.176
C1	CG2R61	0.130	C7	CG2R67	0.495	C4	CG2RC0	0.281
C2	CG2R67	0.243	C8	CG2R61	-0.0813	C5	CG2RC0	0.242
C3	CG2R61	-0.265	C12	CG2R61	0.0227	C6	CG2R61	-0.354
C7	CG2R67	-0.0233	O1	OG311	-0.599	C9	CG2R61	-0.0748
O	OG311	-0.678	H9	HGP1	0.298	H2	HGR61	0.244
H6	HGP1	0.472	S	SG302	0.353	S	SG302	-0.148
			O2	OG2P1	-0.701	O1	OG2P1	-0.223
			O3	OG2P1	-0.701	O2	OG2P1	-0.223
						N	NG3C51	-0.338
						H7	HGP1	0.378
BCA			BIPH			NAPO		
Atoms	Atom type	Charges	Atoms	Atom type	Charges	Atoms	Atom type	Charges
			C1	CG2R61	-0.155	C1	CG2RC0	0.373
			C2	CG2R67	-0.0036	C2	CG2R61	-0.329
			C3	CG2R61	0.212	C3	CG2R61	-0.192
C1	CG2R61	0.142	C7	CG2R67	-0.0036	C5	CG2R61	-0.0514
C2	CG2R67	-0.215	C8	CG2R61	0.212	C8	CG2R61	-0.449
C3	CG2R61	-0.0363	C12	CG2R61	-0.155	C9	CG2RC0	0.399
C7	CG2R67	0.113	O1	OG311	-0.547	C10	CG2RC0	0.103
C12	CG2R61	-0.113	H8	HGP1	0.374	H1	HGR61	0.305
O1	OG311	-0.586	O2	OG311	-0.547	S	SG302	-0.154
O2	OG2D1	-0.499	H10	HGP1	0.374	O1	OG2P1	-0.167
						O2	OG2P1	-0.167
						O3	OG3C51	-0.202

Table 3.3 Optimized bond and angle parameter for each molecule. Bonds and angles have been optimized as described here.

(a)HBP			(b)BCA		
Angle Parameters	K_{θ} (kcal/mol/rad ²)	θ_0 (°)	Angle Parameters	K_{θ} (kcal/mol/rad ²)	θ_0 (°)
C2 C1 O1	43.330	110.545	C1 C2 C7	36.487	114.086
(c)HBPS			(d)BIPH		
Angle Parameters	K_{θ} (kcal/mol/rad ²)	θ_0 (°)	Angle Parameters	K_{θ} (kcal/mol/rad ²)	θ_0 (°)
C7 C8 O1	85.005	118.771	C7 C8 O1	55.730	115.855
C2 C3 S	15.123	112.738			
(e)NTAM			(f)NAPO		
Bond Parameters	K_b (kcal/mol/Å ²)	b_0 (Å)	Bond Parameters	K_b (kcal/mol/Å ²)	b_0 (Å)
C3 S	134.889	1.843	C1 S	135.879	1.829
N S	126.251	1.765	O3 S	88.596	1.755
Angle parameters	K_{θ} (kcal/mol/rad ²)	θ_0 (°)	Angle Parameters	K_{θ} (kcal/mol/rad ²)	θ_0 (°)
C2 C3 S	81.319	125.877	C2 C1 S	61.166	121.527
C3 C4 C5	94.337	114.328	C1 C10 C9	112.940	112.353
C4 C3 S	136.049	113.105	C10 C1 S	140.678	108.541
C5 N S	73.526	110.111	C9 O3 S	62.850	115.994
S N H	70.303	110.249	C1 S O2	75.040	96.537
C3 S N	199.775	96.373	C1 S O3	212.043	94.171
C3 S O2	105.177	114.081	O2 S O3	101.025	98.555
N S O2	69.221	111.276			

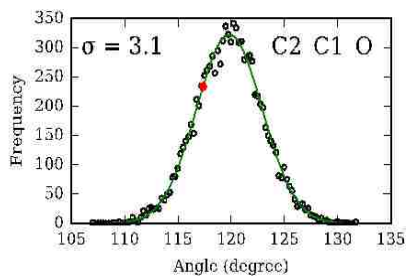
Table 3.4 Optimized dihedral parameters for each molecule. The atom types are given in Table 3.2. Dihedral have been optimized as described in the Method section.

(a) HBP				(b) BCA			
Dihedral parameters	K_φ (kcal/mol)	n	δ	Dihedral parameters	K_φ (kcal/mol)	n	δ
O C1 C2 C3	2.992	2	180	O2 C13 C1 C2	1.934	2	180
O C1 C2 C7	3.000	2	180	O1 C13 C1 C2	0.122	2	0
C2 C1 O H	0.746	2	180	C13 C1 C2 C3	2.727	2	180
				C13 C1 C2 C7	2.484	2	180
(c) HBPS				(d) BIPH			
Dihedral parameters	K_φ (kcal/mol)	n	δ	Dihedral parameters	K_φ (kcal/mol)	n	δ
S C3 C2 C1	1.010	2	180				
O1 C8 C7	3.000	2	180				
C12	1.976	2	0	O1 C8 C7 C12	3.000	2	180
C2 C3 S O2	0.751	2	180	O1 C8 C7 C2	3.000	2	180
C2 C3 S O2	2.240	2	180	C7 C8 O1 H	1.372	2	180
S C3 C2 C7	0.549	2	180				
O1 C8 C7 C2	0.708	2	180				
C7 C8 O1 H	0.888	4	0				
C7 C8 O1 H							
(e) NTAM				(f) NAPO			
Dihedral parameters	K_φ (kcal/mol)	n	δ	Dihedral parameters	K_φ (kcal/mol)	n	δ
C2 C3 C4 C5	2.970	2	180				
C5 N S O2	1.728	3	0				
C5 C4 C3 S	0.684	2	0	C2 C1 C10 C9	0.847	2	0
C4 C5 N S	2.997	3	180	C9 C10 C1 S	1.607	2	180
C4 C3 S O2	0.673	6	180	C10 C1 S O2	0.301	6	180
C3 C4 C5 N	0.851	2	0	C10 C1 S O3	0.638	2	0
C2 C3 S N	2.475	2	0	C8 C9 O3 S	1.871	2	180
H N S O2	1.056	3	0	C3 C2 C1 S	2.935	2	180
H N S O3	1.160	1	0	C9 C3 S O2	1.926	3	0
H N S O3	0.742	2	0	C10 C9 O3 S	2.603	2	180
H N S O3	0.662	3	180	C9 O3 S C1	0.810	3	0
C1 C2 C3 S	2.999	2	180	H C2 C1 S	2.340	2	180
C4 C3 S N	1.485	2	0	C1 C10 C9 O3	2.254	2	180
C5 N S C3	2.574	1	0	C5 C10 C1 S	2.207	2	180
C5 N S C3	1.494	2	180	C2 C1 S O2	0.013	6	180
H C2 C3 S	1.508	2	180	C2 C1 S O3	2.103	2	180
C9 C3 C3 S	0.054	2	180				
C6 C5 N S	0.711	2	0				
C2 C3 S O2	0.053	6	180				

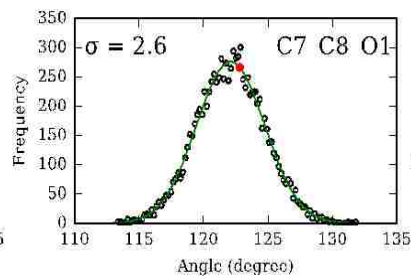
3.4.3 Comparing QM and MM equilibrium bond and angle values

To determine whether force field parameters were capable of sufficiently describing the optimized geometries, and thus the strength of the assigned force constants, we conducted 100-ns MD simulations of each molecule in explicit solvent; the newly optimized parameters were used as input to the simulations. From these simulations, we evaluated how the parameterized bonds and angles changed with respect to time. The data were binned into histograms for comparison with the value obtained from QM geometry optimization (Figure 3.4). In general, we expect the MM-sampled distribution will be centered around the QM value, representing a good fit.

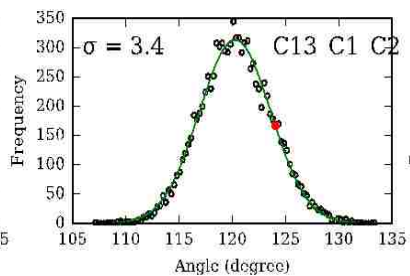
For each parameterized angle, the values sampled over the course of the 100-ns MD simulations exhibited a Gaussian distribution, which encompassed the QM-determined equilibrium value. For most, the distribution maximum aligned well with the QM equilibrium angle. In cases where the maximum was not identical to the QM-optimized geometry, for example the C13-C1-O2 angle of BCA (Figure 3.4(c)), competing constraints on the surrounding geometry made it difficult to enforce the optimum angle without application of large force constants. Despite not ideally replicating the optimized angle in every case, the QM optimized angle values all lie within the 95% confidence interval of the Gaussian distribution, suggesting the optimized parameters sufficiently describe molecular geometries.



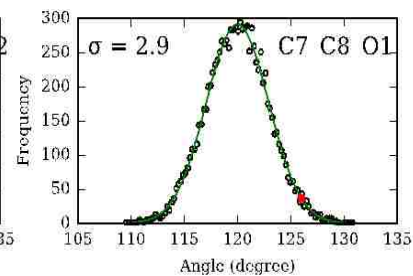
(a)HBP



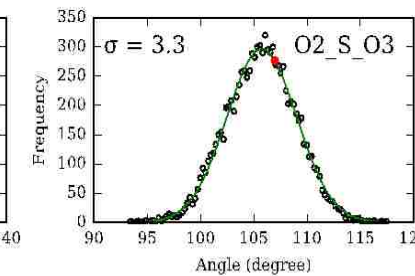
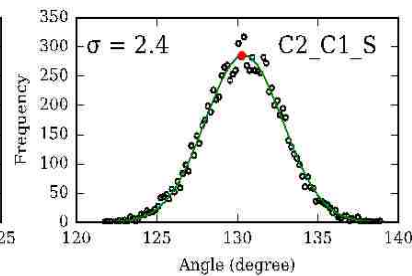
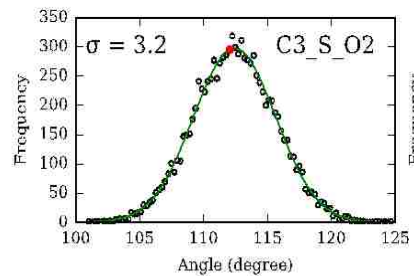
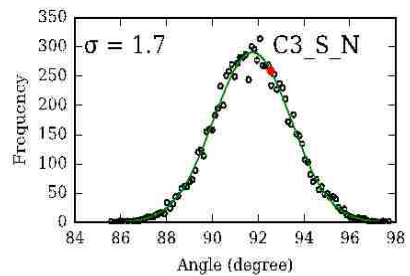
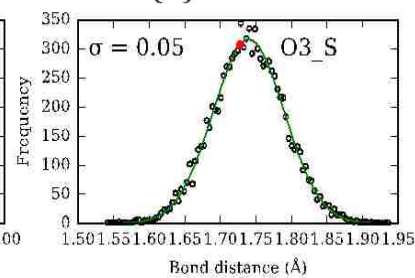
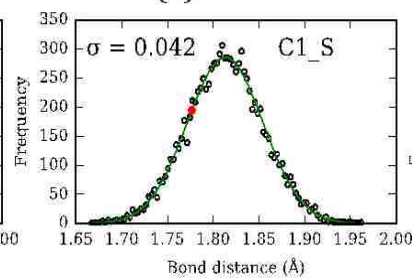
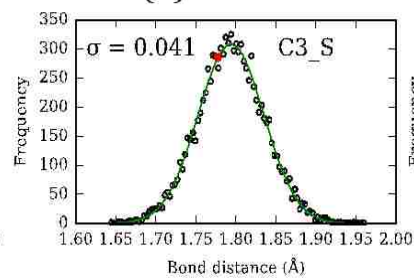
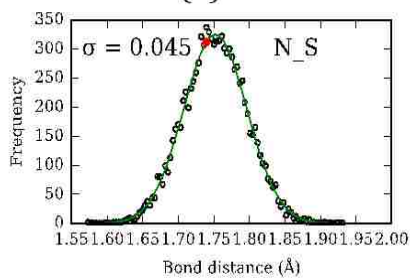
(b)HBPS



(c)BCA



(d)BIPH



(e)NTAM

(f)NAPO

Figure 3.4 Comparison of the QM-optimized bonds and angles to the bond distances and angles sampled over the course of a 100-ns MD simulation for (a) HBP, (b) HBPS, (c) BCA, (d) BIPH, (e) NTAM, (f) NAPO. The parameterized bonds and angles were measured in each frame of the MD simulation; the measurements were binned into histograms (unfiled blue circles) and fit with a Gaussian distribution (green line). The σ given in the top left of each plot is the standard deviation of the histogram fit to a Gaussian distribution. The x-axes represent either the value of the bond distance (\AA) or the angle ($^\circ$) and the y-axes represent frequency of a measurement, as sampled over a 100-ns MD simulation. The red filled circles represent the QM optimum value for a given bond length or angle.

3.4.4 Comparison of calculated and experimental IR spectra

To validate the optimized parameters, we calculated IR spectra for each of the molecules from implicit solvent MD trajectories obtained using the parameters. The calculated spectra were directly compared with experimental spectra (Figure 3.5). For all molecules studied, calculated and experimental spectra were in good agreement in low wavenumber ranges ($<2000\text{ cm}^{-1}$) which were attributed to stretching modes between heavy atoms and C-H bending modes (Table 3.5). This agreement largely supports vibrational similarity between the parameterized model and the heavy atom structure of these compounds.

However, for the higher frequencies ($>2800\text{ cm}^{-1}$), where C-H and O-H bond stretching appears, the calculated frequencies do not align well with experimental peaks, varying by up to 150 cm^{-1} . The largest mismatches are for the O-H stretching

modes. The position of O-H stretching modes are highly sensitive to solvent interactions. In particular, intramolecular and intermolecular hydrogen bonding in protic solvents give distinct shifts. While calculations were completed with an implicit protic solvent, the available experimental data were obtained using aprotic solvents (HBP), solid state (HBPS), and the gas phase (BCA, BIPH); solvation state data for NTAM and NAPO experimental spectra were unavailable. The aromatic C-H stretching calculations were in good agreement (typically <math><50\text{ cm}^{-1}</math> higher) with the experimental data. These inaccuracies at higher frequencies also arise from anharmonicity effects resulting from the manner in which theoretical spectra are calculated [107, 108]. In the spectra calculations, the system was modeled as harmonic. In the low frequency ranges, the harmonic model is capable of accurately predicting IR spectra by modeling bond stretches as a mass-spring system, satisfying experimental agreement. However, the harmonic mass-spring model begins to break down at higher frequencies, as the relationship between force and displacement is nonlinear. An anharmonic oscillator can be used to represent regions not well described by a harmonic model, but this also non-ideal. The anharmonic nonlinear region is not easily calculated. Thus, we applied the harmonic model to calculate the high frequency region, resulting in the energy being coupled to the fundamental vibration frequency, and thus, overestimating wavenumbers.

Table 3.5 Experimental IR spectra frequency values compared with theoretical predictions using MM-computed parameters.

HBP		
Assignment	Experimental frequency (cm^{-1})	Calculated frequency (cm^{-1})
Aromatic; <i>sp</i> ² ; bending C-H	750	758

Aromatic C-O	1182	1187
Aromatic stretching C=C	1480	1364
Biphenyl; stretching C-C	1590	1489
Aromatic; <i>sp</i> ² ; overtones C-H	1660 to 2000	-
Aromatic; <i>sp</i> ² ; stretching C-H	3049 & 3077	3060 & 3110
O-H	3561 & 3611	3473 & 3755

HBPS

Aromatic; <i>sp</i> ² ; bending C-H	756	772
Aromatic C-O	1002	982
C-S	1296	1290
Stretching S=O	1462	1464
Aromatic; stretching C=C	1460	1432
Biphenyl; stretching C-C	1593	1517
Artifact from diamond ATR surface	1786 to 2372 *	-
Aromatic; <i>sp</i> ² ; stretching C-H	3051	3045 & 3111
O-H	3051 2670 to 3650	3760

* The peaks at 1980, 2130, and 2160 are artifacts from the diamond ATR surface

BCA

Aromatic to Acyl; orthosubstituted	694	627
C-C	750	763
Aromatic; <i>sp</i> ² ; bending C-H	1346	1262
Acid; C-O	1482	1327
Aromatic; stretching C=C	1550	1447
Biphenyl; stretching C-C	1770	1757
C=O	3034 & 3070	3046 & 3178
Aromatic; <i>sp</i> ² ; stretching	3572	3673
O-H		

BIPH

Aromatic; <i>sp</i> ² ; bending C-H	750	765
Aromatic C-O	1194	1180
Aromatic; stretching C=C	1474	1432
Biphenyl; stretching C-C	1586	1512
Aromatic; <i>sp</i> ² ; stretching C-H	3046 & 3074	3053 & 3110
O-H	3578	3757

NTAM

Bending SO ₂	602	705
Stretching C-S	755	755
Stretching S-N	812	826
Ring breathing	1056	998
Symmetric stretching SO ₂	1142	1177
Stretching C-N	1289	1299
Asymmetric stretching SO ₂	1363	1347
Aromatic; stretching C=C	1488	1427
Bending N-H	1590	1564
Stretching N-H	3244	3422
Aromatic; <i>sp</i> ² ; stretching C-H	3082	3110

NAPO

Bending SO ₂	520 & 620	697
Stretching C-S	755	732
Stretching S-O	812	782
Stretching C-O	977	1013
Ring breathing	1034	1045
Symmetric stretching SO ₂	1187	1209
Asymmetric stretching SO ₂	1357	1331
Aromatic; stretching C=C	1485	1502
Aromatic; <i>sp</i> ² ; stretching C-H	3077	3106
Unassigned sultone	3444 (broad peak)[109, 110]	-

Table 3.5 gives a side-by-side comparison of the vibrational phenomena observed in the experimental and calculated IR spectra. We discuss only the low frequency range phenomena given the limitations of theoretical spectra calculations described above. Within the low frequency range, the biphenyl compounds all exhibit aromatic and biphenyl carbon-carbon stretching and aromatic carbon-hydrogen stretching. In all compounds, carbon-hydrogen bond stretching in the aromatic rings is reasonably well predicted by calculation. Carbon-carbon bond stretching, both aromatic and biphenyl, is perhaps the least accurate vibrational phenomena in the low frequency region. As the bond and angle parameters primarily defining aromatic carbon-carbon stretching have not been modified from the CGenFF parameters, this small discrepancy is likely common to all aromatic molecules modeled with the CGenFF force field. Otherwise, the biphenyl compound parameters largely reproduce the vibrational phenomena observed in the experimental spectra. The planar naphthenic parameters are capable of reproducing the low frequency ring-based, carbon-sulfur vibrational phenomena, and SO₂ asymmetric and symmetric stretching and bending. Overall, the optimized

parameter set adequately predicts primary vibrational phenomena in classical molecular simulation.

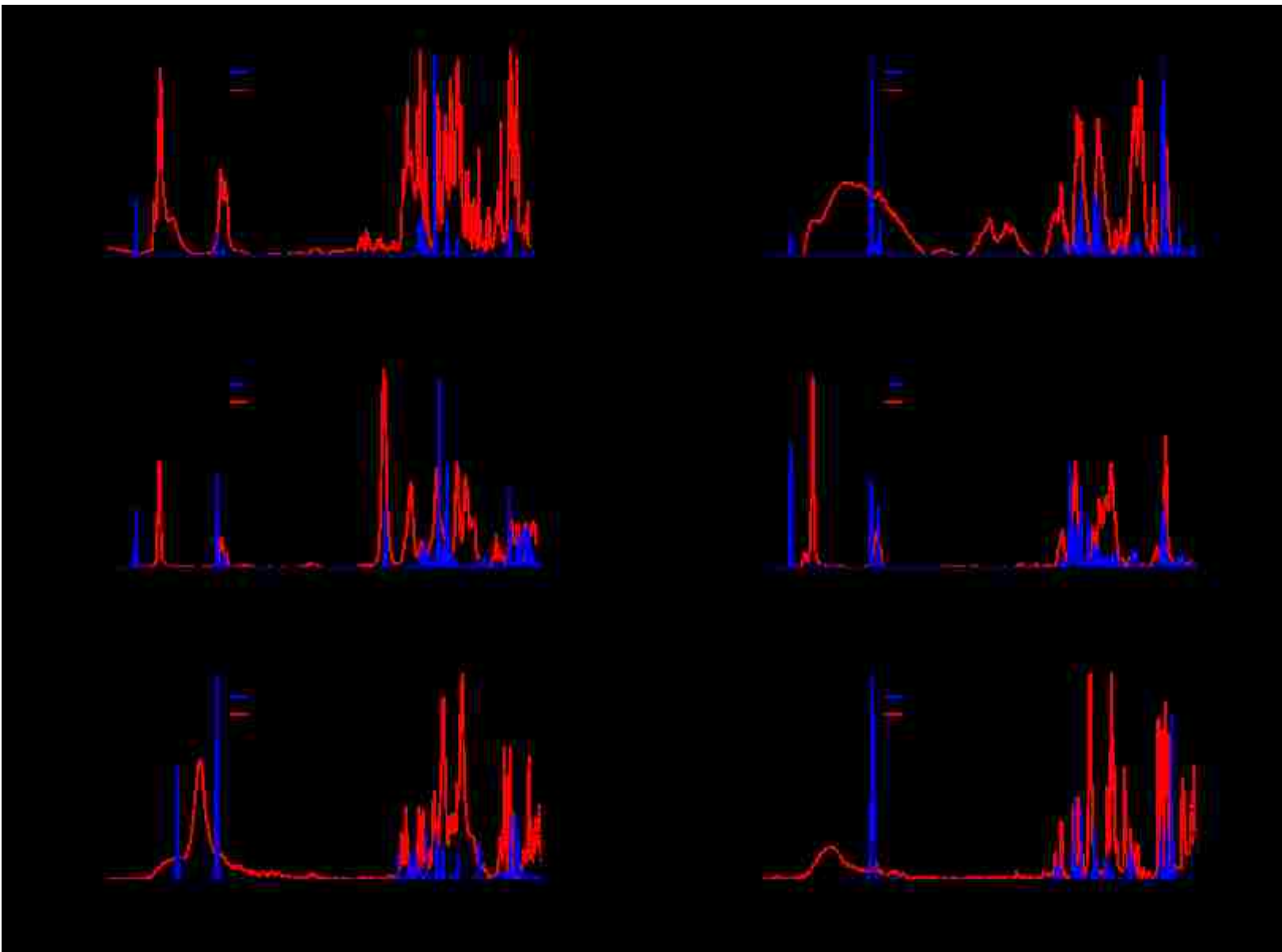


Figure 3.5 Experimental and theoretical IR spectra for HBP, HBPS, BCA, BIPH, NTAM, and NAPO. The HBPS spectrum was obtained as described in the Methods section. The left and right-hand y-axes represent spectral density (computational) and % absorbance/transmittance (experimental), respectively. The magnitude of the peaks has no physical meaning. Rather, the existence of a peak at an appropriate wavenumber, representative of a vibrational phenomenon, is the primary point of comparison.

