University of Kentucky
UKnowledge

Theses and Dissertations--Chemical and Materials Engineering

Chemical and Materials Engineering

2019

IMPACT OF CONFORMATIONAL CHANGE, SOLVATION ENVIRONMENT, AND POST-TRANSLATIONAL MODIFICATION ON DESULFURIZATION ENZYME 2'-HYDROXYBIPHENYL-2-SULFINATE DESULFINASE (*DSZB*) STABILITY AND ACTIVITY

Landon C. Mills
University of Kentucky, lcmi222@g.uky.edu
Digital Object Identifier: https://doi.org/10.13023/etd.2019.329

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation
Mills, Landon C., "IMPACT OF CONFORMATIONAL CHANGE, SOLVATION ENVIRONMENT, AND POST-TRANSLATIONAL MODIFICATION ON DESULFURIZATION ENZYME 2'-HYDROXYBIPHENYL-2-SULFINATE DESULFINASE (*DSZB*) STABILITY AND ACTIVITY" (2019). *Theses and Dissertations--Chemical and Materials Engineering*. 105.
https://uknowledge.uky.edu/cme_etds/105

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.

Landon C. Mills, Student

Dr. Christina M. Payne, Major Professor

Dr. Steve Rankin, Director of Graduate Studies

IMPACT OF CONFORMATIONAL CHANGE, SOLVATION ENVIRONMENT, AND POST-TRANSLATIONAL MODIFICATION ON DESULFURIZATION ENZYME 2'-HYDROXYBIPHENYL-2-SULFINATE DESULFINASE (*DSZB*) STABILITY AND ACTIVITY

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

Landon Clay Mills

Lexington, Kentucky

Co-Directors: Dr. Christina M. Payne, Adjunct Associate Professor of Chemical and Materials Engineering,

and Dr. Derek Englert, Assistant Professor of Chemical and Materials Engineering,

Lexington, Kentucky

Copyright © Landon Clay Mills 2019

Abstract of Dissertation

IMPACT OF CONFORMATIONAL CHANGE, SOLVATION ENVIRONMENT, AND POST-TRANSLATIONAL MODIFICATION ON DESULFURIZATION ENZYME 2'-HYDROXYBIPHENYL-2-SULFINATE DESULFINASE (*DSZB*) STABILITY AND ACTIVITY

Naturally occurring enzymatic pathways enable highly specific, rapid thiophenic sulfur cleavage occurring at ambient temperature and pressure, which may be harnessed for the desulfurization of petroleum-based fuel. One pathway found in bacteria is a four-step catabolic pathway (the 4S pathway) converting dibenzothiophene (DBT), a common crude oil contaminant, into 2-hydroxybiphenyl (HBP) without disrupting the carbon-carbon bonds. 2’-Hydroxybiphenyl-2-sulfinate desulfinase (*DszB*), the rate-limiting enzyme in the enzyme cascade, is capable of selectively cleaving carbon-sulfur bonds. Accordingly, understanding the molecular mechanisms of *DszB* activity may enable development of the cascade as industrial biotechnology. Based on crystallographic evidence, we hypothesized that *DszB* undergoes an active site conformational change associated with the catalytic mechanism. Moreover, we anticipated this conformational change is responsible, in part, for enhancing product inhibition. *Rhodococcus erythropolis* IGTS8 *DszB* was recombinantly produced in *Escherichia coli* BL21 and purified to test these hypotheses. Activity and the resulting conformational change of *DszB* in the presence of HBP were evaluated. The activity of recombinant *DszB* was comparable to the natively expressed enzyme and was competitively inhibited by the product, HBP. Using circular dichroism, global changes in *DszB* conformation were monitored in response to HBP concentration, which indicated that both product and substrate produced similar structural changes. Molecular dynamics (MD) simulations and free energy perturbation with Hamiltonian replica exchange molecular dynamics (FEP/λ-REMD) calculations were used

to investigate the molecular-level phenomena underlying the connection between conformation change and kinetic inhibition. In addition to the HBP, MD simulations of *DszB* bound to common, yet structurally diverse, crude oil contaminates 2’2-biphenol (BIPH), 1,8-naphthosultam (NTAM), 2-biphenyl carboxylic acid (BCA), and 1,8-naphthosultone (NAPO) were performed. Analysis of the simulation trajectories, including root mean square fluctuation (RMSF), center of mass (COM) distances, and strength of nonbonded interactions, when compared with FEP/λ-REMD calculations of ligand binding free energy, showed excellent agreement with experimentally determined inhibition constants. Together, the results show that a combination of a molecule’s hydrophobicity and nonspecific interactions with nearby functional groups contribute to a competitively inhibitive mechanism that locks *DszB* in a closed conformation and precludes substrate access to the active site.

Limitations in DszB’s potential applications in industrial sulfur fixation are not limited to turnover rate. To better characterize DszB stability and to gain insight into ways by which to extend lifetime, as well as to pave the way for future studies in inhibition regulation, we evaluated the basic thermal and kinetic stability of DszB in a variety of solvation environments. Thermal stability of DszB was measured in a wide range of different commercially available buffer additives using differential scanning fluorimetry (DSF) to quickly identify favorable changes in protein melting point. Additionally, a fluorescent kinetic assay was employed to investigate DszB reaction rate over a 48 hr time period in a more focused group of buffer to link thermal stability to DszB life-time. Results indicate a concerningly poor short-term stability of DszB, with an extreme preference for select osmolyte buffer additives that only moderately curbed this effect. This necessitates

a means of stability improvement beyond alteration of solvation environment. To this end, a more general investigation of glycosylation and its impact on protein stability was performed.

Post-translational modification of proteins occurs in organisms from all kingdoms life, with glycosylation being among the most prevalent of amendments. The types of glycans attached differ greatly by organism but can be generally described as protein-attached carbohydrate chains of variable lengths and degrees of branching. With great diversity in structure, glycosylation serves numerous biological functions, including signaling, recognition, folding, and stability. While it is understood that glycans fulfill a variety of important roles, structural and biochemical characterization of even common motifs and preferred rotamers is incomplete. To better understand glycan structure, particularly their relevance to protein stability, we modeled and computed the solvation free energy of 13 common N- and O-linked glycans in a variety of conformations using thermodynamic integration. N-linked glycans were modeled in the β-1,4-linked conformation, attached to an asparagine analog, while O-linked glycans were each modeled in both the α-1,4 and β-1,4-linked conformations attached to both serine and threonine analogs. Results indicate a strong preference for the β conformation and show a synergistic effect of branching on glycan solubility. Our results serve as a library of solvation free energies for fundamental glycan building blocks to enhance understanding of more complex protein-carbohydrate structures moving forward.

KEYWORDS: Biodesulfurization, DszB, molecular dynamics, protein engineering

Landon Mills
July 10, 2019

IMPACT OF CONFORMATIONAL CHANGE, SOLVATION ENVIRONMENT, AND POST-TRANSLATIONAL MODIFICATION ON DESULFURIZATION ENZYME 2'-HYDROXYBIPHENYL-2-SULFINATE DESULFINASE (*DSZB*) STABILITY AND ACTIVITY

By

Landon Clay Mills

Dr. Christina M. Payne

Co-Director of Dissertation

Dr. Derek Englert

Co-Director of Dissertation

Dr. Steve Rankin

Director of Graduate Studies

July 10th, 2018

Date

To my parents, without your continued love and support, none of this would have been possible. Thank you for being there with good food and warm hugs through all of the small victories and tiny challenges. You are the best family anyone could hope to have!

Acknowledgements

The road to writing this document has been long but wonderful. Over the past 5 years, I have experienced a number of successes, ten times as many failures, and have learned a lot along the way. None of this, however, would have been remotely possible without all the amazing people in my life helping along the way.

I cannot begin to express my thanks to my advisor, Dr. Payne, to whom I am deeply indebted. Your immense knowledge and relentless support are the lion's share of the reason I am able to write this dissertation. You remained with me despite your own career changes and continued to help guide my research and career decisions, affording me educational opportunities literally halfway around the world, all while remaining willing to laugh along with me at my numerous mistakes. Your advice has been invaluable to my research and beyond. You show me by example each day what a good scientist should be.

Thank you to Dr. Englert for your willingness to advise despite the distance between us. You were even gracious enough to let me stay in your home when I needed to work on the Paducah campus and treated me to a home cooked meal. Your assistance and patience have been much appreciated.

A big thank you to my committee members, Dr. David Rodgers and Dr. Tom Dziubla, the role you all have played in guiding my research and teaching me has been critical. Dr. Dziubla, you taught me the fundamentals of thermodynamics from which much of this dissertation is built and have introduced me to a number of great board games throughout my time at the University of Kentucky. Your contributions to my learning and social life have both been significant. A special thanks to Dr. Rogers and Dr. Martin Chow at the

University of Kentucky Protein Core for helping fill the gaps my unique advising situation created. Your equipment and expertise were invaluable. Additionally, I'd like to thank Dr. Brad Berron; you provided me with lab space, equipment, and a team of fellow grad students to help me learn and get started.