3.5 Conclusions

Here, we have determined CHARMM-compatible molecular mechanics force fields for HBP, HBPS, BIPH, BCA, NTAM, and NAPO using the ffTK optimization tools available in VMD. The force field parameters are provided in the separated files, including a .psf file with the partial atomic charges and the bond, angle, and dihedral parameters in a .par file for each molecule. Explicit solvent molecular simulation confirmed the ability of the implemented parameter force constants to maintain the optimized geometry equilibrium values within the 95% confidence interval. The optimized parameters were used to calculate theoretical IR spectra, which were then compared to experimental spectra from literature or generated as part of this study. In general, the theoretical spectra were in good agreement with experiment at lower frequencies, but calculated spectra were not particularly accurate in the high frequency ranges as a result of both solvent effects and anharmonicity not captured by harmonic models. Overall, the parameters sufficiently capture the equilibrium geometry and vibrational phenomena of a subset of biphenyl and planar naphthenic molecules that naturally occur in unrefined crude oil. Thus, these parameters can be implemented in biomolecular simulations of 4S

pathway enzymes such as *DszB* to understand the molecular-level interactions driving substrate binding and product inhibitions.

Chapter 4 — Conformational change of 2'-hydroxybiphenyl-2-sulfinate desulfinate upon ligand binding

In this chapter, we applied the CHARMM-compatible force fields generated in Chapter 3 to perform MD simulations on *DszB*-HBPS and *DszB*-HBP. The purpose of such simulations was to identify secondary structure changes that occur in *DszB* as a function of ligand binding. The work in this chapter provides insight into the molecular-level mechanisms governing substrate specificity and product inhibition of *DszB*.

4.1 Summary

Biodesulfurization of petroleum products using the 4S enzymatic pathway can effectively remove sulfur from refractory thiophenic molecules at ambient conditions. 2'-hydroxybiphenyl-2-sulfinate (HBPS) desulfinate (*DszB*) catalyzes the cleavage of the carbon-sulfur bond to produce 2-hydroxybiphenyl (HBP) and sulfite in the final step of the 4S pathway, and it exhibits the slowest reaction rate, partially as a result of product inhibition. Here, we examine a large-scale protein conformational change, which has implications for understanding the mechanism of product inhibition. Molecular Dynamics (MD) simulation was used to analyze how HBPS and the inhibitory product, HBP, bind with *DszB* in terms of binding position, orientation, and differences in molecular interactions. Three loops, including residues 50-60 (loop 1), 135-150 (loop 2), and 180-200 (loop 3), were identified as participating in the dominant conformational change (open-close) upon binding. Umbrella sampling provided a quantitative assessment of the energetic penalty for the gate conformational change, which was 2.4 kcal/mol and independent of bound

ligand and solvent conditions. We also performed MD simulations in hexane-water (0.5 vol/vol) biphasic solution and observed HBP leaving binding pocket through the proposed gate. The dynamic changes of three gate loops are likely not involved in product inhibition, but the high proportion of non-polar residues along the binding pocket exit provided strong non-bonded interactions and readily available hydrogen-binding partners that made it difficult for HBP to escape the binding pocket. We anticipate these latter interactions are primary contributors to HBP inhibition in *DszB*.

4.2 Introduction

Currently, the most widely implemented method for desulfurization is metal-catalyzed hydrodesulfurization (HDS), which effectively removes sulfur from light refinery fractions, but leaves a significant amount of polycyclic organic sulfur, such as dibenzothiophene (DBT), in the middle distillate and heavy oil fractions [8-11, 111]. DBT, in particular, is considered a model compound for polycyclic organic sulfur compounds in fossil fuels, and degradation of DBT by microbial organisms has been mainly studied for biodesulfurization (BDS) [112, 113]. BDS is an enzymatic process by which organosulfur compounds recalcitrant to HDS can be removed at ambient conditions [11, 114, 115]. Therefore, BDS can potentially be used in conjunction with hydrodesulfurization to reduce sulfur content to regulatory levels [8, 15]. However, improving the overall rate of enzymes that desulfurize polycyclic molecules remains a primary technical challenge preventing economical implementation of biodesulfurization [16, 17].

Among different microbial desulfurization methods, the 4S pathway, a sulfur-

specific DBT degradation pathway capable of transforming DBT to sulfite and 2-hydroxybiphenyl (HBP) without affecting fuel heating value, is the most widely studied (Figure 1.1) [18, 116]. However, in the 4S pathway, the final catalytic step has been widely identified as rate-limiting, making *DszB* a logical target for activity improvements [20, 115]. An additional hindrance to overall *DszB* conversion rates is product inhibition, where HBP is responsible for the observed reduction in biocatalyst activity concomitant with HBP generation [117].

There is a significant conformational change upon substrate binding in *DszB* that may be important for positioning the catalytic residue in a productive conformation within the active site. The conformational change was first observed in by Lee *et al.*, who solved three crystal structures to examine the effects of substrate binding on *DszB* structure, including the unbound (apo) structure of *DszB* (PDB code 2DE2) and a C27S catalytically inactive variant bound with HBPS (PDB code 2DE3) [39]. Superimposition of these two structures revealed a distinct conformational change, which was particularly noticeable in three distinct residue loops regions. In Figure 1.4, the superimposed structures are shown in yellow and red cartoon illustrating apo-*DszB* and HBPS-bound *DszB*, respectively. The previously extended loop1 structure (residues 50-60) in the apo structure became an α -helix which introduced H60 into active site; loop 2 (residues 135-150) kept an α -helix form but moved upwards; and loop 3 (residues 180-200), similar to loop 1, changed from an extended structure to an α -helix. As a result, we proposed that the three loops (Figure 1.4) formed a “gate” that changes from “open” to “closed” upon substrate binding; moreover, this gate region is thought to be responsible for ligand

egress and, accordingly, product inhibition.

We anticipate using Molecular Dynamics (MD) simulations to qualitatively analyze how HBPS and the inhibitory product, HBP, bind with *DszB* in terms of binding positions varieties and protein loop dynamics differences. Biphasic MD simulation will be applied on *DszB*-HBP to observe how HBP exit the binding pocket and validate the proposed gate region. Umbrella sampling will be applied to obtain quantitative results on how much energy associate with the gate area conformational change.

4.3 Methods and Materials

4.3.1 Molecular dynamic simulation

MD simulations were run for 200 ns in triplicate for both *DszB*-HBPS and *DszB*-HBP in aqueous buffers. Three additional MD simulations for *DszB*-HBP in biphasic buffer were conducted. Simulations were constructed starting from changing the bound state *DszB* mutation S27 back to its catalytically active state, C27, based on the *DszB*-HBPS structure (PDB code 2DE3) [39]. Although a *DszB*-HBP crystal structure is not available, HBP is the product of the reaction, therefore, removing the sulfite group from HBPS resulted in the *DszB*-HBP structure. The CHARMM General Force Field (CGenFF) with force field parameters version 3.0.1, generated from our previous work, were applied to small molecules HBPS and HBP, and CHARMM36 force field with the CMAP correction was used to protein [35, 47], [87, 89]. Protonation states of all the titratable residues in *DszB* were calculated by the H++ web server from the experimental suggested optimal pH=7.4 [20, 118].

Aqueous solution solvation and minimization: Each protein-ligand complex was solvated in a cubic periodic cell of TIP3P water model using the Multiscale Modeling Tools for Structural Biology (MMTSB) Tool Set [99]. Additionally, 0.3M NaCl, and 50mM HPO₄²⁻ were included in the solvent to mimic experimental media; extra sodium ions were added to neutralize the overall system charge (*DszB* is -17e and HBPS is -1e). The distance between the protein and the edge of the periodic cell was 13Å, resulting in a periodic system of 90 Å x 90 Å x 90 Å. After solvation, the system was minimized in CHARMM with 2000 steps of Steepest Descent (SD) algorithm holding the protein and ligand fixed and another 2000 steps of SD with only the protein restrained. Finally, 2000 steps of SD and 5000 steps adopted basis Newton-Raphson (ABNR) algorithm were applied to the whole system with no applied restraints.

Biphasic solution solvation and minimization: *DszB* was treated as a sphere, and its radius was determined by selecting the distance between the center of *DszB* and its farthest atom. A hexane molecule was built using Avogadro molecular editor (v.1.1.1) to generate a coordinate file (.pdb file format) [119]. The hexane molecule structure was optimized quantum mechanically using Gaussian09 and the MP2/6-31G* basis set [90, 91, 120]. Given an aqueous solution periodic cell size (90Å), a 0.5 vol/vol of hexane to water requires 1111 geometry optimized hexane molecules and 16254 TIP3P water molecules, which were randomly placed inside the cubic box avoiding the spherical *DszB* region. Similarly, 0.3M of NaCl, 50mM HPO₄²⁻, and additional sodium ions were included to neutralize the charge. Since there were several low-density regions surrounding the protein as a result of

the solvation procedure, a more extensive minimization step was performed than for the aqueous systems. First, the biphasic system was minimized with 2000 steps of SD followed by 5000 steps of ABNR, with only hexane molecules flexible. This was followed by another 2000 steps of SD and 5000 steps of ABNR with the water molecules and ions free to move. Following with 2000 steps of SD and 5000 steps of ABNR applied with only protein-ligand restrained. Finally, 5000 steps of SD and 5000 steps of ABNR applied to the whole unrestrained system.

The solvated and minimized systems were then density-equilibrated prior to the MD production simulations with hexane CGenFF force field version 3.0.1. Systems were first heated from 90K to 300K in 50K increments over 100ps in the canonical ensemble. This was followed by isothermal-isobaric (NPT) simulation using the Nosé-Hoover thermostat and barostat in CHARMM for 200 ps at 300 K and 1 atm. Production MD simulation of the density-equilibrated systems was performed in NAMD[53] for 200 ns with 2 fs time step at 300 K in the canonical constant volume NVT ensemble in NAMD. The trajectory was saved every 5000 steps for aqueous systems and 50000 steps for biphasic systems.

4.3.2 Principle component analysis (PCA)

Principle component analysis (PCA) is a statistical technique for finding patterns of a high-dimensional dataset [121]. The PCA was performed using the R Bio3D package, version 2.3.1 [122], and it was based on MD simulation Cartesian coordinate snapshots in which PCs describe the concerted atomic displacement to highlight the major conformational change of the system.

The mathematical description of PCA is shown in Eq. 4.1, in which C is the data covariance matrix, Λ is the diagonal matrix containing the eigenvalues as diagonal entries, and V contains the eigenvectors. If the eigenvectors are sorted in eigenvalues decreasing order, the eigenvector with the largest eigenvalue is called the first principle component (PC1), and it accounts for the most variance of the data. The second component is orthogonal to the first one (PC2) and accounts for the second largest variance and so on.

$$C = V\Lambda V^T \quad 4.1$$

PCA was applied to the Cartesian coordinates of the backbone carbon atoms of *DszB* along the MD trajectory, resulting in a lower dimensional representation of the structural dataset. We selected the first two PCs (which represent more than 40% of total structure variances) on which to project MD conformers; the results divide MD conformers into several different groups on a 2D plot (see Figure 4.1). Therefore, protein conformational changes along MD trajectories can be revealed on a low-dimension plot.

4.3.3 Clustering

Clustering analysis is another statistical technique for finding patterns. This method groups similar objects into subgroups by minimizing intra-cluster and maximizing inter-cluster differences. The hierarchical average-linkage clustering algorithm was applied based on the first 3 PCs. According to PCA results, MD conformers have been divided into distinct groups onto 2D plot. Combined with clustering analysis, each group with its corresponding time frame was obtained

from clustering dendrograms along the MD trajectory. Each clustering subgroup structures with its typical represented time frame was obtained in order to generate corresponding grouped structures. Clustering analysis was performed using R Bio3D package version 2.3.1.

4.3.4 Umbrella sampling

Umbrella Sampling was used to quantitatively determine the amount of energy required for gate changes from close to open. The root mean square deviation (RMSD) values of backbone atoms from three loops (total were 48 atoms), traversing from the two-known end-point conformations, was used as the reaction coordinate. The equilibrated holo structure (closed state) served as the initial structure, and the open state, with PDB code 2DE2, served as final structure. Targeted MD (TMD) simulations were performed using NAMD. The windows for umbrella sampling were determined from the range of RMSD along the reaction coordinate, 4.84 Å to 0.74 Å. Forty-one windows were generated with RMSD change from 4.84 Å to 0.74 Å with 0.1 Å decrement. The force applied was 20 kcal/(mol•Å²) with 400-ps MD simulation carried out in each window at 0.1 Å RMSD decrement for the “window making” process. The end point coordinates from each window were selected for the umbrella sampling, and the harmonic biasing force applied to each window was 5 kcal/ (mol•Å²) to restrain the structure at its final RMSD value of each window for a 10-ns simulation. The potential of mean force (PMF) profiles were calculated using the weighed histogram analysis method (WHAM) version 2.0.9, and error analysis was applied by computing the autocorrelation function and determining the correlation time to apply Monte Carlo Bootstrap Error Analysis in

WHAM [123]. Convergence analysis of umbrella sampling is illustrated in Figure S.1 of the Appendix.

4.4 Results and Discussions

4.4.1 MD simulations of *DszB*-HBPS and *DszB*-HBP in aqueous solution

PCA applied to MD trajectories revealed four major conformational representatives along MD trajectories for both *DszB*-HBPS and *DszB*-HBP in aqueous buffer. Each conformational structure representative was obtained by averaging structures from suggested represented time steps and superimposition group 1 (blue) and 4 (red) (Figure 4.1) corresponding to MD conformers changes from initial to final. For PCA, results showed here only represent one of the MD simulations from each system among three parallel simulations, the rest were presented in appendix Figure S.2 and S.3. The first two components of PCA revealed about 40% of the total variance for both systems, see supplemental material Figure S.2 (a)(b). After PCA, trajectories data were plotted onto a 2d plane against the first two PC scores (Figure 4.1(a) (b)), both systems structures went through four major conformational changes represented with the color change from blue to red. Hierarchical cluster analysis applied on the first three PCs revealed four major group clusters, and the represented groups with the corresponding time frame were applied onto RMSD plots (Figure 4.1 (c)(d)).

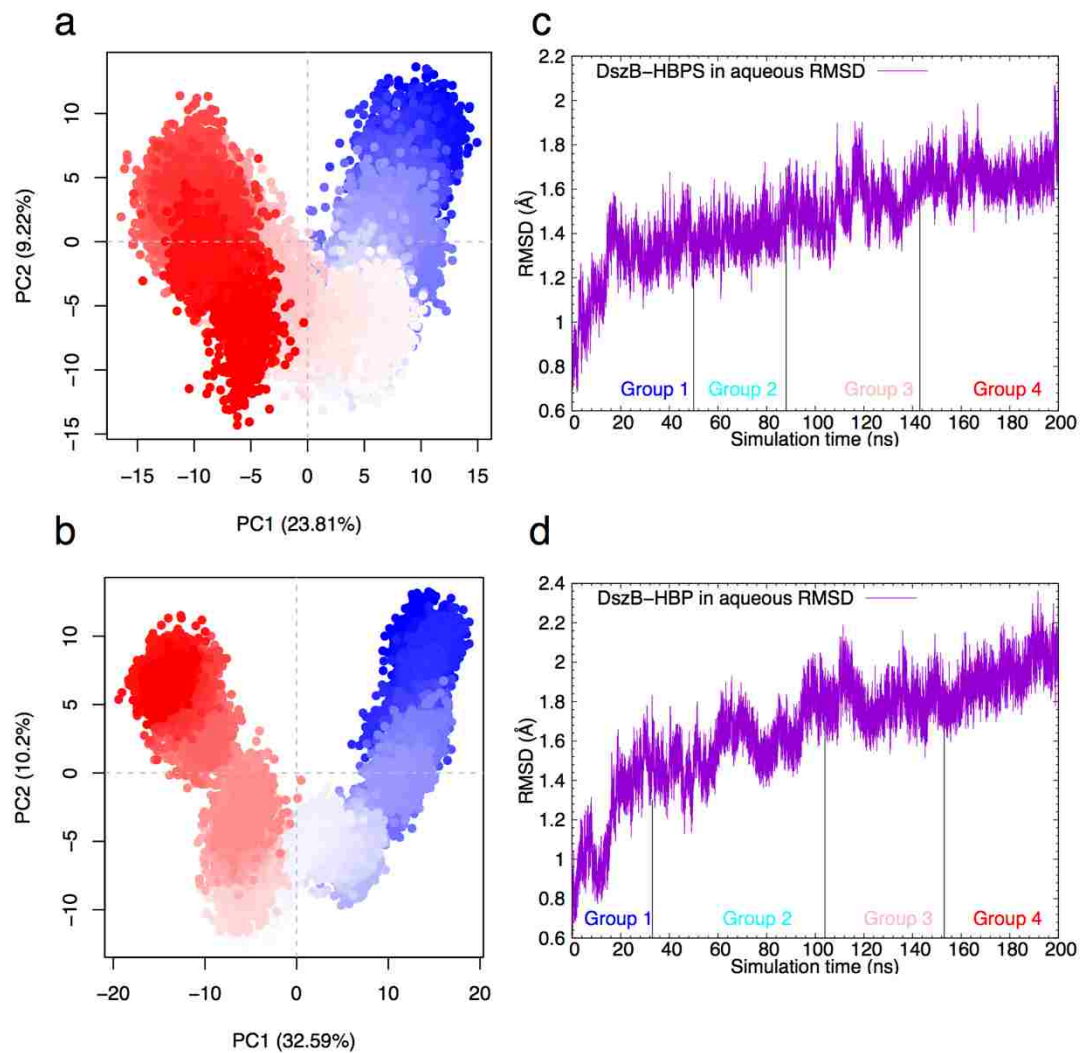


Figure 4.1 (a) (b) PCA results of *DszB*-HBPS and *DszB*-HBP from application of MD trajectories. PC2 is plotted against PC1, where the color transition from blue to red corresponds to MD simulation time from start to end. The % shown on the x, y axes represent PCA scores. (c)(d) RMSD with time shown with the clustering analysis-suggested groupings along the corresponding time step.

MD conformers superimposition analysis. MD simulations suggest that HBPS and HBP bind at different positions inside *DszB*. When HBPS in the active site, it stayed stable (Figure 4.2(a)) with loop 1 and 2 kept in helical structure. Catalytic

residues H60 remained inside the binding pocket of *DszB* without position change and maintained a short distance between further N on the imidazole ring, denoted as N_T , of H60 and sulfinate O of HBPS to be 5.5 Å. The charged catalytic residue R70 maintained a close interaction with HBPS based on ionic interaction, where the distance between N_η from R70 and O from HBPS remained about 2.7 Å. As a result, along the 200-ns MD simulation HBPS stayed very stable in the binding site, all catalytic residues (C27, H60, and R70) remained in their original positions to maintain interactions to HBPS sulfinate group. However, when HBP was in the active site (Figure 4.2(b)), it dislodged from the bound state during the initial 90 ns. After 90 ns, HBP migrated to the proposed gate region and stayed stable during the rest of simulation, which suggested that the proposed gate area provided more favorable binding residues to HBP than the active site of *DszB*. Catalytic residue H60 exhibited high mobility on rotation of the imidazole ring when HBP moved towards to the gate region, and the distance between N_T -H from H60 to O from hydroxyl group of HBP was 2.1 Å. Based on binding position change of HBP, C27, which is the experimentally determined catalytic acid [39], moved further away from active site with distance between S-H from C27 to O from hydroxyl group of HBP changing from 2.3 Å to 11.1 Å, which suggested that high HBP mobility in the binding site but high stability around the gate enabled catalytic residues located close to binding site such as C27 and H60 to become highly mobile.

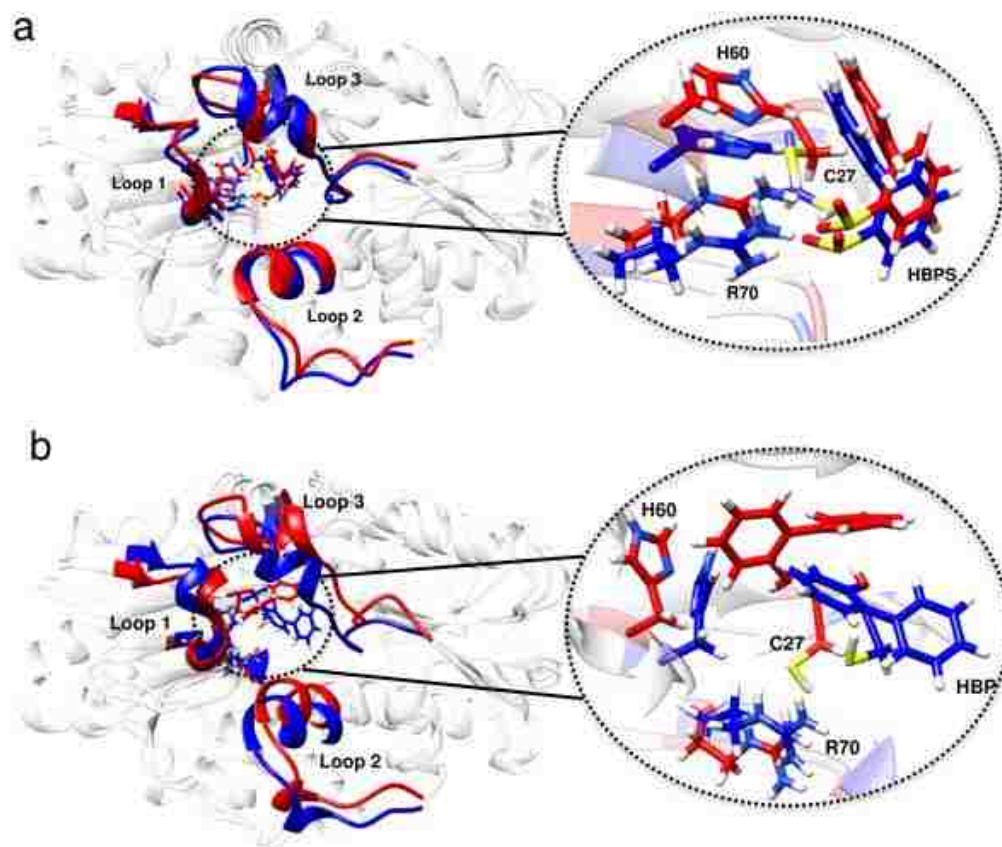


Figure 4.2 (a) (b) Superimposition of group structures 1 and 4 for *DszB*-HBPS and *DszB*-HBP in aqueous solution, averages of 5000 structures from 0–50 ns and 5700 structures from 143–200 ns for HBPS, and 3300 structures from 0-33 ns and 4700 structures from 153-200ns for HBP respectively. Blue and red structural elements correspond to group 1 and group 4, respectively. The proposed gate loops are highlighted and correspond to the dominant conformational change that occurred over the MD simulation trajectories. Important catalytic residues, C27, H60 and R70, are labeled on the right insets.

Secondary structure analysis: The proposed gate area is the most active region among all systems regardless of the bound ligands throughout the 200-ns MD simulation in aqueous solution. Binding induced conformational change revealed

three loops that may potentially be responsible for ligand egress. When ligands HBPS and HBP bound inside *DszB*, the highlighted three loops exhibited the most significant flexibility. More specifically, loop 1 always kept turn structure but constantly changed between helical to turn structure. Loop 3, however, slowly became disordered, and after 100 ns, loop 3 became partially disordered (Figure 4.3(b)). When nothing was bound inside *DszB*, the originally extended loop 1 with H60 pointed out from the active site actively changed between turn structures to disordered structure during the MD simulation (Figure 4.3(c)); therefore, the high flexibility of loop 1 may be responsible for ligand ingress into active site. When colored by root mean square fluctuation (RMSF), right panel of Figure 4.3, the color change from red, yellow, to green as well as thickness change from thin to thick represent the least to the most fluctuating regions during the MD trajectories; Figure 4.3 clearly shows that, for all the systems, the proposed gate area was the most active region along the simulation. As a result, the highlighted loop areas had the most significant flexibilities, and the high activity of the three loops during the simulations potentially related to ligands egress/ingress functionality.