A huge thanks are in order for my fellow lab mates. When minor programing errors cause major problems, you guys are the best to call on and complain with. Japheth, your skills with Python are border line black magic, and I cannot thank you enough for helping me transition from bench top to *in silico* studies. Additionally, thanks to Yue Yu. We spent literally two years writing that paper together, and you deserve far more than just a publication for putting up with me through it all. Thanks for the laughs and the lifelong friendships!

Also, to the unsung heroes who got me started on this path so many year ago, many thanks to my former teachers. You filled me with curiosity and revealed the opportunities that the world affords to those who ask questions. Thank you to Mr. Carmen for piquing my interest in science through your non-traditional methods and encouraging middle schoolers to embrace their quirkiness. Thank you, Mrs. Taylor, for instilling a love of chemistry that has flourished over time and to Mr. Newton and Mr. Tanner who, together, introduced me to my first programming languages. Thank you, Mrs. Sydnor (Miles), for teaching me how to type. I struggled to make my fingers work early on, but your continued patience and assistance helped develop a skill that I cannot imagine attempting to write this document without. Ms. Miller, I may not be a distinguished writer, but I am certainly far better thanks to you. I could not have dreamed of writing a dissertation without the

command of the grammar and mechanics of the English language that you imparted on me (or at least tried to impart anyway).

Finally, I would like to acknowledge the financial support and computational time provided by the Extreme Science and Engineering Discovery Environment (XSEDE) [National Science Foundation (NSF) grant number ACI-1548562], the donors of the American Chemical Society Petroleum Research Fund (53861-DNI4), the National Science Foundation (CBET-1552355), and the Center for Computational Sciences DLX cluster at the University of Kentucky.

Table of Contents

List of Tables

List of Figures

Chapter 1 – Introduction to Biodesulfurization

1.1. Sulfur in Liquid Transportation Fuels

Comprising the majority of energy utilized by the transportation sector in the United States, liquid petroleum-based fuels play a critical role in the mobility of people and products alike.[1] While many environmental concerns center around the use of crude oil-based liquid transportation fuels, their association with sulfur emissions ranks among the most significant.[2] Refinement of crude petroleum leads to the transfer of a variety of naturally present sulfur-containing molecules to the resulting fuel, where, upon use, combustion forms gaseous sulfur oxides (SOx). SOx compounds represent a significant threat to environment, as they are capable of forming sulfuric acid with atmospheric moisture, resulting in acid deposition (acid rain). Additionally, SOx poses a risk to human health, causing respiratory irritation and being classified as acutely toxic according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS).[3] Recent studies even imply a link between SOx exposure during pregnancy and low infant birth rate as well as preterm birth.[4]

The current industrial means of sulfur removal from refined fuel products, hydrodesulfurization, utilizes a sulfur-linked metal catalyst (usually molybdenum) to capture simple sulfurs (linear hydrocarbon bound sulfur as well as thiophene and its single ring derivatives) on its surface via the presence of anion vacancy.[5-6] Catalyst-bound sulfur-containing molecules are cleaved via hydrogenolysis leaving the sulfur atom bound in the previous anion vacancy for later release as hydrogen sulfide (H_2S) during catalysis regeneration (Figure 1.1).

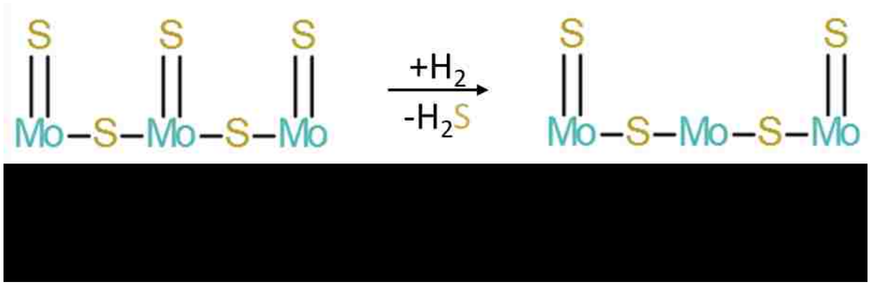

Figure 1.1. Hydrodesulfurization catalysis regeneration mechanism

Complex cyclic sulfur-containing molecules such as thiophene can be targeted, albeit non-specifically, as part of the hydrodesulfurization process, rendering the vast majority of the original sulfur contained in pre-processed crude oil catalyst-bound. Limitations with this process arise in the face of more complex aromatic sulfur molecules, primarily dibenzothiophene (DBT) and its derivatives (structure depicted in Figure 1.2). [7-8] While these DBTs account for only a small population of the total sulfur contained in crude oil, their impact on refinement is significant.

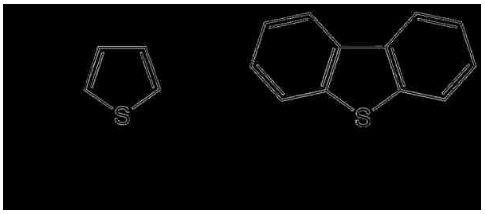

Figure 1.2. Common cyclic crude oil contaminants in liquid transportation fuels

DBT-bound sulfur is highly resistant to hydrodesulfurization and passes through industrial desulfurization processes virtually unimpeded, passing directly into product fuels; 60% of the total sulfur emissions from liquid transportation fuel combustion is attributed to DBT and its derivatives.[9-14] While the remaining cyclic sulfur in petroleum-based fuels represents only a small portion of the original total sulfur content, what remains is still quite detrimental when combusted, thus demonstrating a need for further more complete sulfur removal.

1.2. Biodesulfurization Research Motivation

Biodesulfurization, the enzymatic cleavage of sulfur atoms from larger host molecules, can be achieved through a variety of methods but principally occurs through the introduction of microorganisms into sulfur-rich environments or through the combination of isolated biocatalysts with a sulfur containing solution. In either case, this enzyme-driven process enables rapid, highly specific, and complete cleavage of otherwise difficult-to-access sulfur bonds.[5-6, 15-16]

Biodesulfurization reactions occur at near ambient temperatures and pressures, and the specificity for select substrates makes them ideal additive processes to complement current gaps in existing industrial systems. A variety of microorganisms share a similar catabolic pathway, aptly named the "4S pathway," for sulfur fixation (in sulfur-limited conditions) consisting of four steps and acting on DBT and DBT derivates to selectively cleave the central ring carbon-sulfur bonds.[17-19]

Considering the pathway specificity for what are currently the most prevalent liquid transportation fuel contaminants, in addition to the mild operating conditions associated with biological reactions, biodesulfurization via the 4S pathway presents an ideal solution to mitigate sulfur emissions.[15, 20] However, significant technological barriers exist which preclude direct industrial application of the 4S pathway as found in nature. The final and rate-limiting step of the pathway catalyzed by 2'-hydroxybiphenyl-2-sulfinate (HBPS) desulfinase (*DszB*) has a relatively low turnover number, experiences significant product inhibition, and poor stability (unfolding shortly after purification).[21-25] Most investigations into *DszB* seek to improve enzyme performance by directed evolution, increased enzyme expression, and/or alternative strain selection; however, characterization with the intent of rational design remains largely uninvestigated, and the precise mechanism governing *DszB*'s performance limitations remain largely unknown.[26-28] Accordingly, fundamental understanding of the molecular mechanisms of *DszB* inhibition and stability must be developed.

Improvement of the catalytic efficiency of *DszB* as well as reducing the effects of product inhibition are mandatory before the 4S pathway can be implemented in industrial biodesulfurization processes. By investigating conformational changes during catalysis and in response to the product, as well as how those changes relate to kinetic parameters, we develop a more complete understanding of the molecular mechanisms of *DszB*. Additionally, by observing changes in thermal stability and kinetic activity of *DszB* over time, we generate a more complete understanding of the catalytic lifetime of *DszB* and the means by which we may extend it.

1.3. Background

1.3.1. 4S Pathway

The mechanisms by which microorganisms consume and utilize sulfur vary considerably, but the fundamental ability to consume sulfur is shared by most all of the Bacteria kingdom.[29] Regardless of subtle differences between sulfur fixation approaches such as substrate preferences, reaction conditions, and metabolic phase, mechanistic approaches to cyclic sulfur removal principally occurs via one of two mechanisms: accessing the ring-bound sulfur by either disassembly of a the molecule in a "ring destruction" approach or in a minimal bond breaking, sulfur-specific "ring opening" approach.[30-32] In the case of the former ring destruction mechanism, access to sulfur is obtained by disruption of multiple bonds within in the ring to create a new geometry with a more accessible or exposed sulfur atom. As the name implies, this ring structure is destroyed the process. Given that much of a fuel’s heating value is derived from carbon-carbon bounds, the most ideal mechanistic approaches in the context of fuel treatment fall under the latter category. With sulfur-specific ring opening mechanisms, only bonds with the target sulfur atom are disrupted, leaving the ring largely intact, as depicted in the 4S pathway shown in Figure 1.3.[17]

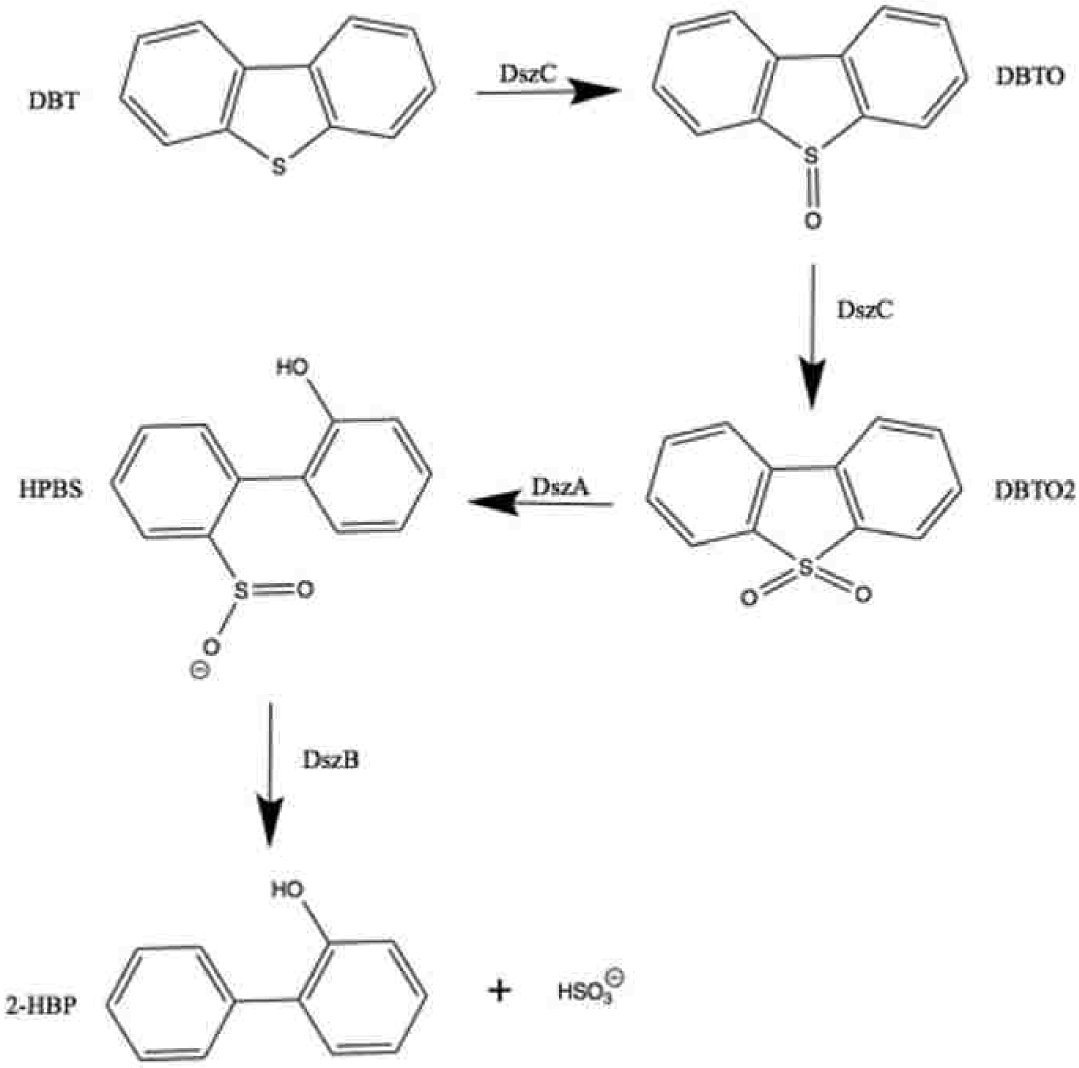

Figure 1.3. R. erythropolis 4S pathway desulfurization of DBT

Acting on DBT and DBT derivatives, the 4S pathway succeeds at removing the ring bound sulfur as sulfite, breaking only C-S bonds and leaving the heating value of the remainder of the molecule relatively undisturbed in the form of the pathway product 2-hydroxybiphenyl (HBP).[17-18] Many microorganisms utilize some form of the 4S pathway as a means of sulfur fixation in sulfur-limiting conditions, [33] but most research regarding the pathway is performed with the organism in which it was first isolated, *Rhodococcus erythropolis*, strain IGTS8.[34] In the case of non-native host production of 4S pathway proteins, the sequences from the IGTS8 are most commonly used. This focus on the original strain's enzymes can be attributed to more than a seminal standard metric. In most organisms in which the 4S pathway is expressed, diffusion of extracellular substrate through the cell membrane limits reaction rate far more than any individual pathway step.[1, 35-37] This limitation is not present in *R. erythropolis*, providing both a better starting point from the more popular research in directed evolution as well as a less complex means of directly assessing pathway performance.[1, 26, 38]

Four enzymes work in concert to achieve desulfurization in the 4S pathway. The monooxygenase, DszC, begins catalysis with oxidation of the still ring-bound sulfur of DBT. The subsequent step is a repetition of the first where, again, DszC catalyzes additional oxidation of the same sulfur atom. In the third step, a second monooxygenase, DszA, catalyzes the oxidative cleavage of the first carbon-sulfur bond, opening the DBT ring. While not depicted in Figure 1.3, the oxidoreductase, DszD, aided by flavin mononucleotide (FMN) and nicotinamide adenine dinucleotide (NAD) cofactors, provides the corresponding reduction to balance the three oxidation steps that begin the 4S pathway. The final and rate-limiting step catalyzed by desulfinase *DszB* results in the hydrolytic cleavage of the second carbon-sulfur bond, liberating the sulfur in the form of water soluble sulfite.[17, 39]

1.3.2. Product Inhibition

The impact of product, HBP, on the 4S pathway is well known, with recent studies reporting its inhibiting effect on each enzyme in the pathway to at least some degree; [40] [25, 35] however, the general mechanisms by which this inhibition is achieved and the corresponding relevance to the regulation of and/or evolutionary significance to the 4S pathway has yet to be conducted.[24, 40-42] The lack of significant characterization of isolated proteins within the 4S pathway limit initial understanding of HBP's inhibitive potential, as evaluations of this phenomenon *in vivo* are obscured by periplasmic and/or membrane retention of HBP (as well as DBT) due to its hydrophobic nature.[1]

Solvent environment also plays a crucial role in observed inhibition, as the presence of an oil-water interface affects clearance/availability of HBP. Biphasic system studies performed on whole cell 4S pathway systems reveal that for organisms such as the IGTS8 strain, i.e., capable of acting at the oil-water interface to remove hydrophobic molecules from the oil phase, product inhibition was stifled by the presence of an oil phase for concentrations up to 140 μM.[1] Outside of these biphasic studies, attempts to reduce the effects of product inhibition of *DszB* are sparsely documented in literature, with no attempts to execute site directed mutagenesis.

1.3.3. Crystallographic Study of DszB

Solved by x-ray crystallography with resolutions ranging from 1.6 Å to 1.8 Å, three structures of *DszB* have been deposited into the Protein Data Base (PDB) and are shown in Figure 1.4.[43] The first of these structures depict an unbound wild-type *DszB* from the IGTS8 strain. The remaining two were obtained from a catalytically-inactive mutants bound to two different substrates, HBPS and 2-biphenyl sulfinic acid (BPS). The structure of primary interest shows HBPS-bound *DszB* (2DE3), though a near identical structure (2DE4) illustrates *DszB* bound to an alternate substrate, 2-biphenyl sulfinic acid. 2-biphenyl sulfinic acid has been confirmed to also elicit *DszB* activity/HBP formation in kinetic studies.[44]

Cys27 was confirmed as the catalytic amino acid, and C27S mutations were used to create the catalytically-inactive mutants used to obtain substrate bound structures. Of additional note, mutations to the His60 residue produced a ~17-fold reduction in *DszB* turn-over number.[43] Further investigation of the critical His60 residue's position in the crystal structure provides rationale for its importance in the activity of *DszB*. While unbound, His60 is contained in an unstructured loop on the surface of *DszB*; however, when bound to substrate, the loop folds into a structured α-helix with His60 buried in the active site, where it coordinates the substrate along with Ser27 (what would be Cys27 in the wild type form). The mobile His60 is displayed in stick form in all three structures presented in Figure 1.4. Collectively, crystallographic data implies that a large conformation occurs in this loop and is involved in the catalytic activity of *DszB*.[43, 45]