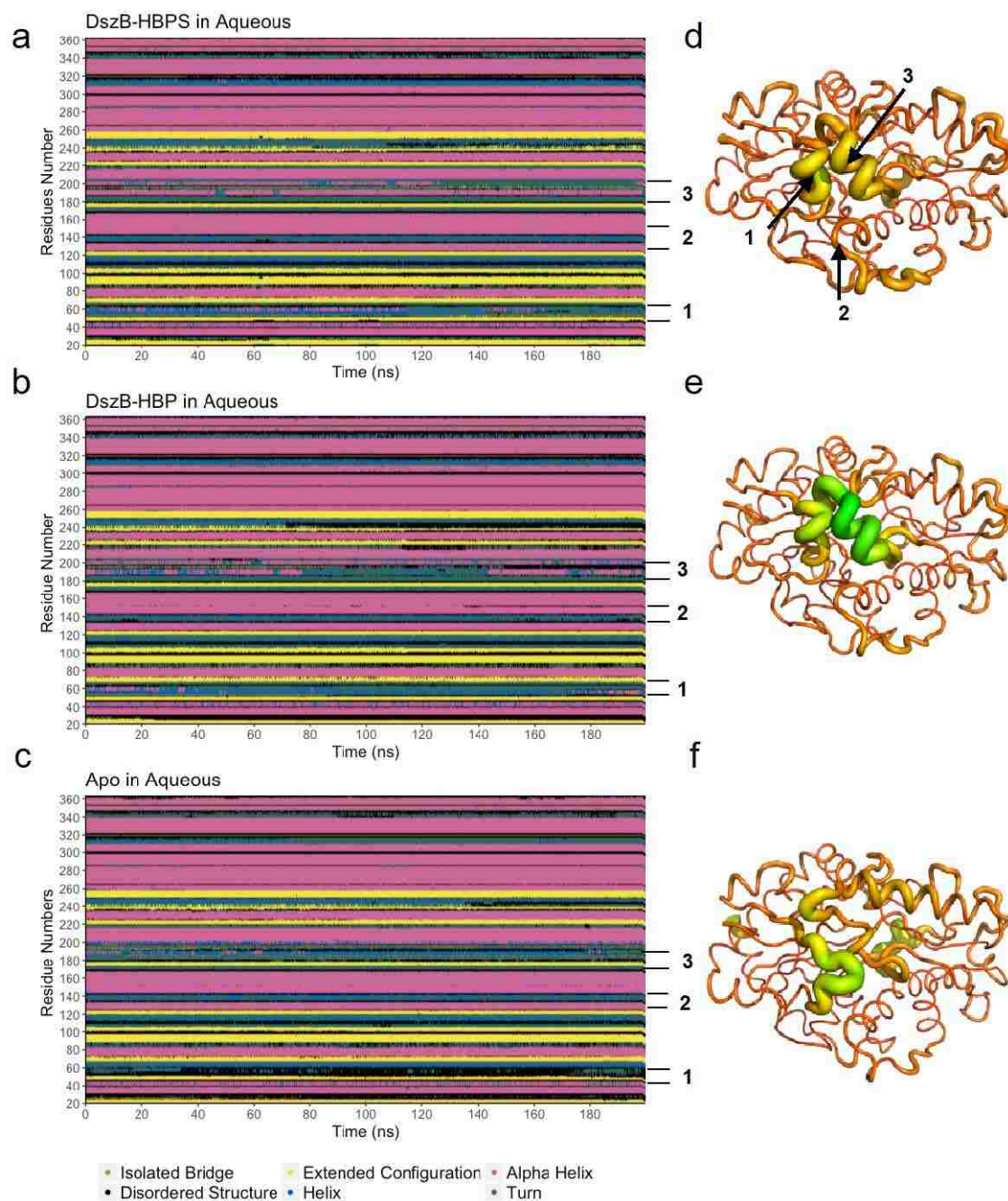


Figure 4.3 Secondary structure analysis along the MD trajectories. Structure analysis was performed using VMD [90]. (a) *DszB*-HBPS in aqueous solution. The three gating loops are labeled as 1, 2, and 3; (b) *DszB*-HBP in aqueous solution; (c) apo-*DszB* in aqueous solution. In panels a-c, the colors represent secondary

structure changes with time. An isolated bridge is shown in gray, disordered regions in black, extended configuration in yellow, helices in blue, α -helices in gold, and β -turns in dark green. The right panels, d-f, are β -factor plots, using calculated RMSF applied as the β -factor, projected onto the *DszB* structures: (d) *DszB*-HBPS, (e) *DszB*-HBP, and (f) apo-*DszB*. The color change from red-yellow-green corresponds to regions of low to high fluctuation. Additionally, structure labeled thin to thick also serves as a visual indicator of flexibility.

Non-bonded interaction energy and hydrogen bond (h-bond) analysis.

HBP had fewer direct residue-residue interactions and formed fewer hydrogen-bonds compared to HBPS inside binding pocket; instead, nearly all the interactions formed between HBP and *DszB* were dominantly van der Waals (VDW)-based and were concentrated at the gate area. Figure 4.4 shows non-bonded interaction energy, including VDW and electrostatic interactions, between the ligand and its nearby residues for both *DszB*-HBPS and *DszB*-HBP in aqueous solutions. For the charged molecule, HBPS, the nearby basic amino acid R70, with a positively charged guanidinium group, resulted in large electrostatic interactions to the negatively charged sulfinate group of HBPS; it was this interaction that was primarily responsible for the stability of HBPS inside the active site. In the case of the neutral non-polar HBP, the charge-charge electrostatic interactions vanished, but nonpolar residues W145 and V186 from loops 2 and 3 produced new VDW interactions.

Non-polar residues within the gate appeared to be responsible for the sluggishness of HBP's expulsion from the protein binding pocket. Even though the non-bonded interaction of HBP within the protein is about 70 times smaller than the

charge-charge electrostatic energy of HBPS with R70 to HBPS, these VDW interactions were numerous enough to prevent HBP from migrating out of binding pocket through the proposed gate region. Similarly, the charged sulfinate group of HBPS formed more h-bonds inside binding site than HBP since the hydroxyl oxygen can only be an h-bond donor (Figure 4.5). All h-bonds formed between HBP and *DszB* were around gate region. Therefore, h-bond interactions between HBP and *DszB* provided additional hindrance for HBP from moving out of the binding pocket through proposed gate. The h-bonds between HBPS and *DszB* inside the catalytic binding site were formed with catalytic residues C27, H60, R70, and G73, all of which are critical for stabilizing the HBPS orientation inside the binding site for reaction. As a result, HBP inhibition of *DszB* is likely related to HBP interactions with the proposed gate area residues that add difficulty for HBP egressing from *DszB* binding pocket. Since HBP showed high flexibility at the active site, and during MD simulation it clearly showed that HBP migrated to the gate area and exhibited higher stability around the gate region rather than active site, we suggest that inhibition of *DszB* by HBP occurs through allosteric competitive binding effects, that HBP binds at the gate area.

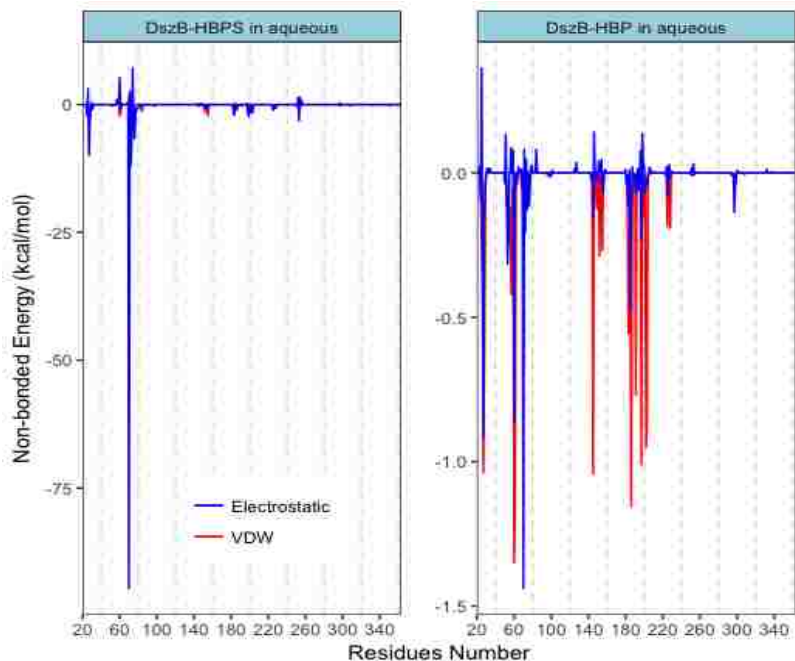


Figure 4.4 Non-bonded interaction energies of *DszB* with HBPS and HBP. Electrostatic interactions (blue) and VDW interactions (red) of the ligand and surrounding residues in the active site are shown as a function of residue number. The left panel is a snapshot of the *DszB*-HBPS interaction energies, and the right panel is *DszB*-HBP interaction energies.

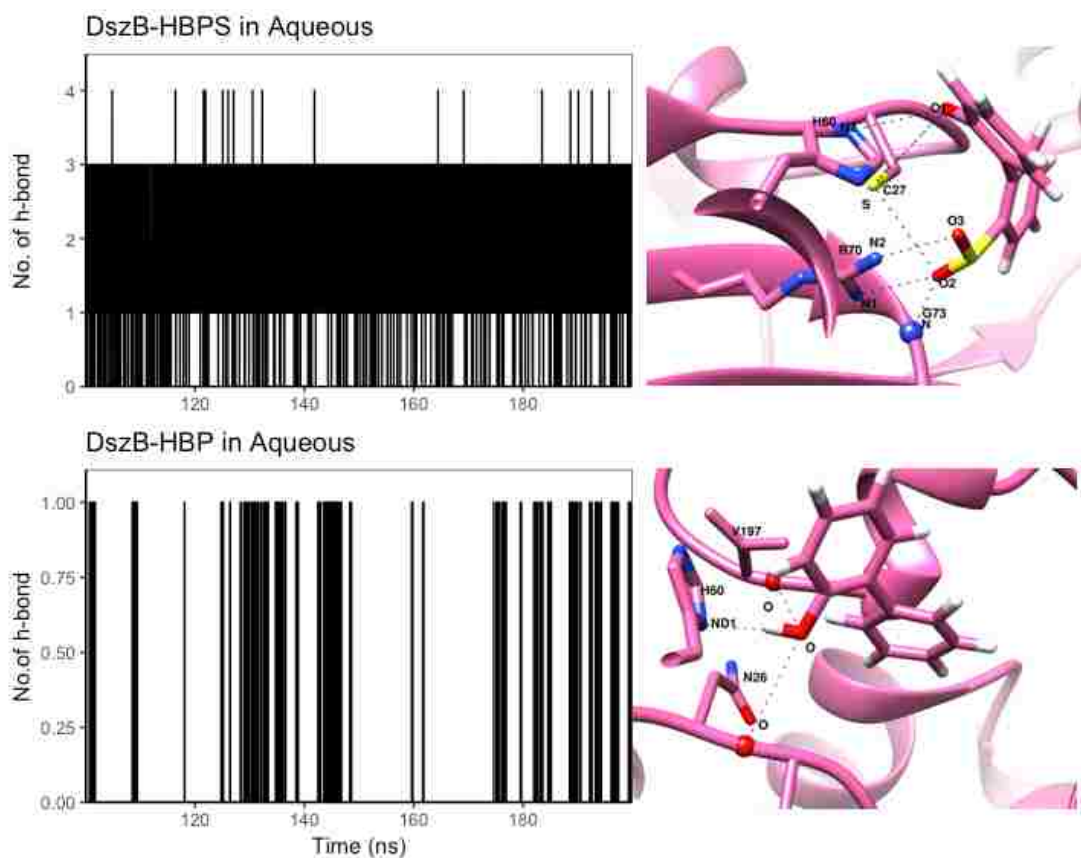


Figure 4.5 (Left) Hydrogen bonds formed along the MD trajectory between the ligand and nearby protein residues inside binding pocket of *DszB*. (Right) All possible hydrogen bonds formed during MD simulation with the ligand and nearby residues. A hydrogen bond was defined as being within 3.0 Å of a polar atom and 20° of the plane.

4.4.2 *DszB*-HBP in hexane-water biphasic solution

HBP high favorability for the hexane-formed organic phase over water enabled HBP to successfully migrate out of the binding pocket through the proposed gate region. Prather *et al.* determined that the partition function of HBP in a hexadecane oil phase and water is $P_{oil/water} = 52$, meaning HBP is about 52 times more likely to remain in the hexadecane organic phase over water [117]. Although we don't have

an experimental partition function for hexane-water, the value would be similar to that of hexadecane and water. As a result, the organic oil phase acted as external driving force to make the product HBP egress the active site. In three parallel, independent MD simulations, two simulations (MD 2nd run and MD 3rd run) showed that HBP successfully left the binding pocket through proposed gate region (Figure 4.6). When hexane molecules aggregated around the gate area, which were more exposed to HBP so that hydrophobic interactions were stronger than non-bonded interactions and h-bonding interactions, HBP migrated out of binding pocket to hydrophobic hexane-formed organic phase. However, when hexane molecules aggregated at the further distance from gate, opposite side from the gate region, HBP was not able to exit the active site; instead, it was trapped underneath the loop 3 region. Hexane aggregation positions from MD simulation are shown in Figure 4.6 in the left panel. Hexane tended to aggregate around the gate region at residues 80-90 and 190-200. However, for MD 1st run, most of the hexane molecules aggregated around residues 300-340 located at the back side of protein, so that HBP stayed inside the binding pocket buried inside loop 3 region corresponding to right panel of Figure 4.6. Therefore, HBP stability around proposed gate region adds additional hindrance to HBP from exiting the binding pocket, and when hexane molecules aggregate around the gate, the hydrophobic interactions can be strong enough to let HBP egress.

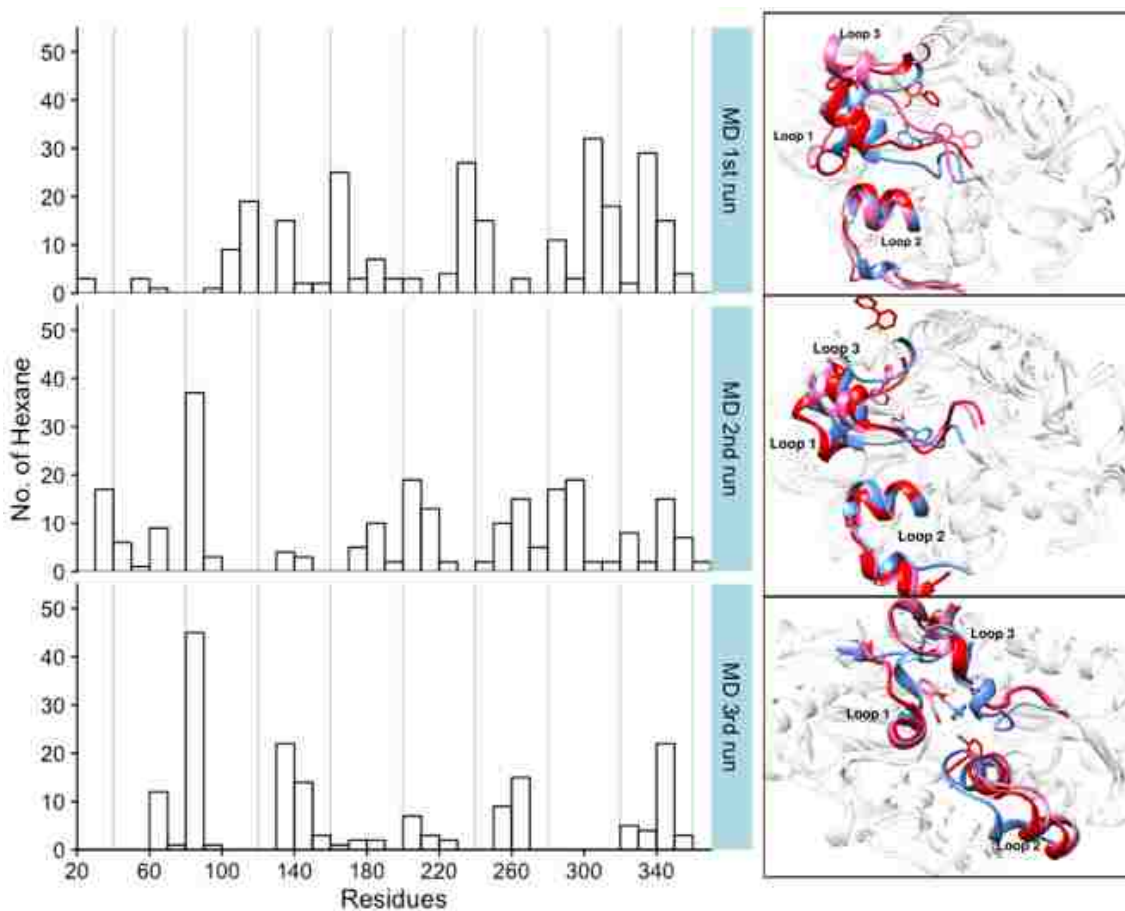


Figure 4.6 (Left) Hexane distribution around the *DszB* protein surface. The y-axis represents the number of hexane molecules within a 9Å cutoff distance from the *DszB* surface at 200 ns of each MD simulations. (Right) MD simulation snapshots of HBP position change. HBP is shown in light blue, pink, to red representing HBP in the initial to final positions, respectively, along the MD simulation. MD 2nd and 3rd runs clearly shows that HBP left the binding pocket through proposed gate region.

4.4.3 Umbrella Sampling

The energy penalty associated with gate conformational change was determined quantitatively from umbrella sampling, and it suggested that the amount of work required for the gate conformation change from closed (bound

DszB) to the open state (apo *DszB*) was 2.4 kcal/mol, which neither depended on bound ligands nor solvent conditions. Umbrella sampling MD simulation was applied, using RMSD of the gate region as the reaction coordinate, to *DszB*-HBPS, *DszB*-HBP, and apo-*DszB* in both aqueous and biphasic solution systems. The potential of mean force calculated for gate conformational change from close to open was the same, 2.4 kcal/mol (Figure 4.7). This observation is consistent with secondary structure comparisons results (Figure 4.3). That is, for all systems during MD simulation in aqueous solution, the gate region always exhibited the most dominant of the secondary structure changes. Therefore, umbrella sampling provides quantitative information of the energy cost for the three loops to undergo conformational change from close to open. The same energy penalty implies that conformational change of gate area is not a critical factor contributing to product inhibition, but rather, HBP interactions with gate area residues hinder HBP egress from binding pocket, likely contributing to product inhibition.

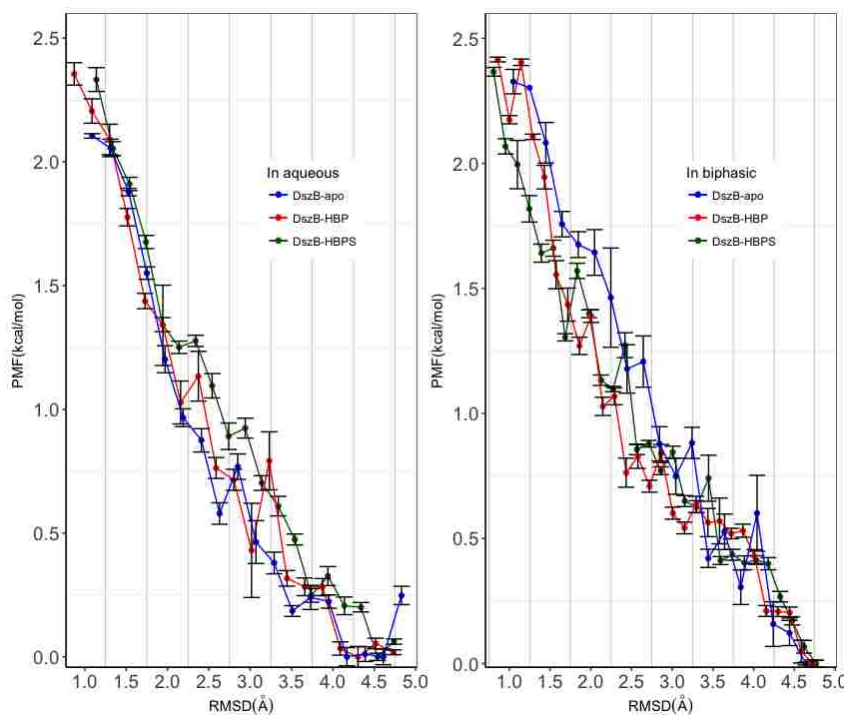


Figure 4.7 Potential of Mean Force (PMF) plots from umbrella sampling MD simulations of *DszB*-HBPS, *DszB*-HBP, and apo-*DszB* in both aqueous and hexane-water (biphasic) buffers. The energy penalty for gate conformation change from close to open corresponding to the RMSD range from 4.8 to 0.7 Å were the same regardless of condition, 2.4 kcal/mol.

4.5 Conclusions

Superimposition of the available crystal structure of *DszB*-HBPS and apo *DszB* revealed that three loops had distinct structure differences, potentially having a gate function that is responsible for ligands egress/ingress. Triplicate independent MD simulations were performed for *DszB*-HBPS and *DszB*-HBP in aqueous solutions, and HBPS showed high stability inside the active site with the catalytic residues C27, H60, and R70 remaining in close contact with HBPS. HBP, however, showed high mobility inside binding pocket, and HBP started migrating to the proposed gate

region. Non-polar residues around gate region produces non-bonded interactions and h-bonding interactions with HBP, such that HBP was not able to move out of binding pocket through the gate region. Hexane-water biphasic solution was applied to *DszB*-HBP system, and, as expected, hexane formed an organic phase that provided an external driving force based on hydrophobic interaction to “pull” HBP out of active site through proposed gate region. Taken together, our data suggests the loop-formed gate region is responsible for ligand ingress/egress out of binding pocket. Umbrella sampling MD simulations of *DszB*-HBPS, *DszB*-HBP, and apo *DszB* in both aqueous and biphasic solution systems was used to quantitatively evaluate the energy penalty associated with gate conformational change, and it was shown that the energy cost for gate conformational change from closed to open is 2.4 kcal/mol, depending on neither surrounding environment nor bound ligands. Gate conformational dynamic change is not critical for product inhibition, but rather HBP interactions with non-polar residues located at the gate area made it difficult to move out from binding pocket through gate which is the prominent factor for product inhibition.

Chapter 5 — Thermodynamic relationships between 2'-hydroxybiphenyl-2-sulfinate desulfinate and putative inhibitors

This chapter includes all previously identified molecules shown in Figure 1.2, including HBP, BIPH, NTAM, BCA, and NAPO and making use of the CGenFF parameters determined in chapter 3, to determine the binding affinities to *DszB* using free energy perturbation methods. The calculated binding free energies can be directly compared to experimentally determined equilibrium constant values from Gibbs equations.

5.1 Summary

The limitations of the biodesulfurization 4S pathway begin with *DszB*, the rate-limiting enzyme in this biocatalytic process, as it exhibits both a low turnover and experiences product inhibition. Kinetic studies performed on purified *DszB* in solution reported 2-hydroxybiphenyl (HBP), 2,2'-biphenol (BIPH), and 1,8-naphthosultam (NTAM) were inhibitors to *DszB* with inhibition constants 0.5 μM , 17 μM , and 1.8 μM , corresponding to -8.6, -6.5, and -7.9 kcal/mol binding free energies. Molecules such as 2-biphenyl carboxylic acid (BCA) and 1,8-naphthosultone (NAPO) were reported non-inhibitory yet non-productive to *DszB*. In this study, we performed molecular dynamic simulations and free energy perturbation (FEP/ λ -REMD) to determine the binding free energies of the selected molecules in *DszB*. For experimentally suggested inhibitors HBP, BIPH, and NTAM the calculated binding free energies were -9.4 ± 0.7 , -5.9 ± 0.8 , and -7.3 ± 0.5 kcal/mol, in excellent agreement with experimental values ($\sim \pm 0.6$ kcal/mol). However, for non-inhibitory molecules, BCA and NAPO, there was a significant discrepancy (>10 kcal/mol) between the theoretical

and experimental values due to high electrostatic interactions with the positively charged R70. Evaluation of the protonation state of the arginine residue in the active site (R70) with the corresponding binding free energy calculations indicated the possibility of a deprotonated R70, which is uncommon in biological molecules.