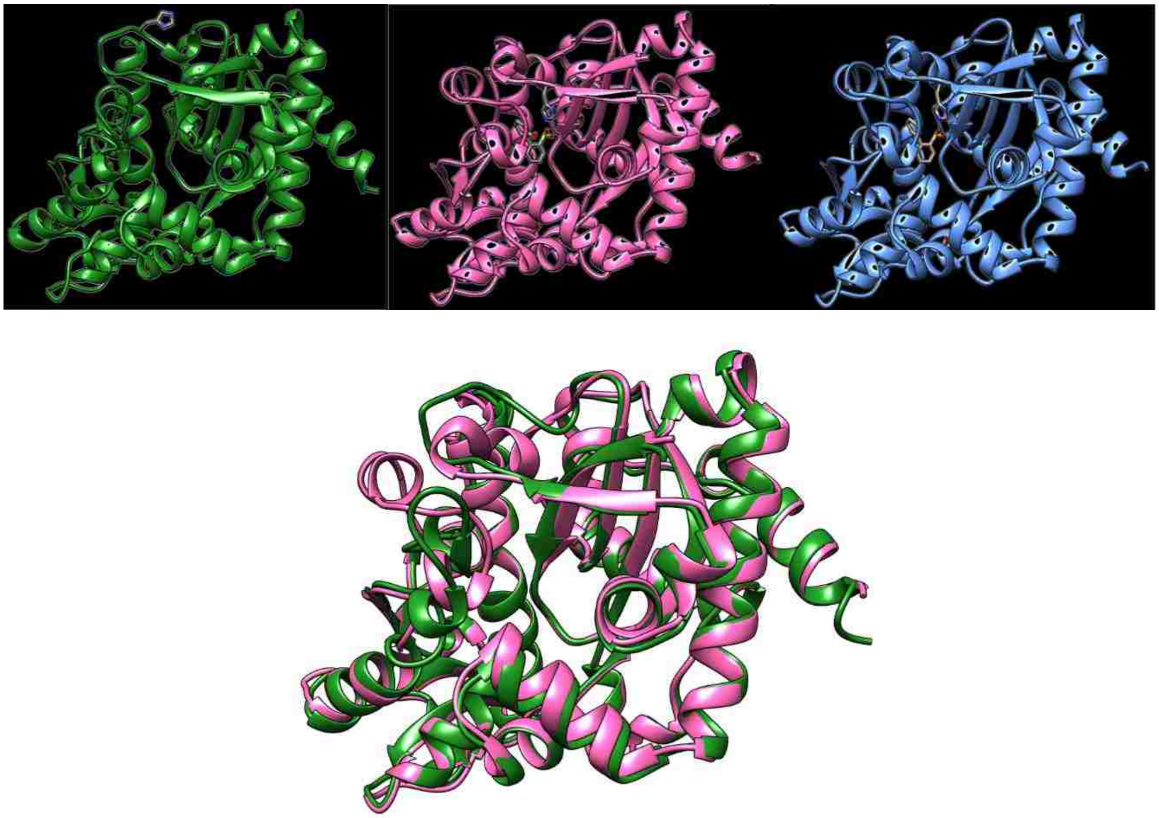

Figure 1.4. Crystal Structures of DszB

PDB Entry 2DE2: Unbound DszB [green], PDB Entry 2DE3: C27S mutated substrate (HBPS) bound DszB [pink], PDB Entry 2DE4: C27S mutated alternate substrate (2-biphenyl sulfinic acid) bound DszB [blue], Overlay of 2DE2 and 2DE3 shown enlarged

1.3.4. Desulfurization Mechanism

Two prominent *DszB* desulfurization mechanisms have been proposed, including nucleophilic addition and electrophilic aromatic substitution.[43, 46] HBPS positioned in the active site relative to local residues of interest as determined by x-ray crystallography is displayed in Figure 1.5 for reference.

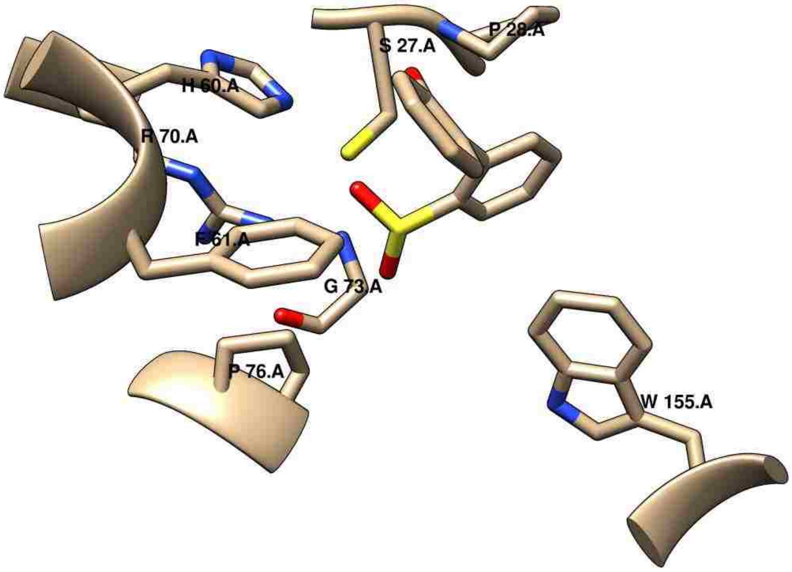

Figure 1.5. Crystallographicly determined HBPS bind position in DszB active site

In the nucleophilic addition mechanism proposed by Lee *et al.*, after having moved into the active site, His60 coordinates the substrate such that the sulfate group from HBPS is in proximity to Cys27. This causes deprotonation of sulfur from Cys27 to form a nucleophile through a substrate-assisted catalysis. Simultaneously, the sulfur nucleophile attacks the HBPS sulfur to form a disulfide bond. The resulting charged state is stabilized by both the guanidinium from Arg70, and the backbone amine from Gly73. The resulting sulfinate formed after cleavage of the carbon sulfur bond is then released as sulfite with a water molecule, completing the reaction.[43]

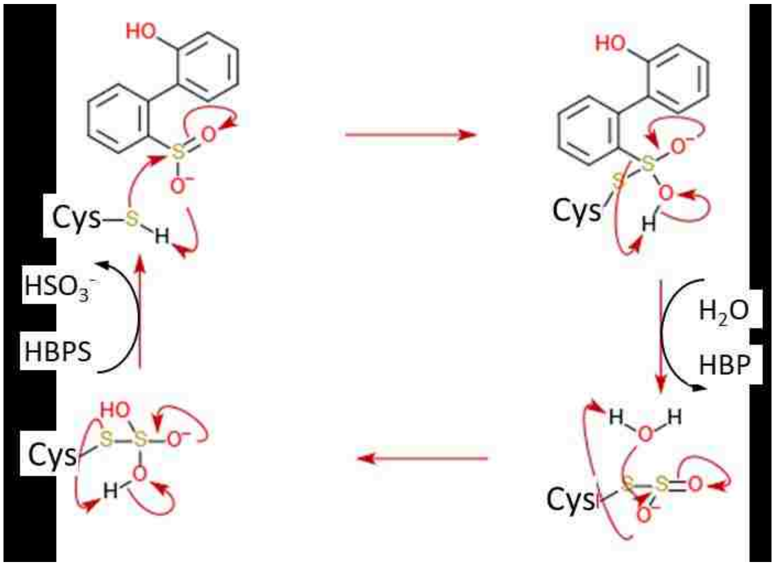

Figure 1.6. Putative neucleophilic addition mechanism

Mechanism proposed based on first crystallographic evidence, adapted from Geronimo et al. [46]

Recent density functional theory studies of the *DszB* mechanism, however, reassert the more probable desulfurization mechanism occurs via electrophilic aromatic substitution, originally proposed by Gray *et al.* (shown in Figure 1.7).[39, 46] In this mechanism, Cys27 serves as a proton donor rather than a nucleophile behaving similarly to tyrosine phenol-lyase, [47] hydrolytically cleaving sulfur in the form of SO_2 where subsequent reaction with water from the surrounding aqueous environment forms HSO_3^-.

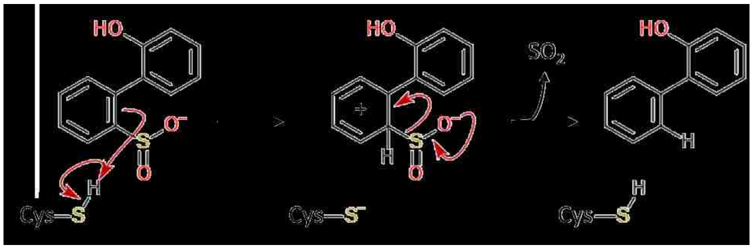

Figure 1.7. Alternate electrophilic aromatic substitution DszB reaction mechanism.