5.2 Introduction

Liquid petroleum products represent a large majority of fuels consumed in the transportation sector and are derived from sulfur-containing crude oils [124]. During the refinement process, crude oil sulfurs are passed to the fuel due to insufficient purification processes, where they are then combusted to form sulfur oxides (SO_x). These SO_x compounds are classified as acutely toxic under the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS) and pose a variety of concerns relating both to the environment and human health including respiratory irritation and acid deposition (acid rain) [125].

Biodesulfurization via enzyme catalysis has the potential for highly specific, rapid thiophenic desulfurization, occurring at ambient temperature and pressure [12]. These mild operating conditions, coupled with the fact the heating value of the fuel is left virtually undisturbed, collectively make biodesulfurization an advantageous additive process from both an initial capital and operational cost stand point. The overall mechanistic approach by which microbial organisms may access sulfur bonds in aromatic molecules can be broken down into two categories: ring destruction and sulfur specific [18, 27]. In the former, as the name implies, the ring is destroyed for ease of access to the sulfur. In the latter, the ring is opened, disturbing only bonds shared with the target sulfur, as in the 4S pathway depicted in

Figure 1.1. Given that the “ring destruction” approach potentially disturbs carbon-carbon bonds from which most of a fuel’s heating value is derived, interest in biodesulfurization is focused on ring opening mechanisms that break as few bonds as possible to achieve sulfur removal [18, 116].

Limitations of the 4S pathway begin with *DszB*, the rate-limiting enzyme in this biocatalytic process, as it exhibits both a low turnover and experiences product inhibition, the inhibitory effect of the presence of product HBP. Kinetic studies performed on purified *DszB* in solution indicate the possibility of product competitive inhibition. Besides HBP, Watkins *et. al.* suggested that several different molecules, HBP analogs, with varying functional group attachments, also inhibited *DszB* over a wide range of inhibition constants (K_i). Within the two-functional group class, the most inhibitory of the tested analogs was 2’2-biphenol ($K_i = 17 \mu\text{M}$). In the planar naphthenic class, the most strongly inhibiting compound was 1,8-naphthosultam ($K_i = 1.8 \mu\text{M}$); however, 2-biphenyl carboxylic acid and 1,8-naphthosultone were non-inhibitory, yet not productive (Figure 1.2) [20]. Thus far, there relatively little molecular-level analysis on how these putative inhibitors bind inside *DszB*. Moreover, characterization of the inhibition mechanism of these inhibitors, i.e., allosteric or competitive binding, has not been reported

In this study, we employed a computational approach towards determining absolute binding free energies for the selected molecules shown in Figure 1.2 using the free energy perturbation with Hamiltonian replica exchange MD (FEP/ λ -REMD) method and compared the free energy values with experimentally reported dissociation constants. We performed molecular dynamics (MD) simulations of *DszB*

bound with the putative inhibitors to provide insights into how the various functional groups of the selected ligands contribute to binding with *DszB*, as well as the active site dynamic contributions to ligand binding stability. The free energy changes compared with the existing experimental values provides information on how the thermodynamic signatures are related to enzyme functionality.

5.3 Method

5.3.1 Molecular dynamic simulations for putative inhibitors bound to *DszB*

MD simulations were used to examine the selected potential inhibitors bound in the active site of *DszB*. The *DszB*- complex systems were constructed based on the availability of crystal structure *DszB*-HBPS with Protein Data Bank (PDB) code 2DE3 [39]. We manually deleted the bound substrate, HBPS, and docked the selected molecules, BIPH, BCA, NAPO, and NTAM, into the binding pocket based on standard affinity-based docking calculations by AutoDock version 4.2.6 [126]. The docked binding positions for the selected inhibitors is shown in Figure 5.1. For HBP, we used the original HBPS position but deleted sulfinate group, representative of the catalytic process. The *DszB* bound system was solvated in a $90 \text{ \AA} \times 90 \text{ \AA} \times 90 \text{ \AA}$ cubic box filled with TIP3P water molecules [127, 128]. The CHARMM 36 force field was applied to model the protein [129-131], combined with the optimized CHARMM general force field (CGenFF)-based force field parameters [89] for the HBPS analogues from our previous work (chapter 3) [35]. Protonation states of all the titratable residues were calculated using H++ web server [118] at the optimal pH=7.4 [20]. Additionally, 0.3M NaCl and 50mM HPO_4^{2-} were put in the periodic box

to mimic experimental media, and extra Na⁺ ions were added into system to neutralize the charge (*DszB* is -17e⁻ and HBPS is 1e⁻).

The solvated system went through minimization and equilibration before MD production runs. Each system was minimized using CHARMM with 2000 steps of Steepest Descent (SD) algorithm with the protein and ligand restrained, following with 2000 SD and 5000 steps adopted basis Newton-Raphson (ABNR) algorithm to the whole system. The minimized systems were then equilibrated by slowly heating the system from 90 K to 300 K with 50 K increments over 100 ps in the canonical ensemble. This was followed by isothermal-isobaric (*NPT*) simulation using Nosé-Hoover thermostat and barostat in CHARMM for 200 ps at 1 atm to equilibrate systems density [47, 132, 133]. The density-equilibrated systems were finally gone through MD production for 200 ns using NAMD [134] in the *NVT* ensemble with a 2-fs time step at 300 K. Temperature was controlled using the Langevin thermostat in NAMD.

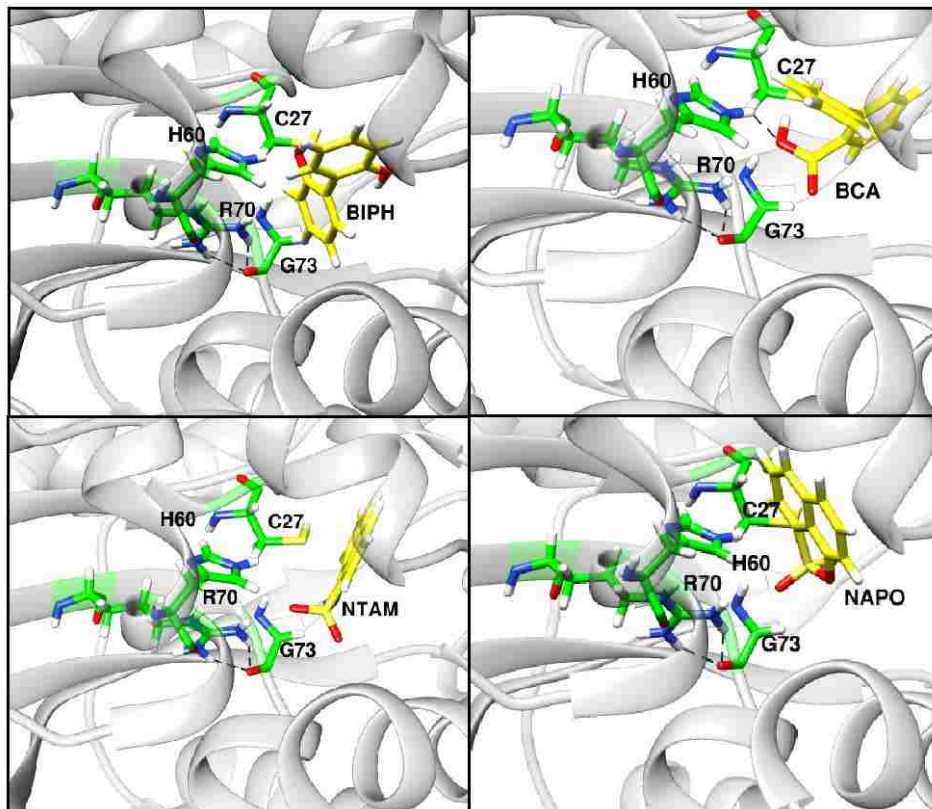


Figure 5.1 Docking calculations predicted binding positions for selected inhibitors (BIPH, BCA, NTAM, and NAPO) inside binding pocket of *DszB*. Catalytic residues, C27, H60, R70, and G73, are highlighted in green stick representations, black dashed lines represent the potential hydrogen-bonding interactions between bound ligand and *DszB*.

5.3.2 Free energy calculation FEP/ λ -REMD

For quantitative examination of thermodynamic preference of bound ligands to *DszB*, we performed absolute binding free energy calculations using FEP/ λ -REMD, an enhanced sampling free energy calculation method [76]. The selected inhibitors absolute binding free energies were calculated with the aim to compare with experimental values.

The FEP/ λ -REMD calculation method couples free energy perturbation with Hamiltonian replica-exchange molecular dynamics to enhance Boltzmann sampling [64]. Insertion ligand into binding pocket was split into two thermodynamic paths (Figure 2.2), with the assistance of three thermodynamic coupling parameters, λ_{repu} , λ_{disp} , and λ_{elec} , to control the non-bonded interactions of the molecule with its surrounding environment, and an additional parameter, λ_{rstr} , was used to control the translational and orientational restraints. The contributions to overall free energy includes the shifted Weeks-Chandler Anderson repulsive and dispersive components, ΔG_{repu} and ΔG_{disp} , electrostatic contribution, ΔG_{elec} , as well as the additional restraining component, ΔG_{rstr} which hold the center of mass of DszB and bound ligand distance unchanged [76]. The thermodynamic paths include two separate calculations: (1) decoupling the bound ligand from *DszB*; (2) decoupling the solvated ligand from solution. The binding free energy, ΔG_{b} , will be the difference of the two. The restraint potential was only applied to the first thermodynamic cycle.

The free energy calculations were performed for 3 ns with 30 consecutive windows, each of which contains 0.1 ns simulations for both thermodynamic paths. The simulations used a set of 128 replicas (72 repulsive, 24 dispersive, and 32 electrostatic) with the exchange frequency of every 0.1 ps. The complex systems involved a positional restraint by the distance of center of mass between ligand and protein, and this restraint was determined from numerical integration with Simpson's rule [64]. Additionally, molecules inside the binding pocket were exposed to an additional harmonic restraint applied on root mean square deviation (RMSD)

values of the aromatic carbons. This additional restraint was calculated by performing separate MD simulations removing the harmonic restraint on the RMSD for aromatic carbons over 5 ns, and Multistate Bennett Acceptance Ratio (MBAR) method was used to determine the potential of the additional restraints [135]. Individual repulsive, dispersive, and electrostatic energy contributions as well as statistical uncertainties were determined based on the output energies from the replica exchange simulations using MBAR for both thermodynamic paths (Figure 2.2) [135]. Lastly, summing up the three energies (repulsive, dispersive, and electrostatic) for both paths and adding the additional restraint potentials to the path 1 resulted in the total energies of each path (ΔG_1 and ΔG_2). The differences of the two revealed the binding free energy of the *DszB*-ligand complex. The standard deviation over the last 1 ns (the converged data) was used to determine as the binding free energy error. Convergence was determined by monitoring the evolution of the free energy calculations of the two thermodynamic paths over the time (Figure S.5). Experimental K_i values were put in terms of free energy using the Gibbs equation (Eq.5.1). Comparisons with experimental values are listed in Table 5.1.

$$\Delta G = RT \ln(k_d/c) \quad 5.1$$

ΔG is the binding free energy, R is the ideal gas constant, k_d is the dissociation constant which is equivalent to K_i , and c is the standard reference with units of μM .

The binding free energy of deprotonated R70 in *DszB* was investigated for the two non-inhibitory molecules, *DszB*-BCA, and *DszB*-NAPO. Deprotonation of R70 on guanidinium N- η 1 position was performed based on the observation from Geronimo

et. al. that structure deprotonation on N- η 1 was 5 kcal/mol more stable than deprotonated at N- ϵ position [42]. Binding free energy of BCA and NAPO were calculated in deprotonated R70 states based on CHARMM force field for neutral arginine, the results were reported in Table 5.2.

5.4 Results and discussions

5.4.1 Putative inhibitors binding inside *DszB*

Selected molecules binding positions in *DszB* were estimated using AutoDock, and it was shown that selected molecule binding positions were all buried inside binding pocket of *DszB* with no other binding position outside the macromolecule. During MD simulation, the inhibitory compounds, BIPH, remained stable inside the substrate binding site, as compared to substrate HBPS (Figure 5.2). Product HBP, however, started migrating towards the proposed gate area, as described in chapter 4 (Figure 4.2). BIPH and NTAM, if they bind to *DszB* from solution, are then likely to compete with HBPS for the binding site, as in a competitive binding mechanism. Molecule BIPH stayed stable inside the binding site; however, planar molecule NTAM showed an orientation change within the binding site, turning slightly perpendicular compared to the original position to take advantage of aromatic stacking. Aromatic residues were highlighted in Figure 5.2. HBP was previously reported to exhibit high flexibility inside binding pocket, but due to prominent interactions around the proposed gate region, HBP exhibited stability within the gate area, suggestive of allosteric competitive inhibition to *DszB*.

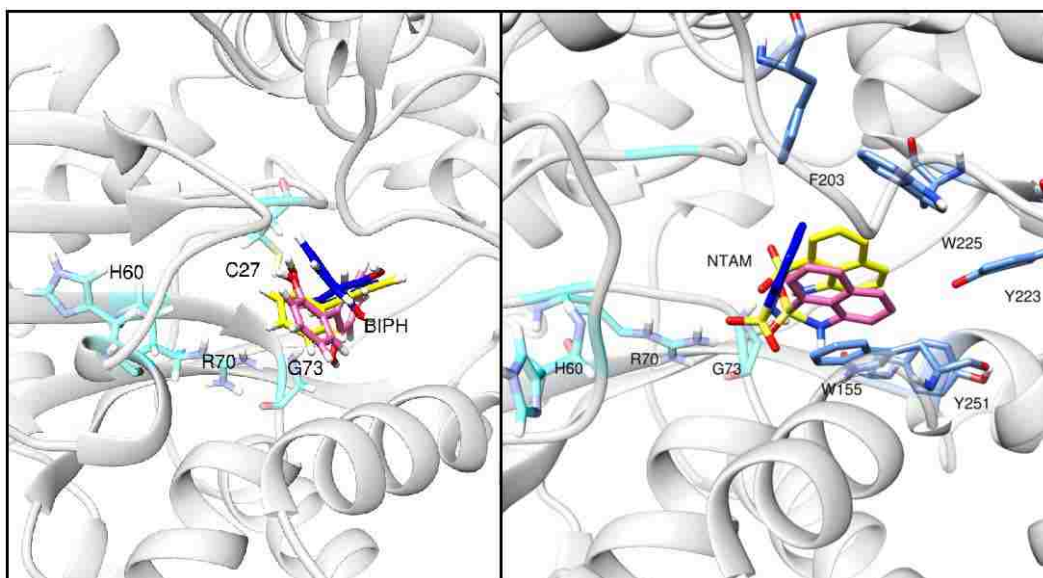


Figure 5.2 Putative inhibitor binding positions inside *DszB*. Catalytic residues are shown in cyan stick with labels. The NTAM binding position is illustrated with aromatic residues highlighted in light blue stick representations. Bound molecules are shown in stick representations with blue, pink, to yellow colors corresponding to ligand position changes at the initial, middle, and final snapshot over the course of 200-ns MD simulations.

Reported non-inhibitory molecules, BCA and NAPO, binding positions during the course of 200-ns MD simulations are shown in Figure 5.3. Interestingly, BCA exhibited high stability inside binding pocket, whereas NAPO showed high flexibility (Figure 5.3). The BCA carboxylic group had a large interaction with net positively charged R70 (-20 kcal/mol non-bonded interaction energy, Figure S.7), which potentially formed a salt bridge that kept BCA very stable inside binding site. Planar molecule NAPO, due to lack of hydrogen-bond donor, exhibited high flexibility and resulted in large protein backbone fluctuations, especially around the proposed gate region, which potentially caused the gate to start unfolding. Therefore, NAPO

binding was unfavorable inside the *DszB* binding site, consistent with non-inhibitory observations from experiment

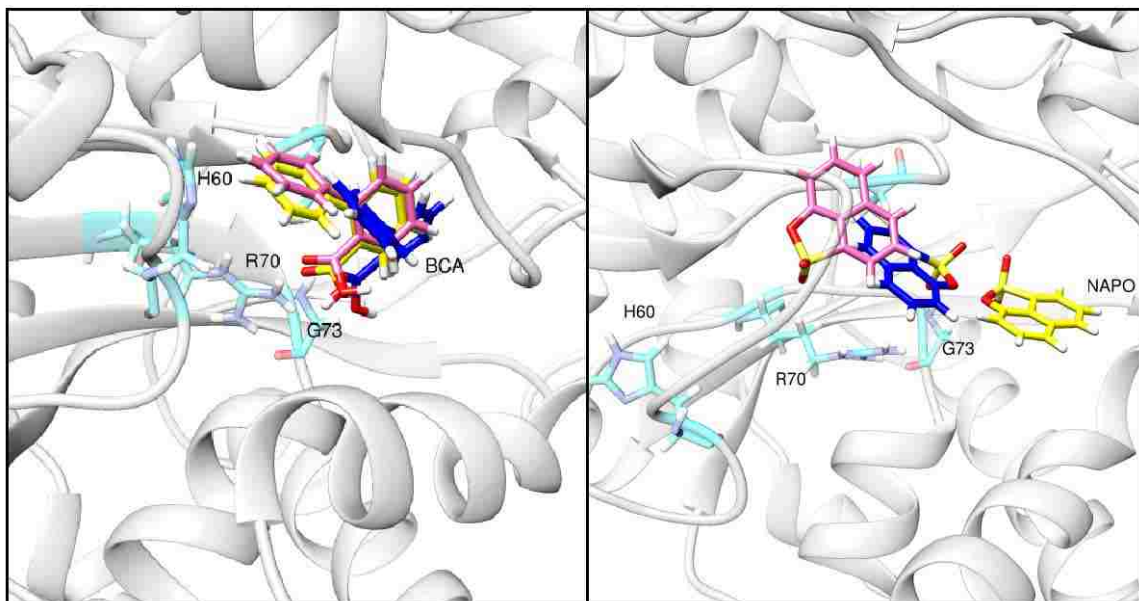


Figure 5.3 Non-inhibitory molecule bound positions inside *DszB*. Catalytic residues are represented in light blue stick with labels; bound ligands are represented in stick representation with blue, pink, and yellow carbons representing initial, middle, and final snapshot of 200-ns MD simulation, respectively.

This high flexibility of NAPO resulted in large protein backbone fluctuation around the proposed gate region, illustrated by root mean square fluctuations (RMSF) values of protein backbone over the course of 200-ns MD simulations. RMSF captured residues 50-60 (corresponding to gate loop 1) and residue 180-200 (corresponding to loop 3) in a dramatic loop dynamic displacement of NAPO. We observed that when NAPO was bound inside the pocket, it moved randomly inside the pocket and resulted in gate area loop 1 and loop 3 starting to change to extended loop structures and exhibited large dynamic fluctuations. However, when BCA was

bound inside, protein backbone fluctuation was prominently reduced, and BCA stayed stable inside binding pocket. The catalytic residues R70, C27, G73 kept close interactions with BCA, especially R70, which potentially formed a salt bridge with carboxylic group of BCA. The high stability of BCA bound inside *DszB* seems to contradict the experimental non-inhibitory observations and may influence the binding free energy value that calculated below.

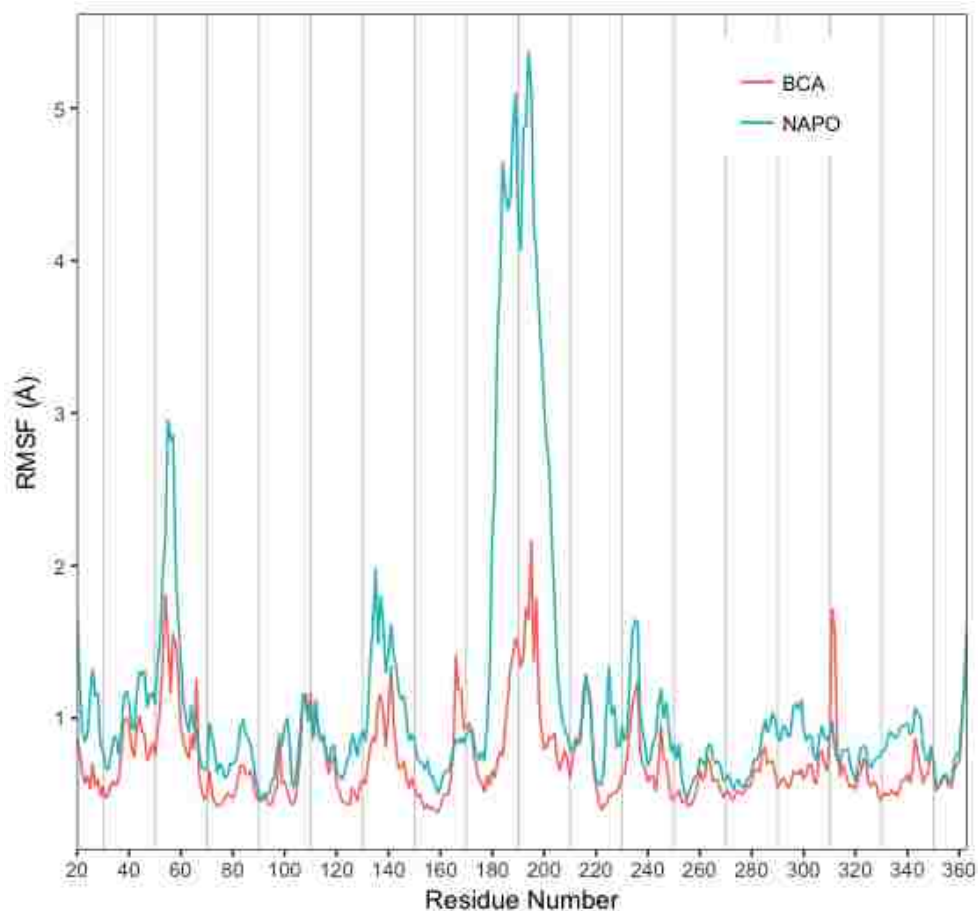


Figure 5.4 Root mean square fluctuations from MD simulation revealed the non-inhibitory molecules bound inside binding pocket resulted in different conformational dynamic changes of protein. Red and blue lines represent BCA and NAPO bound in *DszB* respectively.

5.4.2 Free energy calculations from FEP/ λ -REMD compared with experimentally determined parameters

Binding affinities of putative inhibitors bound inside *DszB* can be determined from binding free energies calculations. The binding free energies values are provided alongside the corresponding repulsive, dispersive, electrostatic, and restraint contributions to the binding free energies. The solvation energies for each ligand are also reported in Table 5.1. Errors associated with each component represent the standard deviation of over the last 1ns of collected data, the error for binding free energy was computed by taking the square root of the sum of the squared standard deviations of the free energy needed to decouple ligand from the enzyme (ΔG_1) and ligand solvation free energy (ΔG_2). The binding free energy as a function of time for convergence validation is provided in Figure S.5.

Table 5.1 Binding free energies calculated from FEP/ λ -REMD represented by ΔG_b° at 300K compared with experimental results ΔG_b^{exp} converted from Watkin et. al. suggested inhibition constants [20].

	ΔG_b° *	ΔG_b^{exp} *	ΔG_{repu} *	ΔG_{disp} *	ΔG_{elec} *	ΔG_{rstr} *
HBP	---	---	22.6 \pm 0.2	-21.3 \pm 0.04	-7.9 \pm 0.1	--
<i>DszB</i> -HBP	-9.2 \pm 0.7	-8.6	20.5 \pm 0.5	-27.7 \pm 0.2	-8.6 \pm 0.2	-0.2
NTAM	--	---	21.8 \pm 0.1	-24.5 \pm 0.04	-13.6 \pm 0.04	--
<i>DszB</i> -NTAM	-7.3 \pm 0.5	-7.9	27.5 \pm 0.3	-34.8 \pm 0.2	-16.1 \pm 0.2	0.3
BIPH	---	---	23.3 \pm 0.1	-22.2 \pm 0.04	-6.6 \pm 0.1	---
<i>DszB</i> -BIPH	-5.8 \pm 0.8	-6.5	20.7 \pm 0.4	-27.9 \pm 0.09	-6.7 \pm 0.2	2.6
BCA	---	---	24.2 \pm 0.1	-22.7 \pm 0.05	-11.4 \pm 0.2	---
^a <i>DszB</i> -BCA	-15.6 \pm 0.9	^b N/A	18.1 \pm 0.5	-30.9 \pm 0.08	-14.1 \pm 0.5	1.4
NAPO	---	---	21.5 \pm 0.1	-23.7 \pm 0.04	-11.9 \pm 0.03	---
<i>DszB</i> -NAPO	-4.4 \pm 0.4	^b N/A	19.8 \pm 0.4	-32.4 \pm 0.2	-7.2 \pm 0.2	1.3

* kcal/mol

^a Exhibited prominent discrepancy with experimental results

^b Experimental reported non-inhibitory molecules

HBP, NTAM, BIPH, BCA, and NAPO, with decreasing inhibitory effects, exhibit K_I values of 0.5 μ M, 1.8 μ M, 17 μ M, and two non-inhibitors, respectively; binding free energy calculated from the inhibition constants is given as ΔG_b^{exp} in Table 5.1.. The calculated absolute binding free energies ΔG_b° matched experimentally suggested values well, within 0.6 kcal/mol, except for non-inhibitory molecule BCA, which showed abnormally strong favorability for binding. The strong binding was also observed with active-site dynamic behavior; BCA showed prominent stability

during the course of MD simulation. The strong interaction between carboxylic group and positively charged R70 kept BCA stable very stable and may possibly lead to abnormally large binding affinity. The possibility of a deprotonated state of R70 was investigated below. The binding free energy of HBPS is not reported here. Finite-size effects significantly affect the accuracy of the FEP/ λ -REMD method for charged ligands in binding free energy calculation [136-139].

5.4.3 Protonation state of R70

Catalytic residues C27, H60, R70, and G73 are highly conserved among homologs of *DszB*. According to Lee, *et. al.*, homologs of *DszB* were selected based on the protein sequence from NCBI database, all catalytic residues, including C27, H60, R70 and G73, were highly conserved, and they also suggested that G73, main chain nitrogen and carboxylic group were required for R70 guanidinium group stability [39]. Hedstrom *et. al.* suggested that all reported neutral arginine acted as a general base in enzyme, although their common structure motif is where arginine is solvent accessible and adjacent to carboxylate groups [140]. In our situation, binding induced conformational changes around gate area were reported independent of bound ligands; therefore, R70 is solvent accessible despite being deep within the binding pocket. Moreover, BCA bound inside introduced an additional carboxylate group combined with G73, favoring the possibility of a deprotonated R70.



Figure 5.5 Multiple sequence alignment of *DszB* homologs. Columns color representation from dark blue to white represents the most to least sequence identity. Catalytic residues C27, H60, R70, and G73 are highlighted with star symbol below. Aligned protein includes *BdsB*, homolog of *DszB* from *Bacillus subtilis* (NCBI accession number BAC20181); *TdsB* homolog of *DszB* from *Paenibacillus* sp. A11-2 (NCBI accession number BAA94832); *Atu3432*, putative protein from *Agrobacterium tumefaciens* Str. C58 (NCBI accession number NP_533929); *SMa2087*, ABC transporter substrate-binding protein from *Acinetobacter* (NCBI accession number WP_004925281). This figure was recreated based on Lee *et. al.*[39].

The binding free energy was investigated for the two non-inhibitory molecules below, BCA and NAPO, were reported in Table 5.2. Deprotonation of R70 led to decreased binding affinity of BCA for *DszB*. The decreased affinity, -5.3 kcal/mol, corresponds to a dissociation constant of 129.8 μ M. Even though an experimental suggested K_I of BCA was unavailable to compare to, the dramatic increased (10 times larger than inhibitory compounds reported K_I) dissociation constant of BCA suggested a decreased inhibitory of BCA consistent with experimental observations. NAPO exhibited a large positive binding free energy indicating the unfavorable binding inside *DszB*. Moreover, the deprotonated systems went through 200-ns MD simulation, and deprotonation of R70 did not perturb the protein structure along the trajectory (simulation snapshots are provided in Figure S.8). Therefore, we

suggest it is quite likely that R70 is deprotonated and deserves further experimental investigations.

Table 5.2 Binding free energy of deprotonated R70 for non-inhibitory molecules.

	ΔG_b^*	ΔG_{repu}^*	ΔG_{disp}^*	ΔG_{elec}^*	ΔG_{rstr}^*
BCA	---	24.2±0.1	-22.7±0.05	-11.4±0.1	--
<i>DszB</i> -BCA	-5.3±0.7	20.5±0.4	-31.7±0.1	-6.3±0.1	2.2
NAPO	--	21.5±0.1	-23.7±0.04	-11.9±0.04	--
<i>DszB</i> -NAPO	23.3±0.9	45.1±0.9	-30.5±0.1	-8.6±0.1	3.3

* kcal/mol

5.5 Conclusions

Binding affinities were determined using FEP/ λ -REMD on selected molecules. According to 200ns-MD simulation results, NTAM and BIPH exhibited competitive inhibitions, seemingly binding at the same binding site compared to substrate HBPS. HBP, on the other hand, bound at a different position during the simulation, at the gate region, suggestive of allosteric competitive inhibitory. The calculated binding free energies matched the experimentally reported values within error, except for BCA. The two non-inhibitory molecules lack of hydrogen-bond donors and an extra carboxylate group raises the possibility of deprotonation on R70. Binding free energies were evaluated based on the deprotonation of R70, and the results showed less favorable binding affinities that coincide with experimental observations. Thus, the possibility that R70 is deprotonated based on different bound ligands need to be further investigated experimentally

Chapter 6 — Conclusions and future directions

Biodesulfurization via the 4S enzymatic pathway has the potential for highly specific and rapid thiophenic organosulfur removal, such as dibenzothiophene (DBT), occurring at ambient temperature and pressure with oil heating value unchanged. It seems to be a perfect downstream implementation to the current existing hydrodesulfurization method. However, improving the overall rate of enzymes that desulfurize thiophenic molecules remains as a big challenge. We applied a computational MD simulation approach and attempted to provide insight into the role of each residue within active site of *DszB* based on substrate and inhibitor binding. We also examined different putative inhibitors bound inside *DszB* qualitatively by active-site dynamic comparisons and quantitatively by their thermodynamic preferences that aimed to provide a molecular-level analysis on inhibitory mechanism of *DszB* and capture more foundational insights to assist rational improvement of desulfinase activity.

In chapter 3, I described a general procedure on CHARMM general Force Field (CGenFF) parameters development for a subset of chemically relevant aromatic molecules. The selections of molecules covered single, double, and planar functional group representatives of HBP analogs with existing experimental reported inhibition constants. This chapter implemented an iterative force field optimization procedure that was clearly presented in Figure 3.1. The optimized parameters were validated using a self-consistent method of comparing calculated MM geometries to the optimized molecular geometries through recapitulation of experimental infrared spectra (IR). Force field parameters are among the most important parameters for

MD simulation. With accurate force field parameters, we will be able to develop molecular models to describe substrate binding and inhibition phenomena in *DszB*. Chapter 3 not only provided important parameters for future simulations, but also summarized Mayne, *et al.* suggested procedure for small molecules force field development [91], and developed a novel validation method. However, we chose IR validation process which is experimentally easier to interpret and generate, but this method is associated with big discrepancies compared to theoretical calculated IR spectra at high frequencies. As discussed in chapter 3, section 3.3.4, the largest mismatch was from O-H bond stretches that theoretical calculation needs to account for anharmonicity effects resulting from high vibrational region of bond stretch. For simplicity, the theoretically calculated IR did not take the anharmonicity effects into account and modeled the system as harmonic, which breaks down the high frequency region results. For future approaches, it is true that implementing anharmonic algorithms for calculating IR is non-trivial, yet it is recommended that we could develop error estimations that evaluate the anharmonic effects based on molecule structure similarity so that the inaccuracies can be reduced based on empirical correction constants. We anticipate that our work on summarizing Mayne, *et al.* outlined procedures [91] and developing novel method for validation process on generated force field parameters will simplify the procedure and motivate more force field developments on small drug-like molecules to expand the existing CHARMM force field database

In Chapter 4, we applied the generated force field to substrate HBPS and product HBP in order to perform MD simulations. Based on the availability of crystal

structure of holo *DszB*-HBPS and apo *DszB*, we discovered prominent conformational differences based on the two structures alignment, and therefore, proposed a gate region that was responsible for ligand egress/ingress. Due to lacking crystal structure availability of *DszB*-HBP, we directly used HBPS binding position but deleted sulfinate group from HBPS and converted to HBP. During the course of 200-ns MD simulation, we observed that HBP migrated towards the gate area and showed more stable binding at the gate region. It characterized HBP as allosteric competitive inhibitor, and it did not compete the exact same binding site with HBPS. Novel biphasic MD simulation method was applied to *DszB*-HBP system in order to observe HBP exiting binding pocket, which is assisted by strong hydrophobic interaction driving force. Fortunately, this novel method allowed us to observe HBP migrating out of the binding pocket through the proposed gate area. Interestingly, umbrella sampling provided a free energy value of gate conformational change from close to open of 2.4 kcal/mol. This value was independent of bound ligands and solvent conditions. This chapter provided insights on product inhibition of *DszB* and identified HBP as a putative allosteric competitive inhibitor based on HBP stability around gate region rather than the original binding site due to high non-bonded interactions around gate. We found there was 53% non-polar residues located at the gate (Figure 6.1(A)), and the most dominant ones are A and V. These non-polar residues provided VDW interactions shown in Figure 6.1(B), provided from gate region. Moreover, this chapter also identified that the dynamics of loop regions is not essential for product inhibition due to same energy penalty values in terms of bound ligands and solvent conditions,

the clear observations of HBP strong interactions with gate residue suggested that product inhibition is related to the gate regions interactions with HBP. It is recommended that for future work, mutation focus on gate area non-polar residues (for example, A188T and F61N) in order to decrease the non-bonded interactions between HBP and gate residues, which may potentially decrease the binding affinity between HBP and *DszB* and, therefore, decrease the product inhibition phenomena

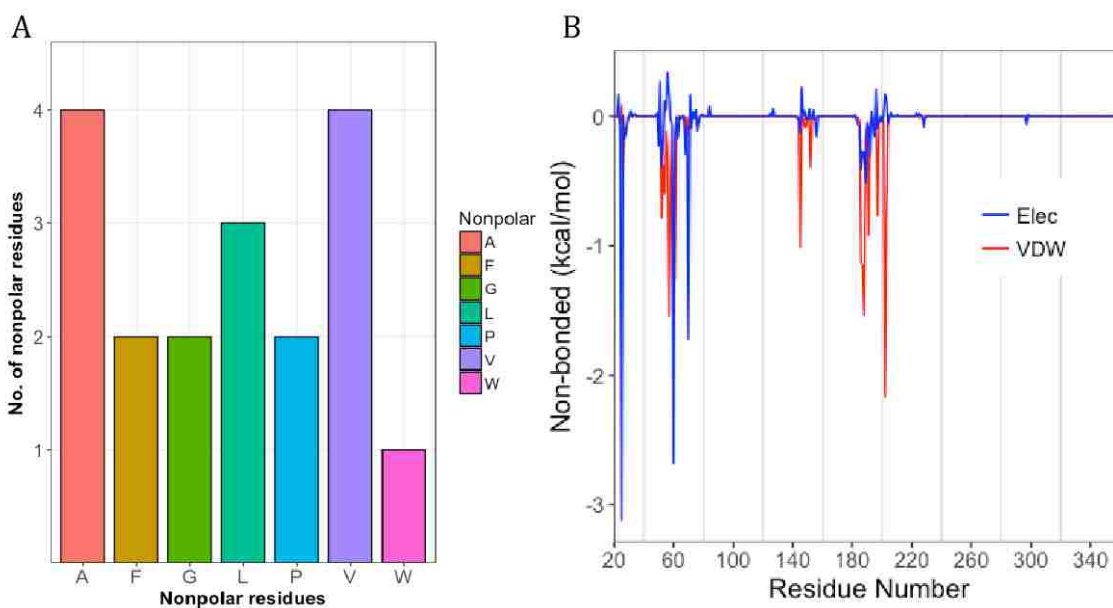


Figure 6.1 (A) Number of non-polar residues located at the gate regions. (B) HBP non-bonded interaction along the course of 200-ns MD simulations.

In chapter 5, we evaluated thermodynamic preferences of bound ligands, which were extended by four more analogs with single, double, and planar functional groups representatives that had force field parameters optimized. We attempted predicting the selected putative inhibitors binding positions inside *DszB*, and performed 200-ns MD simulation each. We observed that for experimental inhibitory molecules, NTAM and BIPH, they exhibited high stability inside the same

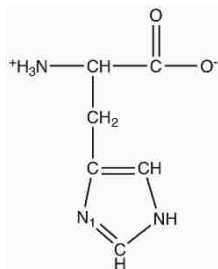
binding site compared to HBPS, which indicated the possibility of competitive inhibition, whereas HBP exhibited allosteric competitive inhibition. The binding affinities calculated from FEP/ λ -REMD matched experimental suggested values with errors less than 0.6 kcal/mol, except for BCA. During MD simulations, BCA showed high stability at the binding site, and exhibited strong interactions with R70 which potentially caused the abnormal binding free energies calculated from FEP/ λ -REMD. *DszB* homolog analysis suggested that R70 and G73 are highly conserved. Therefore, the carboxylate group of G73 was inevitably exposed to R70. When BCA bound inside active site, an additional carboxylic group is exposed to R70 that added more chances of deprotonation of R70. NAPO, similarly, lacks hydrogen-bond donors; therefore, deprotonation state of R70 was evaluated for the two non-inhibitory molecules (BCA and NAPO). It was determined that deprotonation of R70 reduced binding affinity of BCA to -5.3 ± 0.7 which is more consistent with experimental observations. NAPO exhibited positive binding free energy, which indicated NAPO was unfavorable of binding inside *DszB*, consistent with experimental observations. However, we did not report binding free energy for charged HBPS molecule due to finite-size effect of FEP/ λ -REMD on charged ligand that strongly influenced the accuracy of results. There were several schemes reporting the ability to solve the inaccuracy issues by implementing Poisson-Boltzmann calculations on the protein-ligand system. However, due to different ion concentrations added to the solvent, determining the dielectric constant for our solvent is non-trivial. Therefore, development of a correction scheme is definitely worth trying for future approaches to correct the calculated binding free energy from FEP/ λ -REMD for charged HBPS.

Moreover, even though deprotonation states resulted in matched experimental results for non-inhibitory molecules, further structure investigation on deprotonation state of R70 is still needed to be conducted experimentally. For example, evaluating the protonation state of R70 structure by obtaining NMR structures.

In summary, this dissertation applies the computational approaches to gain molecular-level understanding of the product inhibition mechanism of *DszB*. We uncovered a large conformational change upon ligand binding and established that this conformational change functions as a gate to slow the product from exiting the binding pocket. Moreover, we suggested that the product HBP exhibited allosteric competitive inhibition that influences the biodesulfurization last step reaction rate. The thermodynamic characterization of putative inhibitors evaluated the connections between experimental reported inhibition constants and theoretical obtained binding free energy values. This dissertation completed a computational approach cycle from how to obtain computational parameters (force field development) to applying the theoretical generated results to reality and matched the experimental observations.

Appendix

1 Summary of performed MD simulations

Inhibitory molecules				
<i>DszB</i> -	HBPS	HBP	BIPH	NTAM
Crystal Structure	2DE3	2DE3	2DE3 AutoDock	2DE3 AutoDock
Periodic boundary box size (<i>a, b, c</i>)	90Å	90Å	90Å	90Å
Atoms	69182	69169	74271	74343
Sodium ions	194	193	193	193
Chloride ions	132	132	132	132
Phosphate (HPO ₃ ²⁻)	22	22	22	22
*Protonated residues	H60	H60	H60	H60
Non-inhibitory				
<i>DszB</i> -	BCA	NAPO	 <p style="text-align: center;">H60</p>	
Crystal Structure	2DE3 AutoDock	2DE3 AutoDock		
Periodic boundary box size	90Å	90Å		
Atoms	74272	74252		
Sodium ions	193	193		
Chloride ions	132	132		
Phosphate (HPO ₃ ²⁻)	22	22		
*Protonated residues	H60	H60		

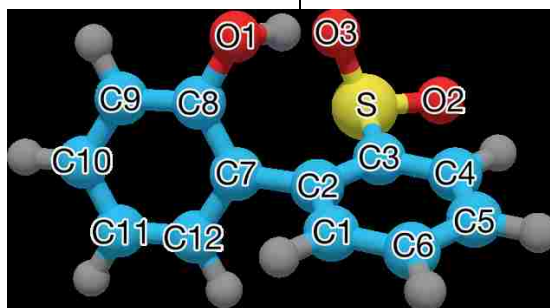
*N₁ position was deprotonated

2 Additional information for chapter 3

2.1 Force field parameters including charges, bonds, angles, and dihedrals for selected molecules

HBPS**Charges**

Atom	Atom type	Charges	Atom	Atom type	Charges
C1	CG2R61	-0.627273	H1	HGR61	0.115000
C2	CG2R67	-0.012273	H2	HGR61	0.115000
C3	CG2R61	0.320727	H3	HGR61	0.115000
C4	CG2R61	-0.115000	H4	HGR61	0.115000
C5	CG2R61	-0.115000	H5	HGR61	0.115000
C6	CG2R61	-0.114000	H6	HGR61	0.115000
C7	CG2R67	0.494727	H7	HGR61	0.115000
C8	CG2R61	-0.081273	H8	HGR61	0.115000
C9	CG2R61	-0.116000	O1	OG311	-0.599273
C10	CG2R61	-0.112000	H9	HGP1	0.297727
C11	CG2R61	-0.114000	S	SG302	0.352723
C12	CG2R61	0.022727	O2	OG2P1	-0.701273
			O3	OG2P1	-0.701273



BONDS

Atom type	Atom type	K_b	b_0
CG2R61	CG2R61	305.000	1.375
CG2R61	CG2R67	305.000	1.375
CG2R61	OG311	334.300	1.411
CG2R61	SG302	190.000	1.730
CG2R61	HGR61	340.000	1.080
CG2R67	CG2R67	300.000	1.490
OG2P1	SG302	630.000	1.440
OG311	HGP1	545.000	0.960

Angles

Atom type	Atom type	Atom type	K_θ	θ_0
CG2R61	CG2R61	CG2R61	40.000	120.000
CG2R61	CG2R61	CG2R67	40.000	120.000
CG2R61	CG2R61	OG311	45.200	120.000
CG2R61	CG2R61	SG302	35.000	119.000
CG2R61	CG2R61	HGR61	30.000	120.000
CG2R67	CG2R61	OG311	85.005	118.711
CG2R67	CG2R67	SG302	15.123	112.738
CG2R61	CG2R67	CG2R61	40.000	120.000
CG2R61	CG2R67	CG2R67	40.000	120.000
CG2R61	OG311	HGP1	65.000	108.000

CG2R61	SG302	OG2P1	60.000	101.000
OG2P1	SG302	OG2P1	85.000	121.000

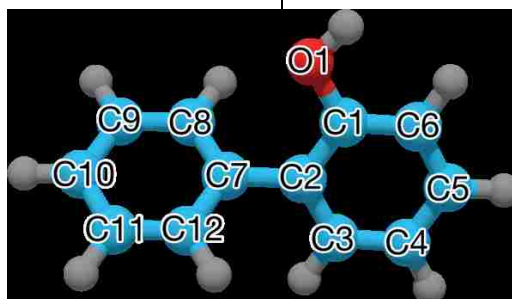
Dihedrals

Atom type	Atom type	Atom type	Atom type	K_{ϕ}	n	δ
CG2R61	CG2R67	CG2R67	CG2R61	0.8900	2	180.00
CG2R61	CG2R61	CG2R67	CG2R67	3.1000	2	180.00
HGR61	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
CG2R61	CG2R61	OG311	HGP1	0.9900	2	180.00
HGR61	CG2R61	CG2R67	CG2R61	4.2000	2	180.00
CG2R67	CG2R61	CG2R61	HGR61	4.2000	2	180.00
CG2R61	CG2R61	CG2R61	SG302	3.0000	2	180.00
SG302	CG2R61	CG2R61	HGR61	2.1000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R67	3.1000	2	180.00
CG2R61	CG2R61	CG2R61	OG311	3.1000	2	180.00
CG2R61	CG2R61	CG2R67	CG2R61	3.1000	2	180.00
CG2R61	CG2R61	SG302	OG2P1	0.0000	6	0.00
HGR61	CG2R61	CG2R67	CG2R67	4.2000	2	180.00
OG311	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2R61	CG2R61	CG2R61	HGR61	4.2000	2	180.00
SG302	CG2R61	CG2R67	CG2R61	1.0100	2	180.00
OG311	CG2R61	CG2R67	CG2R61	3.0000	2	180.00
CG2R67	CG2R61	SG302	OG2P1	1.9760	2	0.00

CG2R67	CG2R61	SG302	OG2P1	0.7510	4	180.00
SG302	CG2R61	CG2R67	CG2R67	2.2400	2	180.00
OG311	CG2R61	CG2R67	CG2R67	0.5490	2	180.00
CG2R67	CG2R61	OG311	HGP1	0.7080	2	180.00
CG2R67	CG2R61	OG311	HGP1	0.8880	4	0.00

HBP**Charges**

Atom	Atom type	Charges	Atom	Atom type	Charges
C1	CG2R61	0.129667	H1	HGR61	0.115000
C2	CG2R67	0.242667	H2	HGR61	0.115000
C3	CG2R61	-0.265333	H3	HGR61	0.115000
C4	CG2R61	-0.114000	H4	HGR61	0.115000
C5	CG2R61	-0.112000	H5	HGR61	0.115000
C6	CG2R61	-0.116000	H6	HGP1	0.471667
C7	CG2R67	-0.022333	H7	HGR61	0.115000
C8	CG2R61	-0.114000	H8	HGR61	0.115000
C9	CG2R61	-0.114000	H9	HGR61	0.115000
C10	CG2R61	-0.115000	H10	HGR61	0.115000
C11	CG2R61	-0.114000	O	OG311	-0.678333
C12	CG2R61	-0.114000			

**BONDS**

Atom type	Atom type	K _b	b ₀
-----------	-----------	----------------	----------------

CG2R61	CG2R61	305.000	1.375
CG2R61	CG2R67	305.000	1.375
CG2R61	OG311	334.300	1.411
CG2R61	HGR61	340.000	1.080
CG2R67	CG2R67	300.000	1.490
OG311	HGP1	545.000	0.960