Mechanism originally proposed by Gray et al, adapted from Geronimo et al. [46]

Recent molecular dynamics (MD) simulations further support the proposed electrophilic aromatic substitution mechanism, confirming the significance of His60 in stabilizing the transition state through hydrogen bonding and charge interactions with both substrate sulfur and Cys27.[46] Simulations went on to emphasize the importance of other active site residues in the creation of a unique microenvironment, indicating that the charge state of R70 was critical to reducing activation energy required to perform catalysis. This is consistent with inactive mutations observed in mutations to the highly conserved H60 and R70 residues and further supports the proposed electrophilic aromatic substitution mechanism.[46]

1.4. *DszB* Characterization Outline

The theme of this dissertation, as it pertains to biodesulfurization, is to deepen understanding of the natural limitations of the rate-limiting enzyme of the 4S pathway in that hope that future research in industrial desulfurization may benefit from a more complete characterization of the *DszB* enzyme. [25, 48-49] To this end, we focus our investigation on three distinct objectives, discussed in greater detail in each of the corresponding chapters:

1.4.1. Evaluate the general mechanisms of DszB product inhibition (Chapter 4)

The 4S pathway product HBP is known to have an inhibitory effect on *DszB,* but the exact mechanism by which it deactivates the enzyme requires clarification.[50] Chapter 4 experimentally demonstrates the kinetic profile of *DszB* in response to HBP, elucidating the class of inhibition. Computational studies then expand determination of inhibitor affinity for *DszB* to include a larger array of common crude oil contaminants.

1.4.2. Determine the effects of HBP on DszB conformation change (Chapter 4)

Linking kinetic properties to structural changes in *DszB* generates a more complete understanding of how specific residue groups influence enzyme activity. Chapter 4 experimentally investigates the conformational response of *DszB* to HBP and extends this research focus to encompass a larger computational investigation into *DszB* structural changes in the presence of multiple known inhibitors.

1.4.3. Assess the kinetic stability of DszB in various environments (Chapter 5)

Understanding the lifetime of *DszB* activity and how solvation environment influences prolonged activity is imperative to developing robust biodesulfurization technology. Chapter 5 characterizes the kinetic stability of *DszB* and investigates the impact of a wide variety of solvation environments on its catalytic lifetime.

Chapter 2 – Introduction to Glycosylation

2.1. Glycosylation as a Post-translational Modification

Proteins across then kingdoms of life rely on post-translational modifications to achieve further diversity of function in order to fulfill their biological roles. Post-translational modifications are linked covalently and a fulfill a variety of purposes with an assortment of attachments, ranging from the common addition of phosphate groups to regulate enzyme activity all the way to lipid attachments to adhere proteins to a membrane. Of the many types of such modifications, glycosylation ranks among the most prevalent and diverse, appearing in each kingdom of life to at least some degree with seemingly countless variations.[51] Typically beginning in the endoplasmic reticulum (ER), or simply intercellularly in the case of prokaryotic organisms, glycosylation is achieved by covalently attaching a monomeric sugar or polysaccharide chain, referred to as a "glycan," to a residue on the resulting glycoprotein complex (Figure 2.1) via either an O-linkage to serine or threonine or an N-linkage to asparagine.[51-53] In higher order organisms, glycosylation may occur at multiple sites on the same glycoprotein, and, often, initially monomeric glycans attachments are continuously glycosylated to build longer, more complex polymeric chains.

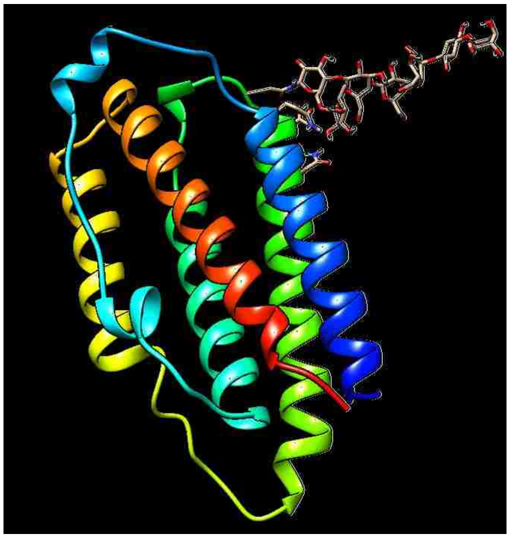

Figure 2.1. Crystal structure of an example glycoprotein complex, human interferon-β [54]

Due to their extremely diverse structure, glycans are able to serve a variety of functions in nature, including signaling, recognition, folding, and stability. In some cases, the effects they are capable of imparting on the final glycoprotein complex can be quite profound. In the case of human interferon-β (depicted in Figure 2.1), a therapeutic glycoprotein used in the treatment of multiple sclerosis, changes in the sequence of the relatively small glycan attached have profound effects on protein activity and stability.[55-58] Studies relating the impact of glycoform on physicochemical and biological properties of human interferon-β reveal much performance variability can be attributed to the degree of sialylation via glycan building block N-acetylneuraminic acid (Neu5Ac). [58] While many of human interferon-β glycoforms begin with a standard N-glycan N-acetyl-glucosamine initiating complex subsequent branching and sialylation vary. Disialylated glycoforms showed greater thermal stability with melt temperatures 1-2 °C higher than Tri and Mono sialylated counter parts. Glycoforms also demonstrated significant differences in binding multiple different receptors.

While human interferon-β may fall nearer to the extreme end of the spectrum regarding a glycan's capacity to impact protein performance, it is by no means the only such protein to rely heavily on specific glycoforms to function.[51]

2.2. Glycan Research Motivation

Proteins are considered an extremely diverse and dynamic biomolecule, performing an array of critical functions within living organisms; however, their respective base structural variability pales in comparison to that of carbohydrates. In the case of proteins, macromolecules are assembled from a relatively small pool of amino acid building blocks linked linearly in a chain through consecutive and fairly rigid peptide bonds. When considering glycans, many more arrangements arise. Assembled from a collection of molecules potentially including any bioavailable monomeric sugar, even relatively short glycans already carry the potential of permutations far exceeding peptides of equivalent length. Additional diversity arises from the manner in which each monomeric addition to a glycan chain can be linked. Glycosidic bonds connecting each new sugar can occur in an α or β conformation, altering the shape and stability of the resulting glycan, and can be linked in one of several positions along to the existing glycan chain, typically to the 2nd, 3rd, 4th, or 6th carbon in a given sugar.[59] Figure 2.2 demonstrates the versatility in connections present in glycans and shows several common linkages.

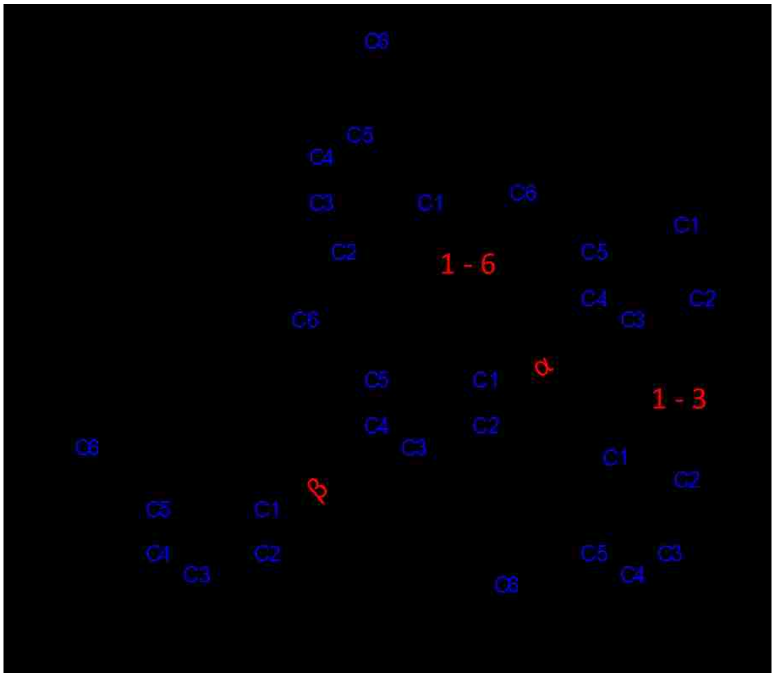

Figure 2.2. Structual examples of glycosidic linkages

Apart from individual linkages, glycans commonly adopt a branched structure in which multiple monomeric sugars are added to the same glycan terminal, providing the single molecule with multiple strands to elongate and build upon.[51] With all this in consideration, the “primary structure” equivalent quickly exceeds the variety observed in proteins even for relatively small molecules. Still further structural variability can be found when considering the glycosidic linkages connecting each glycan are far more dynamic than a peptide bond, allowing for more rotation, producing a unfathomable number of unique rotamers among a single primary sequence. [51] With so many possibilities in both the chemical properties of the various functionalized sugars that comprise them, as well as the size and structure glycans adapt, glycoproteins are particularly challenging to study.