Angles

Atom type	Atom type	Atom type	K_{θ}	θ_0
CG2R61	CG2R61	CG2R61	40.000	120.000
CG2R61	CG2R61	CG2R67	40.000	120.000
CG2R61	CG2R61	OG311	45.200	120.000
CG2R61	CG2R61	HGR61	30.000	120.000
CG2R67	CG2R61	OG311	43.330	110.545
CG2R67	CG2R61	HGR61	30.000	120.000
CG2R67	CG2R67	SG302	15.123	112.738
CG2R61	CG2R67	CG2R61	40.000	120.000
CG2R61	CG2R67	CG2R67	40.000	120.000
CG2R61	OG311	HGP1	65.000	108.000

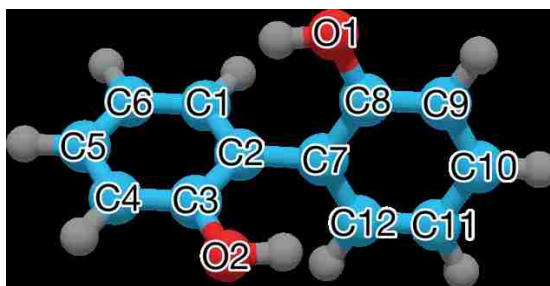
Dihedrals

Atom type	Atom type	Atom type	Atom type	K_{ϕ}	n	δ
CG2R61	CG2R67	CG2R67	CG2R61	0.8900	2	180.00
CG2R61	CG2R61	CG2R67	CG2R67	3.1000	2	180.00
HGR61	CG2R61	CG2R61	HGR61	2.4000	2	180.00

CG2R61	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
CG2R61	CG2R61	OG311	HGP1	0.9900	2	180.00
CG2R67	CG2R61	CG2R61	HGR61	4.2000	2	180.00
HGR61	CG2R61	CG2R67	CG2R61	4.2000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R67	3.1000	2	180.00
CG2R61	CG2R61	CG2R61	OG311	3.1000	2	180.00
CG2R61	CG2R61	CG2R67	CG2R61	3.1000	2	180.00
HGR61	CG2R61	CG2R67	CG2R67	4.2000	2	180.00
OG311	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2R61	CG2R61	CG2R61	HGR61	4.2000	2	180.00
OG311	CG2R61	CG2R67	CG2R61	2.9920	2	180.00
OG311	CG2R61	CG2R67	CG2R67	3.0000	2	180.00
CG2R67	CG2R61	OG311	HGP1	0.7460	2	180.00

BIPH**Charges**

Atom	Atom type	Charges	Atom	Atom type	Charges
C1	CG2R61	-0.154600	O1	OG311	-0.546600
C2	CG2R67	-0.003600	H1	HGR61	0.115000
C3	CG2R61	0.212400	H2	HGR61	0.115000
C4	CG2R61	-0.116000	H3	HGR61	0.115000
C5	CG2R61	-0.112000	H4	HGR61	0.115000
C6	CG2R61	-0.114000	H5	HGR61	0.115000
C7	CG2R67	-0.003600	H6	HGR61	0.115000
C8	CG2R61	0.212400	H7	HGR61	0.115000
C9	CG2R61	-0.116000	H8	HGP1	0.374400
C10	CG2R61	-0.112000	H9	HGR61	0.115000
C11	CG2R61	-0.114000	O2	OG311	-0.546600
C12	CG2R61	-0.154600	H10	HGP1	0.374400

**BONDS**

Atom type	Atom type	K _b	b ₀
CG2R61	CG2R61	305.000	1.375

CG2R61	CG2R67	305.000	1.375
CG2R61	OG311	334.300	1.411
CG2R61	HGR61	340.000	1.080
CG2R67	CG2R67	300.000	1.490
OG311	HGP1	545.000	0.960

Angles

Atom type	Atom type	Atom type	K_{θ}	θ_0
CG2R61	CG2R61	CG2R61	40.000	120.000
CG2R61	CG2R61	CG2R67	40.000	120.000
CG2R61	CG2R61	OG311	45.200	120.000
CG2R61	CG2R61	HGR61	30.000	120.000
CG2R67	CG2R61	OG311	55.730	115.855
CG2R67	CG2R61	HGR61	30.000	120.000
CG2R61	CG2R67	CG2R61	40.000	120.000
CG2R61	CG2R67	CG2R67	40.000	120.000
CG2R61	OG311	HGP1	65.000	108.000

Dihedrals

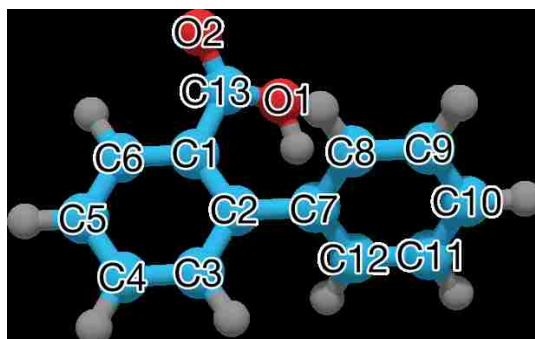
Atom type	Atom type	Atom type	Atom type	K_{ϕ}	n	δ
CG2R61	CG2R67	CG2R67	CG2R61	0.8900	2	180.00
CG2R61	CG2R61	CG2R67	CG2R67	3.1000	2	180.00
HGR61	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
CG2R61	CG2R61	OG311	HGP1	0.9900	2	180.00

CG2R67	CG2R61	CG2R61	HGR61	4.2000	2	180.00
HGR61	CG2R61	CG2R67	CG2R61	4.2000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R67	3.1000	2	180.00
CG2R61	CG2R61	CG2R61	OG311	3.1000	2	180.00
CG2R61	CG2R61	CG2R67	CG2R61	3.1000	2	180.00
HGR61	CG2R61	CG2R67	CG2R67	4.2000	2	180.00
OG311	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2R61	CG2R61	CG2R61	HGR61	4.2000	2	180.00
OG311	CG2R61	CG2R67	CG2R61	3.0000	2	180.00
OG311	CG2R61	CG2R67	CG2R67	3.0000	2	180.00
CG2R67	CG2R61	OG311	HGP1	1.3720	2	180.00

BCA

Charges

Atom	Atom type	Charges	Atom	Atom type	Charges
C1	CG2R61	0.141714	H1	HGR61	0.115000
C2	CG2R67	-0.215286	H2	HGR61	0.115000
C3	CG2R61	-0.036286	H3	HGR61	0.115000
C4	CG2R61	-0.114000	H4	HGR61	0.115000
C5	CG2R61	-0.115000	H5	HGR61	0.115000
C6	CG2R61	-0.119000	H6	HGR61	0.115000
C7	CG2R67	0.112714	H7	HGR61	0.115000
C8	CG2R61	-0.114000	H8	HGR61	0.115000
C9	CG2R61	-0.114000	H9	HGR61	0.115000
C10	CG2R61	-0.115000	C13	CG202	0.537000
C11	CG2R61	-0.114000	O1	OG311	-0.586286
C12	CG2R61	-0.113286	H10	HGP1	0.429000
			O2	OG2D1	-0.499286



BONDS

Atom type	Atom type	K_b	b_0
CG202	CG2R61	254.000	1.480
CG202	OG2D1	750.000	1.220
CG202	OG311	230.000	1.400
CG2R61	CG2R61	305.000	1.375
CG2R61	CG2R67	305.000	1.375
CG2R61	HGR61	340.000	1.080
CG2R67	CG2R67	300.000	1.490
OG311	HGP1	545.000	0.960

Angles

Atom type	Atom type	Atom type	K_θ	θ_0
CG2R61	CG202	OG2D1	70.000	123.100
CG2R61	CG202	OG311	40.000	113.900
OG2D1	CG202	OG311	50.000	123.000
CG202	CG2R61	CG2R61	45.000	120.000
CG202	CG2R61	CG2R67	36.487	114.086
CG2R61	CG2R61	CG2R61	40.000	120.000
CG2R61	CG2R61	CG2R67	40.000	120.000
CG2R61	CG2R61	HGR61	30.000	120.000
CG2R67	CG2R61	HGR61	30.000	120.000
CG2R61	CG2R67	CG2R61	40.000	120.000
CG2R61	CG2R67	CG2R67	40.000	120.000

CG202 OG311 HGP1 55.000 115.000

Dihedrals

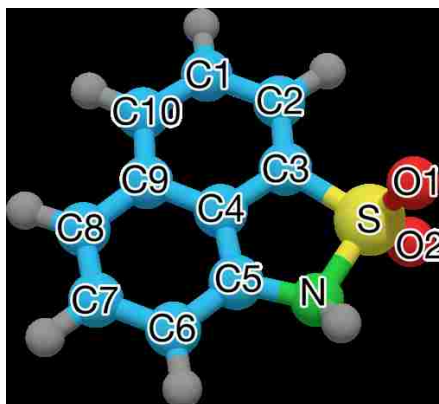
Atom type	Atom type	Atom type	Atom type	K_{ϕ}	n	δ
CG2R61	CG2R67	CG2R67	CG2R61	0.8900	2	180.00
CG2R61	CG2R61	CG2R67	CG2R67	3.1000	2	180.00
HGR61	CG2R61	CG2R61	HGR61	2.4000	2	180.00
OG311	CG202	CG2R61	CG2R61	1.0250	2	180.00
CG2R61	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
HGR61	CG2R61	CG2R67	CG2R61	4.2000	2	180.00
CG202	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
CG2R67	CG2R61	CG2R61	HGR61	4.2000	2	180.00
OG2D1	CG202	OG311	HGP1	2.0500	2	180.00
CG202	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R67	3.1000	2	180.00
CG2R61	CG2R61	CG2R67	CG2R61	3.1000	2	180.00
HGR61	CG2R61	CG2R67	CG2R67	4.2000	2	180.00
OG2D1	CG202	CG2R61	CG2R61	1.0250	2	180.00
CG2R61	CG202	OG311	HGP1	0.9750	1	180.00
CG2R61	CG202	OG311	HGP1	2.7000	2	180.00
CG2R61	CG202	OG311	HGP1	0.0500	3	180.00
CG2R61	CG202	OG311	HGP1	0.2500	6	180.00
CG2R61	CG2R61	CG2R61	HGR61	4.2000	2	180.00
OG311	CG202	CG2R61	CG2R67	1.9340	2	180.00

CG202	CG2R61	CG2R67	CG2R67	2.7270	2	180.00
OG2D1	CG202	CG2R61	CG2R67	0.1220	2	0.00
CG202	CG2R61	CG2R67	CG2R61	2.4840	2	180.00

NTAM

Charges

Atom	Atom type	Charges	Atom	Atom type	Charges
C1	CG2R61	-0.277846	H1	HGR61	0.196000
C2	CG2R61	-0.101846	H2	HGR61	0.244154
C3	CG2RC0	0.176154	H3	HGR61	0.208000
C4	CG2RC0	0.281154	H4	HGR61	0.196000
C5	CG2RC0	0.242154	H5	HGR61	0.196000
C6	CG2R61	-0.353846	H6	HGR61	0.196000
C7	CG2R61	-0.200000	H7	HGP1	0.378154
C8	CG2R61	-0.187000	S	SG302	-0.147846
C9	CG2R61	-0.074846	O1	OG2P1	-0.222846
C10	CG2R61	-0.187000	O2	OG2P1	-0.222846
N	NG3C51	-0.337846			



BONDS

Atom type	Atom type	K _b	b ₀
CG2R61	CG2R61	305.000	1.375

CG2R61	CG2RC0	300.000	1.360
CG2R61	HGR61	340.000	1.080
CG2RC0	CG2RC0	360.000	1.385
CG2RC0	NG3C51	330.000	1.400
CG2RC0	SG302	134.889	1.843
NG3C51	SG302	126.251	1.765
NG3C51	HGP1	450.000	1.018
OG2P1	SG302	630.000	1.440

Angles

Atom type	Atom type	Atom type	K_{θ}	θ_0
CG2R61	CG2R61	CG2R61	40.000	120.000
CG2R61	CG2R61	CG2RC0	50.000	120.000
CG2R61	CG2R61	HGR61	30.000	120.000
CG2RC0	CG2R61	HGR61	30.000	120.000
CG2R61	CG2RC0	CG2RC0	50.000	120.000
CG2R61	CG2RC0	NG3C51	35.000	130.700
CG2R61	CG2RC0	SG302	81.319	125.877
CG2RC0	CG2RC0	CG2RC0	94.337	114.328
CG2RC0	CG2RC0	NG3C51	100.000	109.300
CG2RC0	CG2RC0	SG302	136.049	113.105
CG2RC0	NG3C51	SG302	73.526	110.111
CG2RC0	NG3C51	HGP1	41.000	114.500

SG302	NG3C51	HGP1	70.303	110.249
CG2RC0	SG302	NG3C51	199.775	96.373
CG2RC0	SG302	OG2P1	105.177	114.081
NG3C51	SG302	OG2P1	69.221	111.276
OG2P1	SG302	OG2P1	85.000	121.000

Dihedrals

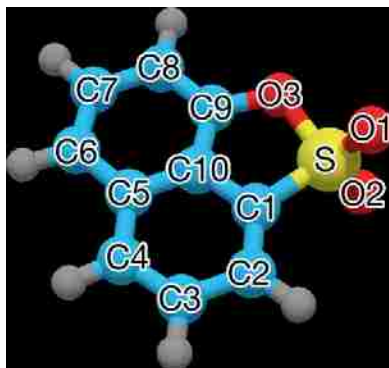
Atom type	Atom type	Atom type	Atom type	K_{ϕ}	n	δ
CG2R61	CG2RC0	CG2RC0	NG3C51	6.0000	2	180.00
HGR61	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2RC0	CG2RC0	NG3C51	HGP1	0.0000	3	0.00
CG2RC0	CG2R61	CG2R61	HGR61	3.0000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
HGR61	CG2R61	CG2RC0	NG3C51	0.0000	2	180.00
HGR61	CG2R61	CG2RC0	CG2RC0	3.0000	2	180.00
CG2R61	CG2R61	CG2R61	CG2RC0	3.0000	2	180.00
CG2R61	CG2RC0	CG2RC0	CG2R61	3.0000	2	180.00
CG2R61	CG2R61	CG2RC0	NG3C51	6.0000	2	180.00
CG2R61	CG2R61	CG2RC0	CG2RC0	3.0000	2	180.00
CG2R61	CG2R61	CG2R61	HGR61	4.2000	2	180.00
CG2R61	CG2RC0	NG3C51	HGP1	0.0000	3	0.00
CG2R61	CG2RC0	CG2RC0	CG2RC0	2.9700	2	180.00
CG2RC0	NG3C51	SG302	OG2P1	1.7280	3	0.00
CG2RC0	CG2RC0	CG2RC0	SG302	0.6840	2	0.00

CG2RC0	CG2RC0	NG3C51	SG302	2.9970	3	180.00
CG2RC0	CG2RC0	SG302	OG2P1	0.6730	6	180.00
CG2RC0	CG2RC0	CG2RC0	NG3C51	0.8510	2	0.00
CG2R61	CG2RC0	SG302	NG3C51	2.4750	2	0.00
HGP1	NG3C51	SG302	OG2P1	1.0560	3	0.00
HGP1	NG3C51	SG302	CG2RC0	1.1600	1	0.00
HGP1	NG3C51	SG302	CG2RC0	0.7420	2	0.00
HGP1	NG3C51	SG302	CG2RC0	0.6620	3	180.00
CG2R61	CG2R61	CG2RC0	SG302	2.9990	2	180.00
CG2RC0	CG2RC0	SG302	NG3C51	1.4850	2	0.00
CG2RC0	NG3C51	SG302	CG2RC0	2.5740	1	0.00
CG2RC0	NG3C51	SG302	CG2RC0	1.4940	2	180.00
HGR61	CG2R61	CG2RC0	SG302	1.5080	2	180.00
CG2R61	CG2RC0	CG2RC0	SG302	0.0540	2	180.00
CG2R61	CG2RC0	NG3C51	SG302	0.7110	2	0.00
CG2R61	CG2RC0	SG302	OG2P1	0.0530	6	180.00

NAPO

Charges

Atom	Atom type	Charges	Atom	Atom type	Charges
C1	CG2RC0	0.372583	H1	HGR61	0.304583
C2	CG2R61	-0.329417	H2	HGR61	0.196000
C3	CG2R61	-0.192417	H3	HGR61	0.196000
C4	CG2R61	-0.187000	H4	HGR61	0.196000
C5	CG2R61	-0.051417	H5	HGR61	0.196000
C6	CG2R61	-0.187000	H6	HGR61	0.278000
C7	CG2R61	-0.153000	S	SG302	-0.154417
C8	CG2R61	-0.449417	O1	OG2P1	-0.167417
C9	CG2RC0	0.399583	O2	OG2P1	-0.167417
C10	CG2RC0	0.102583	O3	OG3C51	-0.202417



BONDS

Atom type	Atom type	K _b	b ₀
CG2R61	CG2R61	305.000	1.375
CG2R61	CG2RC0	300.000	1.360

CG2R61	HGR61	340.000	1.080
CG2RC0	CG2RC0	360.000	1.385
CG2RC0	OG3C51	330.000	1.389
CG2RC0	SG302	135.870	1.829
OG2P1	SG302	630.000	1.440
OG3C51	SG302	88.596	1.765

Angles

Atom type	Atom type	Atom type	K_{θ}	θ_0
CG2R61	CG2R61	CG2R61	40.000	120.000
CG2R61	CG2R61	CG2RC0	50.000	120.000
CG2R61	CG2R61	HGR61	30.000	120.000
CG2RC0	CG2R61	HGR61	30.000	120.000
CG2R61	CG2RC0	CG2RC0	50.000	120.000
CG2R61	CG2RC0	OG3C51	50.000	125.300
CG2R61	CG2RC0	SG302	61.166	121.527
CG2RC0	CG2RC0	CG2RC0	112.940	112.353
CG2RC0	CG2RC0	OG3C51	80.000	114.700
CG2RC0	CG2RC0	SG302	140.678	108.541
CG2RC0	OG3C51	SG302	62.850	115.994
CG2RC0	SG302	OG2P1	75.040	96.537
CG2RC0	SG302	OG3C51	212.043	94.171
OG2P1	SG302	OG2P1	85.000	121.000

OG2P1 SG302 OG3C51 101.025 98.555

Dihedrals

Atom type	Atom type	Atom type	Atom type	K_{ϕ}	n	δ
HGR61	CG2R61	CG2RC0	OG3C51	2.4000	2	180.00
HGR61	CG2R61	CG2R61	HGR61	2.4000	2	180.00
CG2RC0	CG2R61	CG2R61	HGR61	3.0000	2	180.00
CG2R61	CG2R61	CG2R61	CG2R61	3.1000	2	180.00
CG2R61	CG2R61	CG2RC0	OG3C51	2.0000	2	180.00
HGR61	CG2R61	CG2RC0	CG2RC0	3.0000	2	180.00
CG2R61	CG2R61	CG2R61	CG2RC0	3.0000	2	180.00
CG2R61	CG2RC0	CG2RC0	OG3C51	4.0000	2	180.00
CG2R61	CG2RC0	CG2RC0	CG2R61	3.0000	2	180.00
CG2R61	CG2R61	CG2RC0	CG2RC0	3.0000	2	180.00
CG2R61	CG2R61	CG2R61	HGR61	4.2000	2	180.00
CG2R61	CG2RC0	CG2RC0	CG2RC0	0.8470	2	0.00
CG2RC0	CG2RC0	CG2RC0	SG302	1.6070	2	180.00
CG2RC0	CG2RC0	SG302	OG2P1	0.3010	6	180.00
CG2RC0	CG2RC0	SG302	OG3C51	0.6380	2	0.00
CG2R61	CG2RC0	OG3C51	SG302	1.8710	2	180.00
CG2R61	CG2R61	CG2RC0	SG302	2.9350	2	180.00
CG2RC0	OG3C51	SG302	OG2P1	1.9260	3	0.00
CG2RC0	CG2RC0	OG3C51	SG302	2.6030	2	180.00
CG2RC0	OG3C51	SG302	CG2RC0	0.8100	3	0.00

HGR61	CG2R61	CG2RC0	SG302	2.3400	2	180.00
CG2RC0	CG2RC0	CG2RC0	OG3C51	2.2540	2	180.00
CG2R61	CG2RC0	CG2RC0	SG302	2.2070	2	180.00
CG2R61	CG2RC0	SG302	OG2P1	0.0130	6	180.00
CG2R61	CG2RC0	SG302	OG3C51	2.1030	2	180.00

3 Additional information for chapter 4

3.1 Convergence evaluations for umbrella sampling

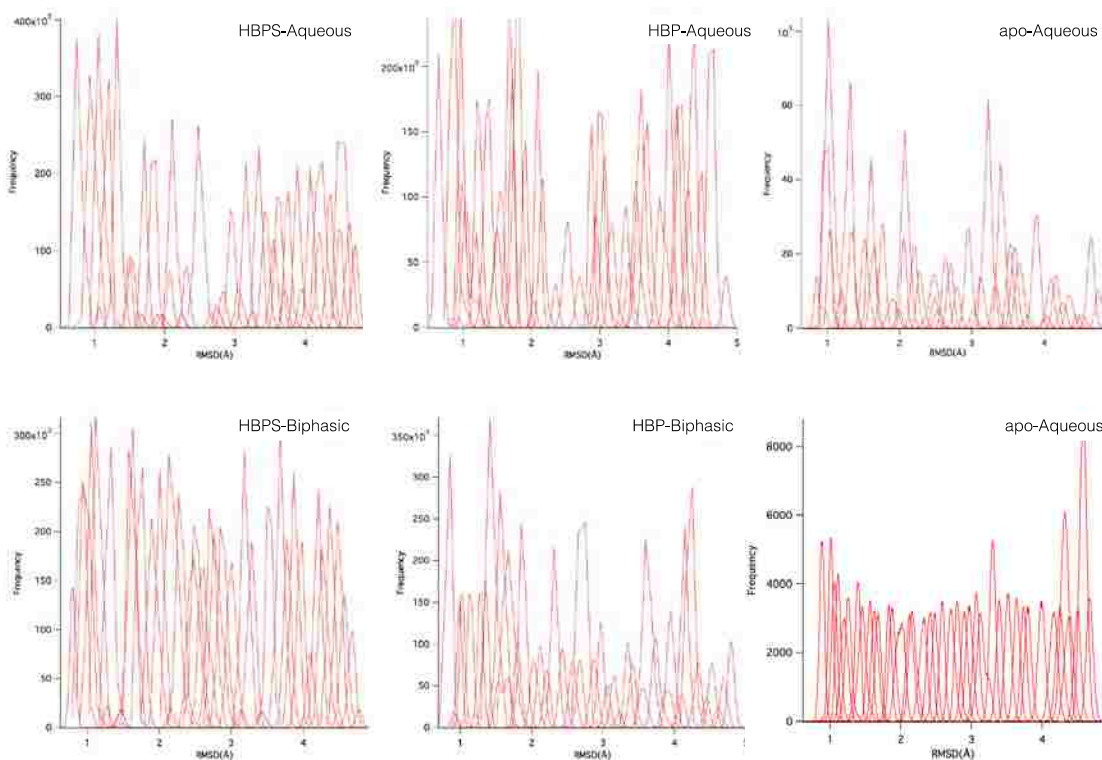


Figure S.1. Umbrella sampling window overlapping evaluations for *DszB*-HBPS and *DszB*-HBP in two solution systems.

3.2 Complete PCA results for of *DszB*-HBPS and *DszB*-HBP in aqueous solution

As described in manuscript in chapter 4, only one out of three parallel MD simulations results was shown, therefore, the other two MD simulation results were shown below for *DszB*-HBPS and *DszB*-HBP in aqueous solutions.