Given the many possible forms glycans can take and the variability between otherwise identical proteins, an understanding of how specific glycans alter host protein properties is critical.[60-61] While many separate studies into independent glycans or glycoproteins have been performed, more general data regarding how particular glycan building blocks or recurring motifs influence their complexed protein is more limited. While glycoproteins' capacity to affect signaling pathways, enzyme activity, molecule recognition may be too situationally specific to discussion in universal terms, their impact on protein stability, an issue of particular interest in the commercial production of enzymes (particularly as pharmaceuticals), can be measured and expressed in more globally applicable terms not limited by specific glycan to protein interactions.[62-63]

2.3. Background

2.3.1. Glycosidic Linkage Conformation in Oligosaccharides

When considering the general nature of oligosaccharides, glycosidic linkages in the β conformation form more energetically favorable complexes due their more linear form limiting the potential for steric clashes between accompanying functional groups along each pyranose ring as well as better positioning neighboring hydroxyl groups from both of the monomeric sugars for more optimal hydrogen bonding to other cellulose polymers. This phenomenon is observed in nature to extreme effect in the plant cell wall component cellulose, which combines the stabilizing benefit of consecutive β linkages between glucose molecules at opposing ends of each sugar through 1,4 linkages (shown in Figure 2.3).

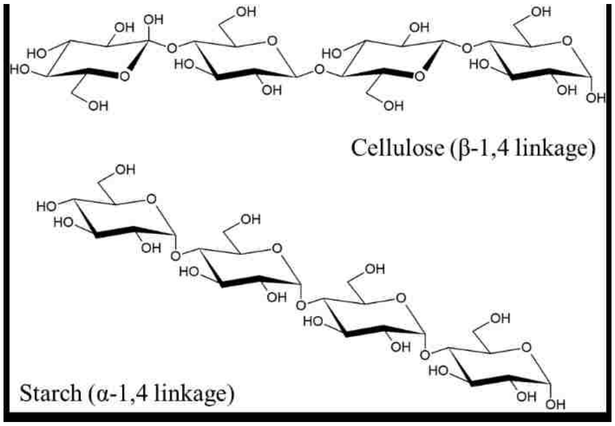

Figure 2.3. Chemical structures of cellulose and starch

The end result is an oligosaccharide assembly so stable it is indigestible by many organisms do to its incredibly long form optimally aligned for hydrogen bonding and linear structure build to avoid folding in on itself and generating steric clash. In contrast, starch (Figure 2.3), another oligosaccharide formed from glucose bound, instead, by α-1,4 linkages is considerably more vulnerable to enzymatic attack and is almost universally digestible among the animal kingdom. The bend introduced in its structure from glycosidic α bonds reduces hydrogen bonding, increases steric hindrance, and leaves linkages more exposed for enzymatic attack. These structural principles observed in simple oligosaccharides can be easily extrapolated to describe the significance of glycosidic linkage conformation in glycans in general.

2.3.2. Common O-Glycosylation in Nature

While protein linked glycans (especially O-linked) exhibit considerably less diversity than glycans as a whole, such post translational modifications still employ a wide range of bioavailable sugar monomers. Many naturally occurring glycoproteins employ only short on monomeric sugar additions, though lengthy chains with branching and variable linkage conformations are still observed, especially in higher order organisms.

Prokaryotes

Glycosylation in prokaryotic organisms occurs within the cytosol via both sequential and simultaneous pathways, meaning that glycans may be build either monomerically piecewise by repetitive attachments to the glycoprotein, or constructed independently from the protein and attached upon completion. [51] A broad range of monosaccharide donors exist throughout prokaryotic kingdoms and vary from organism to organism void of any recurring motifs or globally utilized molecules. Among the more commonly utilized glycans, glucose, galactose, and N-acetyl-galactosamine are commonly bonded to serine and threonine residues via simultaneous addition. [64]

Plantae

Possessing unique O-glycosylation mechanisms, no evolutionary trends are observed in ER-produced glycoproteins. A common fundamental pattern among plant glycosylation shows a preference for targeting the less common hydroxyproline residues and occasionally neighboring serine for glycosylation.[65] Cytosolic glycosylation is common, utilizing N-acetyl-glucosamine and fucose O-linked to the more common serine and threonine residues.[51]

Fungi

Fungi, especially yeast, show almost no diversity in O-glycosylation, utilizing mannose almost exclusively.[51] While galactose and xylose may be employed in some organisms, heavy glycosylation with mannose remains a prevalent theme. Glycans can be comprised of both linear mannose chains and branched structures.[65-67]

Animalia

Highly diverse and increasingly more complex among the higher-order organisms within the kingdom, many O-glycosylation motifs are conserved in animals. Predominantly initiated in the ER, with some organisms initiating select glycosylation in the Golgi apparatus, animals utilize mannose, glucose, galactose, N-acetyl-glucosamine, N-acetyl-galactosamine, xylose, and fucose commonly, with chain elongation being extremely prevalent. Fucose, glucose, and N-acetyl-glucosamine are employed in epidermal growth factor (EGF)-like protein domains as a stabilizing force that is critical for function but not well understood.[51, 68] Galactose glycosylation is used in extra cellular matrix proteins such as collagen, and xylose is applied heavily to proteoglycans as an initial monomer for further and diverse chain elongation. [17]

2.3.3. Natural N-Glycosylation Motif

Organisms across the kingdoms of life utilize a common initiating N-glycan motif which is often further modified for specific function thereafter. [69] Attached at an asparagine residue, this base glycan is assembled separately from the protein and applied afterward en bloc. The motif, depicted in standard block glycan notation below consists of two β-1,4 linked N-acetyl-glucosamines attached to a mannose that is then further elongated and branched by a second and third mannose linked through an α-1,3 and α-1,6 bond, respectively. [62, 69-72]

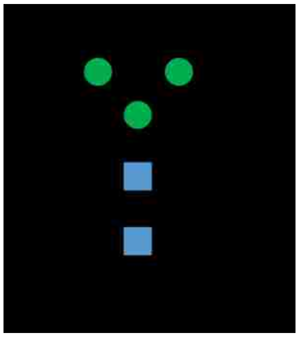

Figure 2.4. Symbolic representation of fundumental N-glycan motif

Symbol nomenclature based on scheme developed in "Essentials of Glycobiology. 2nd edition:" [51] *N-acetyl-glucosamines (blue square); Mannose (green circle); α and β linkage conformations are as noted.*

Following construction and attachment in the ER, the resulting glycoprotein is moved to the Golgi apparatus for additional functionalization of the base motif. The most common additions include: (1) a continual addition of mannose with introduction of new branching points, (2) unique and diverse glycan chains usually beginning with the addition of an N-acetyl-glucosamines to each of the two terminal mannose, and (3) a blending of the first two options in which one of the branched mannose is elongated further with more mannose, and the other is bound with N-acetyl-glucosamine and primed for a unique glycan chain.[62, 69-72]

2.4. Glycan Characterization Outline

The theme of this dissertation, as it applies to glycosylation, is to better characterize their impact on protein stability. We examined the role of several common O- and N-linked glycan motifs in stabilizing protein structures, focusing on simple and single glycoforms to build fundamental understanding that may be extrapolated to more complex glycans. To this end, we focused our investigation on two distinct objectives, discussed in greater detail in each of the corresponding chapters:

2.4.1. Evaluate the impact of monomer diversity and protein-glycan bond conformation on stability of resulting glycoprotein complex (Chapter 6)

Glycans are composed of a variety of monomeric building blocks and linked to various residues in one of two conformations. Chapter 6 investigates, computationally, how these factors impacts the aqueous stability of the glycoprotein complex and references naturally occurring glycans to form hypotheses as to why each may be employed.

2.4.2. Assess the effects of glycan chain elongation and branching present in most naturally occurring N-glycans on glycoprotein stability (Chapter 6)

Many N-glycan motifs are generated from a common base structure and elongated to form more elaborate glycans. Chapter 6 elucidates the effect of chain elongation and branching on glycoprotein stability with molecular dynamics simulations.

Chapter 3 – Methods

3.1. Materials

Electrophoresis gels and reagents as well as immobilized metal affinity column (IMAC) packing were purchased from Millipore Sigma’s TruPAGE™ line. Plasmid preparation enzymes and reagents were purchased from IBI Scientific and NEB as specified throughout this chapter. Select reagents and standards were purchased through ThermoFisher Scientific and Hampton Research. All other chemicals and materials without specifically stated origins were purchased from VWR.

3.2. Protein Expression and Purification

The contents of this section describe the preparation of the *DszB* plasmid and subsequent expression and purification of *DszB* from an *Escherichia coli* host. The schematic below shown in Figure 3.1 briefly outlines the process, with greater detail provided in subsequent subsections.