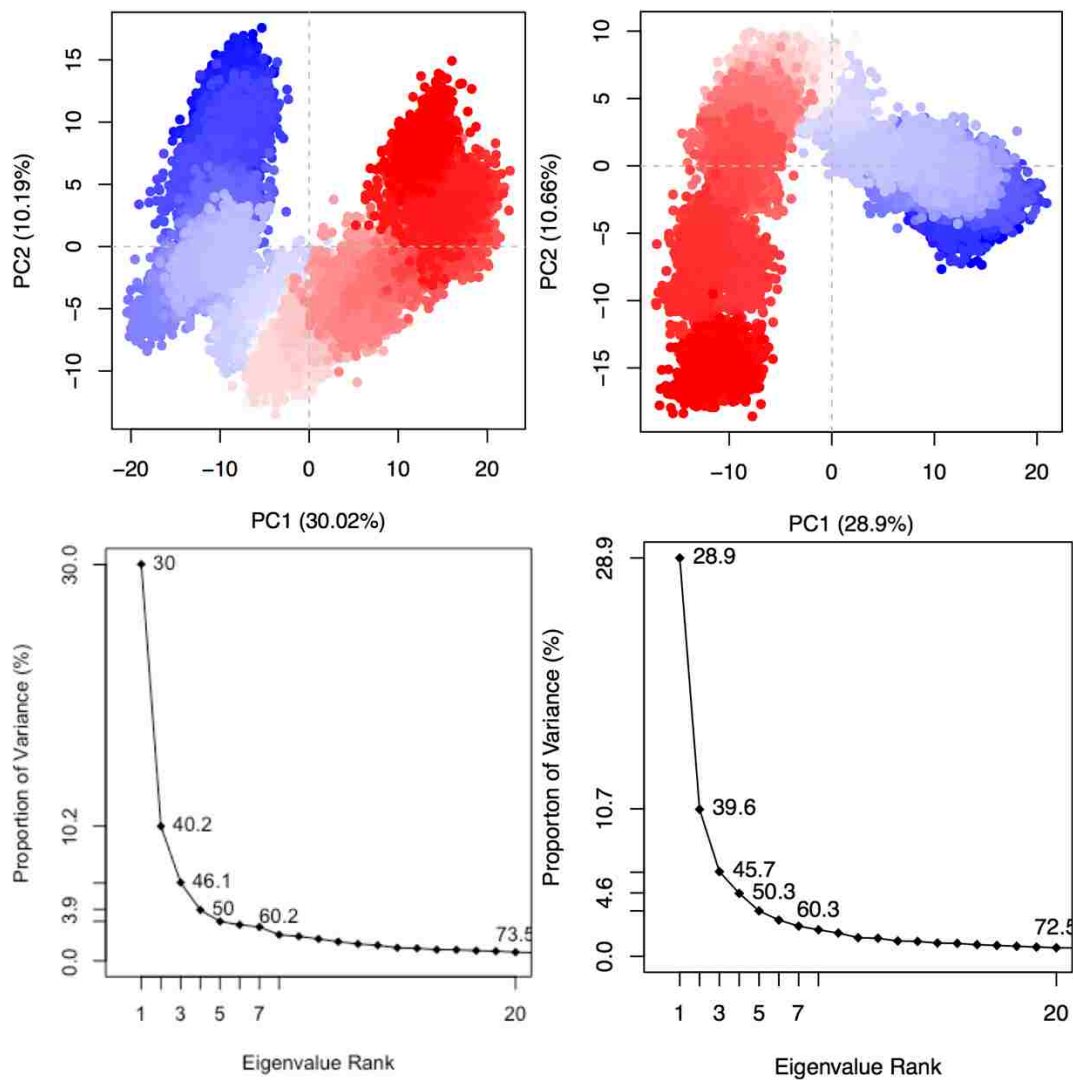


Figure S.2. Top panel, *DszB*-HBPS 2nd MD simulation and 3rd run PCA results, all MD conformers can be divided into four groups shows with color change from blue to red represent MD initial to final timeframe; The bottom plots represented the first three PCs represented by eigenvalue rank contribute 46% of the total conformational changes for *DszB*-HBPS MD 2nd and 3rd runs.

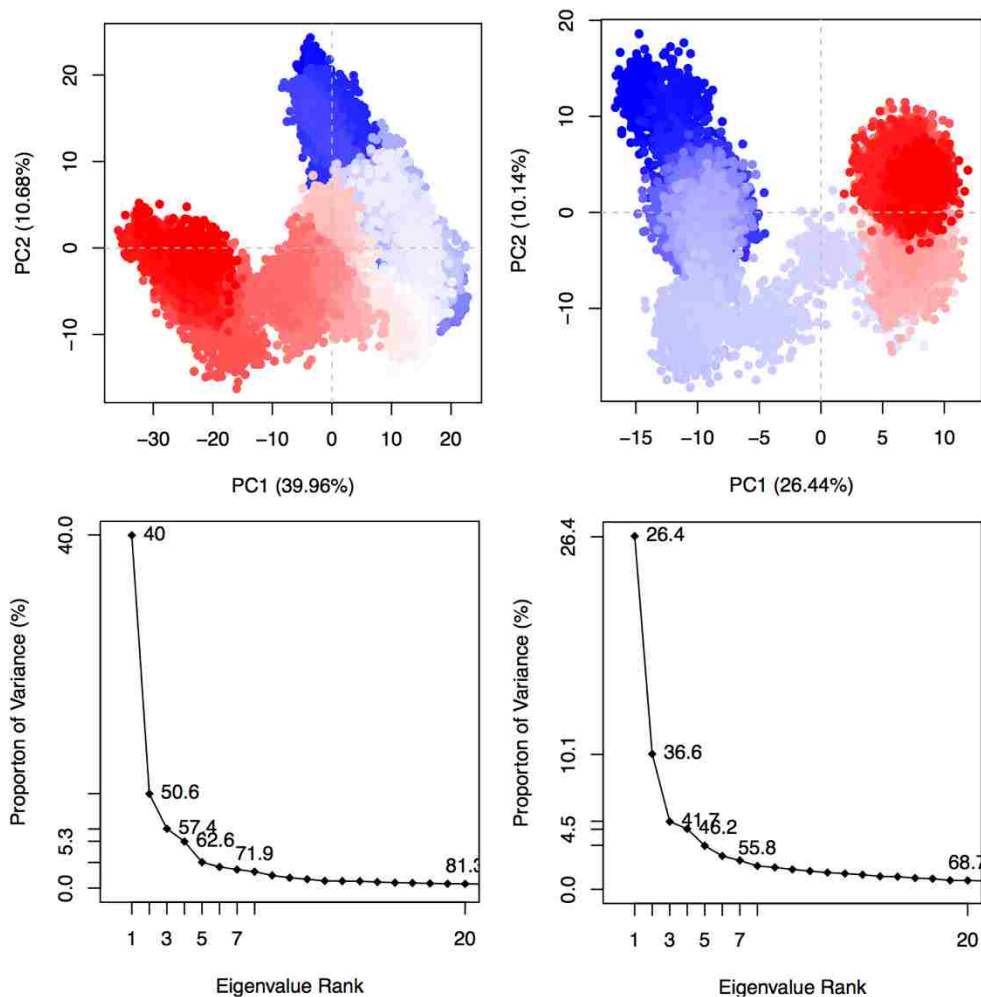


Figure S.3. Top panel, *DszB*-HBP 2nd MD simulation and 3rd run PCA results, all MD conformers can be divided into four groups shows with color change from blue to red represent MD initial to final timeframe; The bottom plots represented the first three PCs represented by eigenvalue rank contribute 57.4% and 41.7% of the total conformational changes for *DszB*-HBPS MD 2nd and 3rd runs respectively.

3.3 RMSD plots for *DszB*-HBPS and *DszB*-HBP in different solutions

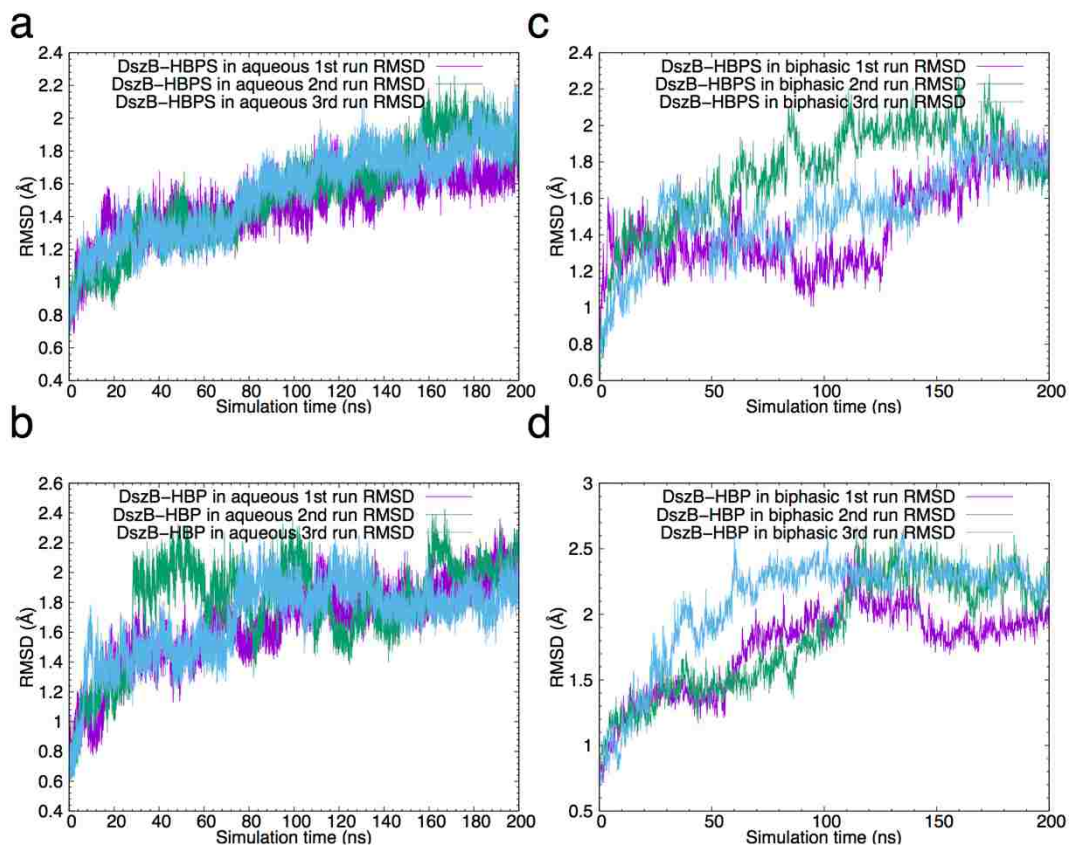


Figure S.4. (a)(b) represent *DszB*-HBPS and *DszB*-HBP in aqueous solution, both system reached to equilibrium after 130ns with RMSD values were not changing with respect to time; (c)(d) *DszB*-HBPS and *DszB*-HBP in biphasic solution, all systems reached to equilibrium after 140ns.

4 Additional information for chapter 5

4.1 Thermodynamic convergence

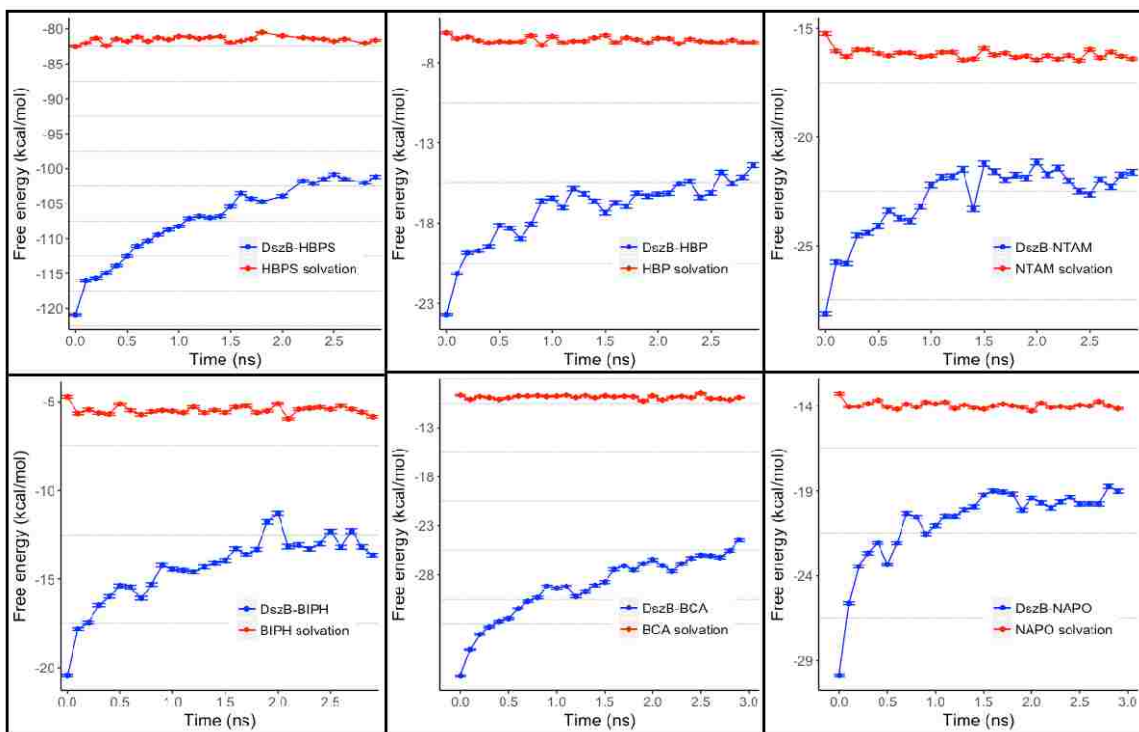


Figure S.5. Calculated Gibbs free energy over 30 consecutive 0.1-ns calculations using FEP/ λ -REMD for selected systems; blue line represents decoupling the bound ligand from *DszB* from solvate conditions, red line represents ligand solvation energy.

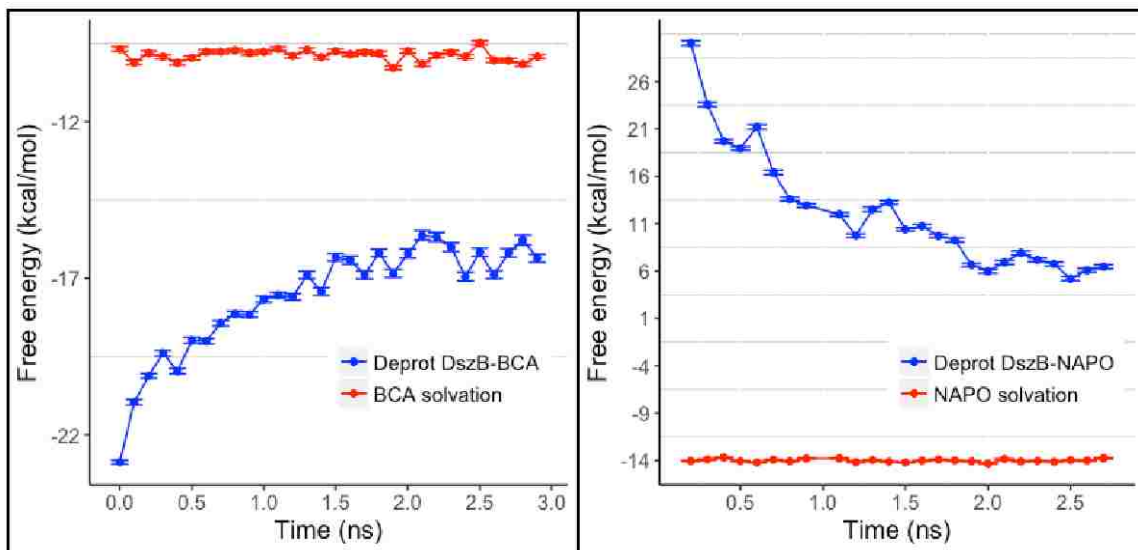


Figure S.6. Calculated Gibbs free energy using FEP/ λ -REMD for deprotonated R70 calculation on *DszB*-BCA (30 consecutive 0.1-ns calculations) and *DszB*-NAPO (25 consecutive 0.1-ns calculations); blue line represents decoupling the bound ligand from *DszB* from solvate conditions, red line represents ligand solvation energy.

4.2 Non-bonded interactions

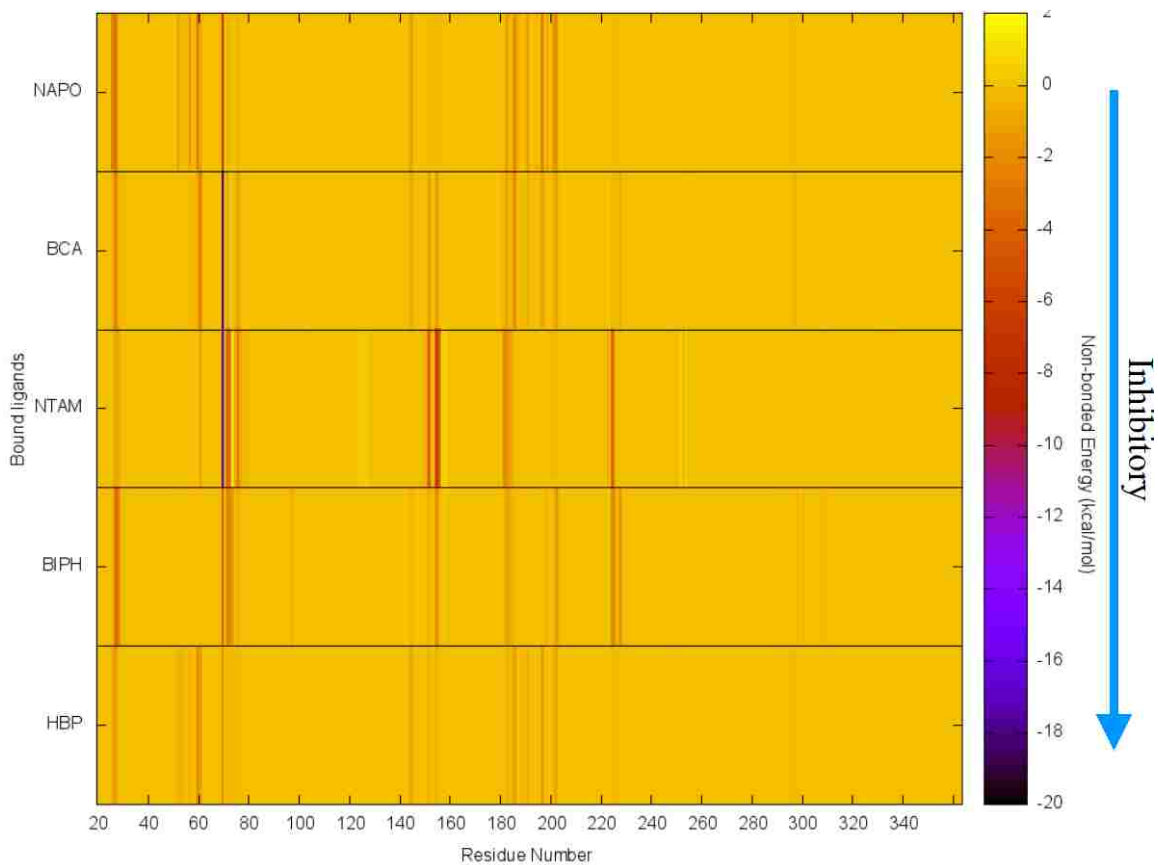


Figure S.7. Non-bonded energies (including VDW and electrostatic) for putative inhibitors with increasing inhibitory from top to bottom; BCA exhibited strong interactions with R70 resulted in large binding free energies calculated from FEP/ λ -REMD compared to experimental values.

4.3 Simulation snapshots for deprotonated R70

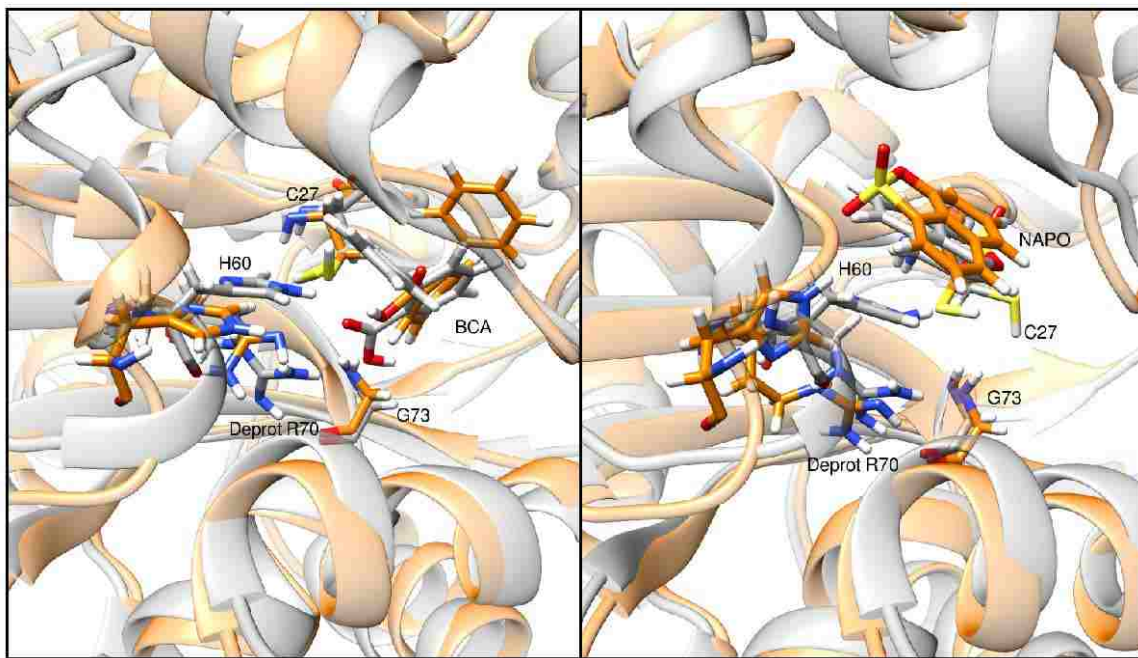


Figure S.8. Snapshots of *DszB*-BCA and *DszB*-NAPO with deprotonated R70 labeled as Deprot R70. Color representation: gray and orange represent the structure alignment at initial and final during 200-ns MD simulation. Deprotonation of R70 did not perturb the protein structure along the time trajectory

References

1. *Annual Energy Outlook 2017, Energy information administration*. Department of Energy, 2017.
2. Obuekwe, C., D. Westlake, and F. Cook, *Corrosion of Pembina crude oil pipeline*. European journal of applied microbiology and biotechnology, 1983. **17**(3): p. 173-177.
3. Takaoka, S., *Process Economics Program, Report 47*. 1969, Stanford Res. Inst. Menlo Park, CA.
4. Carrales, M. and R.W. Martin, *Sulfur content of crude oils*. 1975: US Department of the Interior, Bureau of Mines.
5. *The Availability and Price of Petroleum and Petroleum Products Produced in Countries Other Than Iran, US Energy Information Administration*. Department of Energy, 2013.
6. McFarland, B.L., et al., *Biocatalytic sulfur removal from fuels: applicability for producing low sulfur gasoline*. Critical reviews in microbiology, 1998. **24**(2): p. 99-147.
7. Swaty, T., *Global refining industry trends: the present and future*. Hydrocarbon processing, 2005. **84**(9): p. 35-35.
8. Javadli, R. and A. De Klerk, *Desulfurization of heavy oil*. Applied petrochemical research, 2012. **1**(1-4): p. 3-19.
9. Leprince, P., *Petroleum Refining. Vol. 3 Conversion Processes*. Vol. 3. 2001: Editions Technip.
10. Weast, R., M. Astle, and W. Beyer, *CRC Handbook of Chemistry and Physics (CRC, Boca Raton, FL, 1983)*, p. E-103, 1982.
11. Kabe, T., A. Ishihara, and H. Tajima, *Hydrodesulfurization of sulfur-containing polyaromatic compounds in light oil*. Industrial & engineering chemistry research, 1992. **31**(6): p. 1577-1580.
12. Gray, M.R., et al., *Kinetics of hydrodesulfurization of thiophenic and sulfide sulfur in Athabasca bitumen*. Energy & fuels, 1995. **9**(3): p. 500-506.
13. Denome, S.A., E.S. Olson, and K.D. Young, *Identification and cloning of genes involved in specific desulfurization of dibenzothiophene by Rhodococcus sp. strain IGTS8*. Applied and environmental microbiology, 1993. **59**(9): p. 2837-2843.
14. Linguist, L. and M. Pacheco, *Enzyme-based diesel desulfurization process offers energy, CO2 advantages*. Oil and Gas Journal, 1999. **97**(8): p. 45-50.
15. Agarwal, P. and D. Sharma, *Comparative studies on the bio-desulfurization of crude oil with other desulfurization techniques and deep desulfurization through integrated processes*. Energy & Fuels, 2009. **24**(1): p. 518-524.

16. Kilbane, J.J., *Microbial biocatalyst developments to upgrade fossil fuels*. Current Opinion in Biotechnology, 2006. **17**(3): p. 305-314.
17. Le Borgne, S. and R. Quintero, *Biotechnological processes for the refining of petroleum*. Fuel Processing Technology, 2003. **81**(2): p. 155-169.
18. Oldfield, C., et al., *Elucidation of the metabolic pathway for dibenzothiophene desulphurization by Rhodococcus sp. strain IGTS8 (ATCC 53968)*. Microbiology, 1997. **143**(9): p. 2961-2973.
19. Gray, K.A., et al., *Molecular mechanisms of biocatalytic desulfurization of fossil fuels*. Nature biotechnology, 1996. **14**(13): p. 1705-1709.
20. Watkins, L., et al., *Purification and characterization of the aromatic desulfinase, 2-(2'-hydroxyphenyl) benzenesulfinate desulfinase*. Archives of biochemistry and biophysics, 2003. **415**(1): p. 14-23.
21. Nakayama, N., et al., *A novel enzyme, 2'-hydroxybiphenyl-2-sulfinate desulfinase (DszB), from a dibenzothiophene-desulfurizing bacterium Rhodococcus erythropolis KA2-5-1: gene overexpression and enzyme characterization*. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2002. **1598**(1): p. 122-130.
22. Caro, A., et al., *Description of by-product inhibition effects on biodesulfurization of dibenzothiophene in biphasic media*. Biodegradation, 2008. **19**(4): p. 599-611.
23. Davoodi-Dehaghani, F., M. Vosoughi, and A.A. Ziaee, *Biodesulfurization of dibenzothiophene by a newly isolated Rhodococcus erythropolis strain*. Bioresource technology, 2010. **101**(3): p. 1102-1105.
24. Ohshiro, T., et al., *Novel Reactivity of Dibenzothiophene Monooxygenase from Bacillus subtilis WU-S2B*. Bioscience, biotechnology, and biochemistry, 2009. **73**(9): p. 2128-2130.
25. Reichmuth, D.S., et al., *Biodesulfurization of dibenzothiophene in Escherichia coli is enhanced by expression of a Vibrio harveyi oxidoreductase gene*. Biotechnology and bioengineering, 2000. **67**(1): p. 72-79.
26. Maghsoudi, S., et al., *Biodesulfurization of hydrocarbons and diesel fuels by Rhodococcus sp. strain P32C1*. Biochemical Engineering Journal, 2001. **8**(2): p. 151-156.
27. Kodama, K., et al., *Identification of microbial products from dibenzothiophene and its proposed oxidation pathway*. Agricultural and Biological Chemistry, 1973. **37**(1): p. 45-50.
28. Aggarwal, S., I. Karimi, and G.R. Ivan, *In silico modeling and evaluation of Gordonia alkanivorans for biodesulfurization*. Molecular BioSystems, 2013. **9**(10): p. 2530-2540.
29. Piddington, C.S., B.R. Kovacevich, and J. Rambosek, *Sequence and molecular characterization of a DNA region encoding the dibenzothiophene*

- desulfurization operon of Rhodococcus sp. strain IGTS8*. Applied and environmental microbiology, 1995. **61**(2): p. 468-475.
30. Gupta, N., P. Roychoudhury, and J. Deb, *Biotechnology of desulfurization of diesel: prospects and challenges*. Applied microbiology and biotechnology, 2005. **66**(4): p. 356-366.
 31. Folsom, B., et al., *Microbial desulfurization of alkylated dibenzothiophenes from a hydrodesulfurized middle distillate by Rhodococcus erythropolis I-19*. Applied and environmental microbiology, 1999. **65**(11): p. 4967-4972.
 32. Gray, K.A., G.T. Mrachko, and C.H. Squires, *Biodesulfurization of fossil fuels*. Current opinion in microbiology, 2003. **6**(3): p. 229-235.
 33. Mihara, H., et al., *Cysteine Sulfinase Desulfinase, a NIFS-like Protein of Escherichia coli with Selenocysteine Lyase and Cysteine Desulfurase Activities GENE CLONING, PURIFICATION, AND CHARACTERIZATION OF A NOVEL PYRIDOXAL ENZYME*. Journal of Biological Chemistry, 1997. **272**(36): p. 22417-22424.
 34. Miles, E.W. and A. Meister, *The Mechanism of the Reaction of β -hydroxyaspartate with L-Aspartate β -Decarboxylase. A New Type of Pyridoxal 5'-Phosphate-Enzyme Inhibition*. Biochemistry, 1967. **6**(6): p. 1734-1743.
 35. Yue Yu, I.A.F., Landon C. Mills, Derek L. Englert, Brad J. Berron, Christina M. Payne, *CHARMM force field parameters for 2'-hydroxybiphenyl-2-sulfinase, 2-hydroxybiphenyl, and related analogs*. Journal of Molecular Graphics and Modelling, 2017. **72**: p. 32-42.
 36. Kilbane, J.J., *Sulfur-specific microbial metabolism of organic compounds*. Resources, Conservation and Recycling, 1990. **3**(2-3): p. 69-79.
 37. Mohebali, G. and A.S. Ball, *Biocatalytic desulfurization (BDS) of petrodiesel fuels*. Microbiology, 2008. **154**(8): p. 2169-2183.
 38. Lee, M., J. Senius, and M. Grossman, *Sulfur-specific microbial desulfurization of sterically hindered analogs of dibenzothiophene*. Applied and environmental microbiology, 1995. **61**(12): p. 4362-4366.
 39. Lee, W.C., et al., *Crystal structure and desulfurization mechanism of 2'-hydroxybiphenyl-2-sulfinic acid desulfinase*. Journal of Biological Chemistry, 2006. **281**(43): p. 32534-32539.
 40. Polgár, L., *Mercaptide—imidazolium ion-pair: The reactive nucleophile in papain catalysis*. FEBS letters, 1974. **47**(1): p. 15-18.
 41. Chen, H.Y., T.V. Demidkina, and R.S. Phillips, *Site-directed mutagenesis of tyrosine-71 to phenylalanine in Citrobacter freundii tyrosine phenol-lyase: evidence for dual roles of tyrosine-71 as a general acid catalyst in the reaction mechanism and in cofactor binding*. Biochemistry, 1995. **34**(38): p. 12276-12283.

42. Geronimo, I., S.R. Nigam, and C.M. Payne, *Desulfination by 2[prime or minute]-hydroxybiphenyl-2-sulfinate desulfinase proceeds via electrophilic aromatic substitution by the cysteine-27 proton*. *Chemical Science*, 2017. **8**(7): p. 5078-5086.
43. Galabov, B., et al., *Electrophilic aromatic substitution: new insights into an old class of reactions*. *Acc. Chem. Res*, 2016. **49**(6): p. 1191-1199.
44. Payne, C.M., et al., *Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation*. *Journal of Biological Chemistry*, 2011. **286**(47): p. 41028-41035.
45. Payne, C.M., et al., *Hallmarks of processivity in glycoside hydrolases from crystallographic and computational studies of the *Serratia marcescens* chitinases*. *Journal of Biological Chemistry*, 2012. **287**(43): p. 36322-36330.
46. Bu, L., et al., *Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods*. *Journal of Biological Chemistry*, 2011. **286**(20): p. 18161-18169.
47. Brooks, B.R., et al., *CHARMM: the biomolecular simulation program*. *Journal of computational chemistry*, 2009. **30**(10): p. 1545-1614.
48. Jiang, W., M. Hodoscek, and B. Roux, *Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics*. *Journal of chemical theory and computation*, 2009. **5**(10): p. 2583-2588.
49. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. *Nature Structural & Molecular Biology*, 2002. **9**(9): p. 646-652.
50. Dror, R.O., et al., *Biomolecular simulation: a computational microscope for molecular biology*. *Annual review of biophysics*, 2012. **41**: p. 429-452.
51. Schlick, T., et al., *Biomolecular modeling and simulation: a field coming of age*. *Quarterly reviews of biophysics*, 2011. **44**(2): p. 191-228.
52. Frenkel, D. and B. Smit, *Understanding molecular simulations: from algorithms to applications*. 2002, Academic Press.
53. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. *Journal of computational chemistry*, 2005. **26**(16): p. 1781-1802.
54. Case, D., et al., *AMBER 2015 (University of California, San Francisco)*. There is no corresponding record for this reference, 2015.
55. Berendsen, H.J., D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation*. *Computer Physics Communications*, 1995. **91**(1-3): p. 43-56.
56. Lindahl, E., B. Hess, and D. Van Der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*. *Journal of molecular modeling*, 2001. **7**(8): p. 306-317.

57. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free*. Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.
58. Hess, B., et al., *GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation*. Journal of chemical theory and computation, 2008. **4**(3): p. 435-447.
59. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1**: p. 19-25.
60. Brook, B., R. Brucoleri, and B. Olafson, *States DJ, Swaminathan S, Karplus M: CHARMM: a program for macromolecular energy, minimization and dynamic calculations*. J Comput Chem, 1983. **4**: p. 187-217.
61. MacKerell, A.D., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids*. Biopolymers, 2000. **56**(4): p. 257-265.
62. Berendsen, H., J. Grigera, and T. Straatsma, *The missing term in effective pair potentials*. Journal of Physical Chemistry, 1987. **91**(24): p. 6269-6271.
63. Deng, Y. and B. Roux, *Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules*. The Journal of Physical Chemistry B, 2004. **108**(42): p. 16567-16576.
64. Deng, Y. and B. Roux, *Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant*. Journal of Chemical Theory and Computation, 2006. **2**(5): p. 1255-1273.
65. Wang, J., Y. Deng, and B. Roux, *Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials*. Biophysical journal, 2006. **91**(8): p. 2798-2814.
66. Deng, Y. and B. Roux, *Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations*. The Journal of chemical physics, 2008. **128**(11): p. 03B611.
67. Rick, S.W., *Increasing the efficiency of free energy calculations using parallel tempering and histogram reweighting*. Journal of chemical theory and computation, 2006. **2**(4): p. 939-946.
68. Rhee, Y.M. and V.S. Pande, *Multiplexed-replica exchange molecular dynamics method for protein folding simulation*. Biophysical journal, 2003. **84**(2): p. 775-786.
69. Sugita, Y., A. Kitao, and Y. Okamoto, *Multidimensional replica-exchange method for free-energy calculations*. The Journal of Chemical Physics, 2000. **113**(15): p. 6042-6051.

70. Mitsutake, A. and Y. Okamoto, *From multidimensional replica-exchange method to multidimensional multicanonical algorithm and simulated tempering*. Physical Review E, 2009. **79**(4): p. 047701.
71. Fajer, M., D. Hamelberg, and J.A. McCammon, *Replica-exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic integration*. Journal of chemical theory and computation, 2008. **4**(10): p. 1565-1569.
72. Min, D., et al., *Synergistic approach to improve "alchemical" free energy calculation in rugged energy surface*. The Journal of chemical physics, 2007. **126**(14): p. 144109.
73. Sugita, Y. and Y. Okamoto, *Chem. Phys. Lett.* 1999(314): p. 141-151.
74. Woodcock, H.L., et al., *Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations*. Journal of computational chemistry, 2007. **28**(9): p. 1485-1502.
75. Woodcock, H.L., et al., *Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase*. Theoretical Chemistry Accounts, 2003. **109**(3): p. 140-148.
76. Jiang, W., *Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics*. Journal of Chemical Theory and Computation, 2009. **5**: p. 2583-2588.
77. Jiang, W., *Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics(FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations*. Journal of Chemical Theory and Computation, 2010. **6**: p. 2559-2565.
78. Torrie, G.M. and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*. Journal of Computational Physics, 1977. **23**(2): p. 187-199.
79. Torrie, G.M. and J.P. Valleau, *Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid*. Chemical Physics Letters, 1974. **28**(4): p. 578-581.
80. Kästner, J., *Umbrella sampling*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. **1**(6): p. 932-942.
81. Kumar, S., et al., *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*. Journal of computational chemistry, 1992. **13**(8): p. 1011-1021.
82. Grossfield, A., *WHAM: the weighted histogram analysis method, version 2.0. 9* (<http://membrane.urmc.rochester.edu/content/wham>). There is no corresponding record for this reference.
83. Ferrenberg, A.M. and R.H. Swendsen, *New Monte Carlo technique for studying phase transitions*. Physical Review Letters, 1989. **63**(15): p. 1658.

84. Ferrenberg, A.M. and R.H. Swendsen, *Optimized monte carlo data analysis*. Physical Review Letters, 1989. **63**(12): p. 1195.
85. Roux, B., *The calculation of the potential of mean force using computer simulations*. Computer physics communications, 1995. **91**(1-3): p. 275-282.
86. Rana, M.S., et al., *A review of recent advances on process technologies for upgrading of heavy oils and residua*. Fuel, 2007. **86**(9): p. 1216-1231.
87. MacKerell Jr, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. The journal of physical chemistry B, 1998. **102**(18): p. 3586-3616.
88. Mackerell, A.D., *Empirical force fields for biological macromolecules: Overview and issues*. Journal of Computational Chemistry, 2004. **25**(13): p. 1584-1604.
89. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields*. Journal of computational chemistry, 2010. **31**(4): p. 671-690.
90. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. Journal of molecular graphics, 1996. **14**(1): p. 33-38.
91. Mayne, C.G., et al., *Rapid parameterization of small molecules using the force field toolkit*. Journal of computational chemistry, 2013. **34**(32): p. 2757-2770.
92. ElGamacy, M. and L. Van Meervelt, *A fast topological analysis algorithm for large-scale similarity evaluations of ligands and binding pockets*. Journal of cheminformatics, 2015. **7**(1): p. 42.
93. Vanommeslaeghe, K. and A.D. MacKerell Jr, *Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing*. Journal of chemical information and modeling, 2012. **52**(12): p. 3144-3154.
94. Yu, W., et al., *Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations*. Journal of computational chemistry, 2012. **33**(31): p. 2451-2468.
95. Frisch, M., et al., *Gaussian 09, Revision C. 01, Gaussian, Inc., Wallingford CT, 2009 Search PubMed;(b) Y. Zhao and DG Truhlar*. Theor. Chem. Acc, 2008. **120**: p. 215.
96. Guvench, O. and A.D. MacKerell, *Automated conformational energy fitting for force-field development*. Journal of molecular modeling, 2008. **14**(8): p. 667-679.
97. Nelder, J.A. and R. Mead, *A simplex method for function minimization*. The computer journal, 1965. **7**(4): p. 308-313.
98. Press, W.H., et al., *Numerical Recipes Example Book:(C)*. 1990: Cambridge University Press.

99. Feig, M., J. Karanicolas, and C.L. Brooks, *MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology*. Journal of Molecular Graphics and Modelling, 2004. **22**(5): p. 377-395.
100. Center, N.M.S.D. and S. Stein, director, "Infrared Spectra" in *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, Eds. PJ Linstrom and WG Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899.
101. Tanner, D.E., et al., *Parallel generalized Born implicit solvent calculations with NAMD*. Journal of chemical theory and computation, 2011. **7**(11): p. 3635-3642.
102. Almenningen, A., et al., *Structure and barrier of internal rotation of biphenyl derivatives in the gaseous state: Part 1. The molecular structure and normal coordinate analysis of normal biphenyl and perdeuterated biphenyl*. Journal of molecular structure, 1985. **128**(1-3): p. 59-76.
103. Perrin, M., K. Bekkouch, and A. Thozet, *Structure of 2-hydroxybiphenyl*. Acta Crystallographica Section C: Crystal Structure Communications, 1987. **43**(5): p. 980-982.
104. Chen, X.-M., et al., *2, 2'-Biphenol Monohydrate*. Acta Crystallographica Section C: Crystal Structure Communications, 1996. **52**(7): p. 1727-1729.
105. Cruickshank, D., *A detailed refinement of the crystal and molecular structure of naphthalene*. Acta Crystallographica, 1957. **10**(8): p. 504-508.
106. Balasubramanian, V., *Peri Interaction in naphthalene derivatives*. Chemical Reviews, 1966. **66**(6): p. 567-641.
107. Koch, W., M.C. Holthausen, and M. Kaupp, *BUCHER-A Chemist's Guide to Density Functional Theory*. Angewandte Chemie-German Edition, 2001. **113**(5): p. 989-989.
108. Morrow, T.I. and E.J. Maginn, *Molecular dynamics study of the ionic liquid 1-n-butyl-3-methylimidazolium hexafluorophosphate*. The Journal of Physical Chemistry B, 2002. **106**(49): p. 12807-12813.
109. SDBSWeb, <http://sdbb.db.aist.go.jp>, (National Institute of Advanced Industrial Science and Technology).
110. Li H., Z.X.C., *Synthesis of 1,3-propane sultone*. Journal of Chemical Science and Technology 2, 2013: p. 209-212.
111. Meille, V., et al., *Hydrodesulfurization of alkyl dibenzothiophenes over a NiMo/Al₂O₃ Catalyst: kinetics and mechanism*. Journal of Catalysis, 1997. **170**(1): p. 29-36.
112. D. J. Monticello, W.R.F., *Microbial Desulfurization of Fossil Fuels*. Annual Review Microbiology, 1985. **39**: p. 371-89.

113. Van Hamme, J.D., A. Singh, and O.P. Ward, *Recent advances in petroleum microbiology*. Microbiology and molecular biology reviews, 2003. **67**(4): p. 503-549.
114. McFarland, B.L., *Biodesulfurization*. Current opinion in microbiology, 1999. **2**(3): p. 257-264.
115. Gray, M.R. and A.R. Ayasse, *Kinetics of hydrodesulfurization of thiophenic and sulfide sulfur in Athabasca bitumen*. Energy & Fuels, 1995. **9**(3): p. 500-506.
116. Gallagher, J.R., E.S. Olson, and D.C. Stanley, *Microbial desulfurization of dibenzothiophene: a sulfur-specific pathway*. FEMS Microbiology Letters, 1993. **107**(1): p. 31-35.
117. Abin-Fuentes, A., et al., *Exploring the mechanism of biocatalyst inhibition in microbial desulfurization*. Applied and environmental microbiology, 2013. **79**(24): p. 7807-7817.
118. Gordon, J.C., et al., *H++: a server for estimating pKas and adding missing hydrogens to macromolecules*. Nucleic acids research, 2005. **33**(suppl 2): p. W368-W371.
119. Hanwell, M.D., et al., *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform*. Journal of cheminformatics, 2012. **4**(1): p. 17.
120. Gaussian09, R.A., *1, MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G. Scalmani, V. Barone, B. Mennucci, GA Petersson et al., Gaussian. Inc., Wallingford CT, 2009.*
121. Jolliffe, I.T., *Principal component analysis and factor analysis*. Principal component analysis, 2002: p. 150-166.
122. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures*. Bioinformatics, 2006. **22**(21): p. 2695-2696.
123. Grossfield, A., *WHAM: the weighted histogram analysis method*. version, 2013. **2.0.9**(<http://membrane.urmc.rochester.edu/content/wham>).
124. Perry, H., *Coal in the United States: a status report*. Science, 1983. **222**: p. 377-385.
125. Secretariat, U.N.E.C.f.E., *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)*. 2009: United Nations Publications.
126. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility*. Journal of computational chemistry, 2009. **30**(16): p. 2785-2791.
127. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of chemical physics, 1983. **79**(2): p. 926-935.
128. Durell, S.R., B.R. Brooks, and A. Bennaïm, *Solvent-induced forces between 2 hydrophilic groups*. Journal of Physical Chemistry, 1994. **98**(8): p. 2198-2202.

129. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles*. Journal of chemical theory and computation, 2012. **8**(9): p. 3257.
130. MacKerell, A.D., M. Feig, and C.L. Brooks, *Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations*. Journal of computational chemistry, 2004. **25**(11): p. 1400-1415.
131. MacKerell Jr, A., et al., *J. Wio rkiewicz-Kuczera, D. Yin and M. Karplus*. J. Phys. Chem. B, 1998. **102**: p. 3586.
132. Nosé, S., *A unified formulation of the constant temperature molecular dynamics methods*. The Journal of chemical physics, 1984. **81**(1): p. 511-519.
133. Hoover, W.G., *Canonical dynamics: equilibrium phase-space distributions*. Physical review A, 1985. **31**(3): p. 1695.
134. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. J Comput Chem, 2005. **26**(16): p. 1781-802.
135. Shirts, M.R. and J.D. Chodera, *Statistically optimal analysis of samples from multiple equilibrium states*. The Journal of chemical physics, 2008. **129**(12): p. 124105.
136. Hünenberger, P.H. and J.A. McCammon, *Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: a continuum electrostatics study*. The Journal of chemical physics, 1999. **110**(4): p. 1856-1872.
137. Kastenholz, M.A. and P.H. Hünenberger, *Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods*. The Journal of Physical Chemistry B, 2004. **108**(2): p. 774-788.
138. Kastenholz, M.A. and P.H. Hünenberger, *Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids*. The Journal of chemical physics, 2006. **124**(12): p. 124106.
139. Rocklin, G.J., et al., *Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects*. The Journal of chemical physics, 2013. **139**(18): p. 11B606_1.
140. Schlippe, Y.V.G. and L. Hedstrom, *A twisted base? The role of arginine in enzyme-catalyzed proton abstractions*. Archives of Biochemistry and Biophysics, 2005. **433**(1): p. 266-278.

Vita

Personal Information

Place of Birth Xi'an, Shaanxi, China

Education

2014 – Present Ph.D. Chemical Engineering, University of Kentucky

2011 – 2014 B.E. Chemical Engineering, University of Kentucky

2009 – 2011 B.E. Process Equipment and Control Engineering,
China Univ. of Mining and Tech.

Publications

Yue Yu, I. A. F., Landon C. Mills, Derek L. Englert, Brad J. Berron, and Christina M. Payne, "CHARMM force field parameters for 2'-hydroxybiphenyl-2-sulfinate, 2-hydroxybiphenyl, and related analogs". *Journal of Molecular Graphics and Modelling* 2017, 72, 32-42.

Yue Yu, and Christina M. Payne, "Conformational change of 2'-hydroxybiphenyl-2-sulfinate desulfinate upon ligand binding". (In preparation)

Yue Yu, Landon C. Mills, and Christina M. Payne, "Thermodynamic relationships between 2'-hydroxybiphenyl-2-sulfinate desulfinate and putative inhibitors". (In preparation)