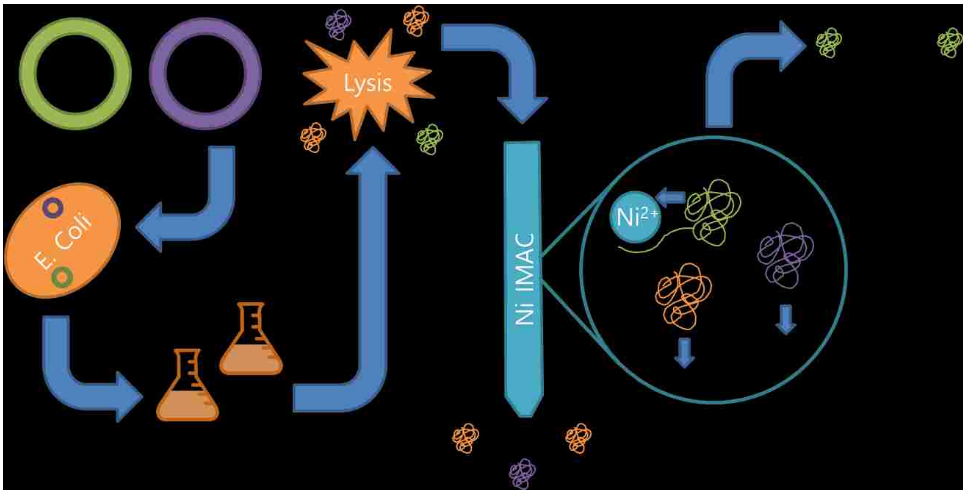

Figure 3.1. Expression and IMAC purifcation of MAT-Tag DszB

3.2.1. Preparation of DszB Plasmid

Genomic DNA from bacterial strain *Rhodococcus erythropolis* IGTS8 (ATCC 53968) was extracted using a gBAC Mini DNA Bacteria Kit (IBI Scientific). The *dszB* gene from the genomic DNA was amplified by Polymerase Chain Reaction (PCR), using primers dszB-forward-HindIII and dszB-reverse-SphI. dszB DNA fragments were purified with DNA Fragments Extraction Kit (IBI Scientific) and digested with restriction enzymes HindIII and SphI simultaneously. Digestion product was isolated via agarose gel electrophoresis and purified via DNA Fragments Extraction Kit (IBI Scientific). A MilliporeSigma pTAC-MAT-2 plasmid was prepared for recombination in a similar fashion, simultaneously digested with restriction enzymes HindIII and SphI. Digested plasmid was also isolated via agarose gel electrophoresis and purified via DNA Fragments Extraction Kit (IBI Scientific). Ligation between purified digestion products occurred at a 3:1 insert to vector ratio using T4 DNA Ligase (NEB) to form working plasmid depicted in Figure 3.2. Sequence conformation was performed by a third party via Eurofins Genomics overnight sequencing.

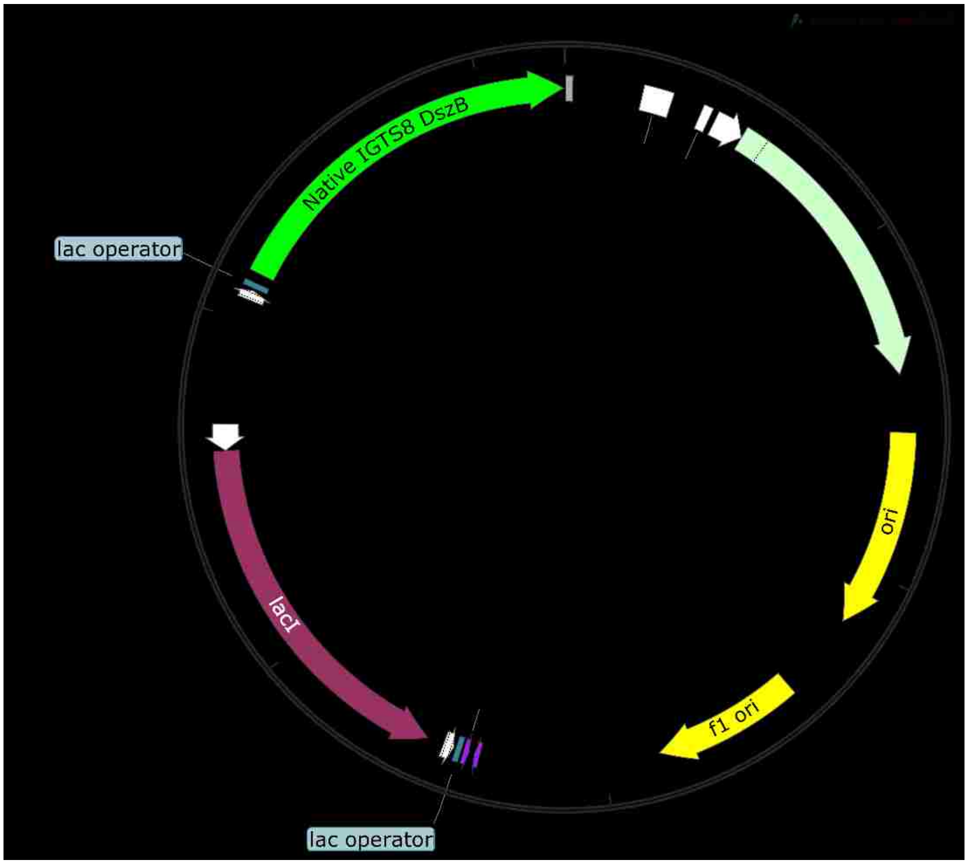

Figure 3.2. Sigma pTAC-MAT plasmid vector with DszB insert

3.2.2. E. coli Transformation

Ligation product was transformed into electrocompetent *E. coli* (NEB5-alpha) via electroporation using a Gene Pulser Xcell (Bio-Rad). Resulting bacteria was then plated on lysogeny broth (LB) (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) agar containing 100 µg/mL ampicillin and allowed to grow. Individual plate colonies were then selected to seed 10 ml LB media (w/ 100 µg/ml ampicillin) cultures for plasmid production. The *DszB* production plasmid (pTAC-MAT-Native-IGTS8-*DszB*) was then harvested using High-speed Plasmid Mini Kit (ISI Scientific). The above process was then repeated using groES/groEL-producing plasmid pG-KJE8 in place of ligation product and 50 µg/ml chloramphenicol in place of 100 µg/ml ampicillin, producing pG-KJE8 and pTAC-MAT-Native-IGTS8-*DszB* concentrate.

Electrocompetent *E. coli* (BL21) was then transformed again via electroporation using a Gene Pulser Xcell (Bio-Rad) with plasmid pG-KJE8 and cultured in LB agar containing 50 µg/ml chloramphenicol. Single colonies were selected and cultured in 10 ml of LB media (w/ 50 µg/ml chloramphenicol). The product culture was then made electrocompetent for secondary transformation. Electrocompetency was restored by chilling the culture to 4 °C and pelleting its cells for media exchange via centrifugation at 1000 x g for 20 min. The resulting supernatant was decanted and cells were resuspended in ice cold sterile 10% glycerol in deionized water. Media exchange with glycerol was repeated 3 times to insure complete removal of LB media components. The final cell pellet was resuspended in 10% glycerol.

Electrocompetent *E. coli* (BL21) with pG-KJE8 plasmid was then transformed a second time with pTAC-MAT-Native-IGTS8-*DszB*. Again, resulting cells were plated on LB agar, now containing both 50 µg/ml chloramphenicol and ampicillin 100 µg/ml. Resulting colonies were used as inoculant for *DszB* production in section 3.2.3.

3.2.3. Cell Culture and Protein Expression

10 ml of (LB) media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) loaded with 50 µg/ml chloramphenicol and ampicillin 100 µg/ml antibiotics was inoculated with a single colony of BL21 *E. coli* as prepared in section 3.2.2. The inoculum was incubated at 37 °C overnight on a New Brunswick Incubator Shaker under agitation at 175 rpm in an Erlenmeyer flask. The resulting culture was then combined with an additional 2 L of LB

media with equivalent antibiotic levels (50 µg/ml chloramphenicol and ampicillin 100 µg/ml] and allowed to grow under identical incubation conditions until its optical density at 600nm (OD600) exceeded 0.8. Afterward, expression of the groES/groEL molecular chaperons was induced by adding arabinose, and co-expression of *DszB* was induced by adding 2 ml of 1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG). Culture temperature was reduced to room temperature and allowed to express overnight.

3.2.4. Cell Lysis

2 L of mature *E. coli* culture were collected after expression, as described in section 3.2.3, and centrifuged at 5,000 x g for 30 min. The cell pellet was resuspended in 40 ml of equilibration buffer (0.3 M NaCl, 50 mM Na_2HPO_4, 10 mM imidazole) and chilled on ice for 10 min. Cell lysis was performed on the concentrated cell pellet via sonication at 50% amplitude in ice over 10 cycles of 10 sec sonication bursts followed by 30 sec cool times. Insoluble cell pellet remnants were removed via centrifugation for 30 min at 32,000 x g chilled to 4 ^{o}C. The resulting lysate supernatant was then collected for IMAC purification.

3.2.5. IMAC

DszB was purified from lysate as generated in section 3.2.4 via IMAC. Approximately 20 ml of a Millipore Sigma Ni IMAC- select affinity gel for purifying MAT-Tag proteins was poured into a glass gravity fed drip column. Storage medium was washed from the column packing with five column volumes of ultra-pure deionized water. The column was then equilibrated by a wash with five column volumes of equilibration buffer (0.3 M NaCl, 50 mM Na_2HPO_4, 10mM imidazole). Centrifuged lysate from section 3.2.4 was then loaded into the column and allowed to flow through completely allowing protein column binding. The column was washed with an additional five column volumes of equilibration buffer to remove residual non-tagged proteins. Finally, purified *DszB* was eluted from the column with 50 ml of elution buffer (0.3 M NaCl, 50 mM Na_2HPO_4, 200 mM imidazole). elute was collected in 5 ml fractions to ensure optimal purity.

3.2.6. Bradford Assay

Protein concentration was assessed by Bradford Assay. A calibration curve to determine the relationship between protein concentration and absorbance was generated from standards of bovine serum albumin (BSA) suspended in column elution buffer (0.3 M NaCl, 50 mM Na_2HPO_4, 200 mM imidazole). BSA stock solutions were created with concentrations of 5, 1, 0.5, and 0.1 mg/ml. 100 µl of each solution along with a blank containing only elution buffer was loaded in triplicate into a black, clear-bottom, 96-well tray for analysis. Additionally, 100 µl of each protein sample subject to concentration testing was also loaded in triplicate. 100 µl of Coomassie Bradford assay reagent was then added to each well, and the plate was loaded into a BioTek Synergy plate reader set to shake internally for 2 min. The absorbance of each well was measured at 595 nm wavelength. The absorbance of BSA standards was fitted to a linear calibration curve, which was subsequently used to determine protein concentration in samples tested.

3.2.7. SDS-PAGE

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was used to assess protein purity. Protein samples were gathered during execution of methods outlined in section 3.2.5 and include resuspended cell pellet post lysis, 1:10 dilute lysate, column wash flow through, and elution fractions of IMAC purified *DszB*. SDS samples were prepared by mixing protein sample with TruPAGE™ 4x sample buffer in a 3:1 ratio to a final volume of 10 μL per sample. 1 μL of β-mercaptoethanol was added to each sample to reduce any remaining disulfide bonds, and samples were heated for 10 min in a 70 °C water bath to completely denature. A 12-well, 12% TEA-Tricine TruPAGE™ gel was loaded into a Bio-Rad gel electrophoreses cell. The inner and out chambers of the electrophoreses cell were filled with 1x TruPAGE™ running buffer. SDS samples were then loaded into each well. A Fisher EZ-Run protein ladder was added to its own well to serve as a standard.

The electrophoresis cell was then connected to power supply and run at constant voltage (180 V) until the dye band reached the bottom of the gel (approximately 45 min). Upon completion, the loaded gel was removed from the electrophoresis cell and freed from its storage cassette. The gel was then submerged in ultrapure deionized water and subjected to 75 rpm on a New Brunswick incubator shaker for 30 min to wash away residual SDS. Wash water wash changed every 10 min. Freshly washed gel was then submerged in Fisher PAGE-Blue gel stain for 1 hr again under gentle agitation on a New Brunswick incubator shaker to stain migrated proteins. Gel stain was drained with residual stain being rinsed away with ultrapure deionized water to clearly resolve protein bands. 40 kDa bands observed in elution fraction samples were assumed to be *DszB*. Conversion of HBPS to HBP observed in activity studies outlined in Section 3.4 was taken as additional conformation of DszB presence.

3.2.8. Protein Dialysis

Dialysis was used to perform buffer exchanges on IMAC-purified protein for experimentation in which imidazole concentration or ionic strength of column elution buffer was insufficient. Thermo Fisher Scientific Snakeskin 16 mm dialysis tubing with 10 kDa cut off was used for dialysis. Dialysis tubing was presoaked in dialysis buffer (10 mM

Na_2HPO_4) for 10 min prior to use. 10 ml of IMAC-purified *DszB* produced as described in section 3.2.5 was loaded into dialysis tubing and submerged in 2 L of dialysis buffer. Dialysis was performed over 6 hrs refrigerated at 4 °C and under constant agitation via magnetic stir plate. Dialysis buffer was exchanged at the 2 and 4 hr time marks.

3.3. Substrate Synthesis

Pure biphenosultine, a precursor of HBPS was obtained as a gift from the Texas A&M University National Products LINCHPIN Laboratory. Biphenosultine to a concentration of 20 mM was combined with 44 mM NaOH at room temperature as outlined in L. T. Harper[73] and left to react overnight on a Fisher Scientific Reliable Tube Rotator in order to synthesize the *DszB* substrate, HBPS.[73] The pH was adjusted post-reaction to 8.5, and the resulting 20 mM stock solution of HBPS was stored as 2 ml aliquots in a standard -20 °C freezer for later use.

3.4. Fluorometric Assays

3.4.1. HBP Calibration Curve

A link between the concentration of HBP of fluorescence was determined by fluorescent spectroscopy. Six HBP stock solutions in column elution buffer (0.3 M NaCl, 50 mM Na_2HPO_4, 200 mM imidazole) were created at concentrations of 20, 10, 5, 2.5, 1.25, and 0.63 μM. 200 μL of each solution was loaded in triplicate into a black, clear-bottom, 96-well tray for analysis. HBP fluorescence was measured at 414 nm wavelength using a BioTek Synergy plate reader with an excitation wavelength of 288 nm. The resulting calibration curve is displayed in Figure 0.1 and was used to evaluate reaction progress in DszB kinetic studies.

3.4.2. Kinetic Activity

Kinetic activity of *DszB* was determined using fluorescent spectroscopy. HBPS solutions were created in imidazole-free column elution buffer (0.3 M NaCl, 50 mM Na_2HPO_4) from frozen stocks as discussed in section 3.3 at concentrations of 25, 12.5, 6.25, 3.13, 1.56, and 0.78 µM. 200 µL of each solution along with a blank containing only column elution buffer was added in triplicate to a black, clear-bottom, 96-well tray. 50 µL of IMAC purified *DszB*, as discussed in section 3.2, suspended in column elution buffer at a concentration of 0.2 mg/ml was added to each well using a multichannel pipette for near simultaneous addition. The plate was immediately loaded in a BioTek Synergy plate reader, and the initial fluorescence response was measured at 414 nm wavelength with an excitation wavelength of 288 nm. Fluorescence of each well was scanned every 45 sec (the highest sampling rate allowed) for 6 min.

Using the HBP calibration curve discussed in section 3.4.1, fluorescent measurements for each well at each time point were converted to corresponding concentrations of HBP. The initial reaction velocity (d[P]/dt) for each well was calculated by linear regression of changes in HBP concentration vs. time for the first 6 min of the reaction. Average initial reaction velocities were determined for each corresponding substrate concentration, and error in reaction rate was determined by the standard deviation of this average across each of the three wells tested.

Michaelis–Menten curves were generated by plotting reaction velocity vs. initial substrate concentration. Michaelis–Menten parameters, maximum rate of reaction (V_{max}) and Michaelis constant (K_m), were determined by curve fitting reaction data to the Michaelis–Menten equation (Equation 3.1) via non-linear regression, minimizing the sum of squared errors with the Generalized Reduced Gradient (GRG) algorithm. For applications requiring *DszB* turn-over number, k_{cat} was calculated by dividing V_{max} by reaction well protein concentration (1μM).

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{[S] + K_m}$$

Equation 3.1. Michaelis–Menten equation

[P] = product concentration (μM), t = time (sec), [S] = substrate concentration (μM), V_{max} = Maximum theoretical rate of substrate saturation (μM per sec), K_m = Michaelis–Menten constant (μM)

3.4.3. Product Inhibition

The effect of HBP inhibition on *DszB* activity was assessed using fluorescent spectroscopy. The kinetic assay described in section 3.4.2 was used to establish a baseline, uninhibited activity for *DszB*. The assay was then repeated an additional five times, with HBPS substrate solutions being spiked with a different initial amount of HBP in each repetition. The full assay was repeated with working reaction well concentrations of HBP at 0.63, 1.25, 2.5, 5, and 10 µM.

The apparent Michaelis–Menten parameters for each different concentration of HBP were determined according to the method outlined in section 3.4.2. The effect of HBP concentration on each parameter was then compared to assess which model best described inhibition (Equation 3.2).

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{[S] + K_m(1 + \frac{[I]}{K_I})}$$

Equation 3.2. Modified Michaelis–Menten equation for competetive inhibition

[P] = product concentration (μM), t = time (sec), [S] = substrate concentration (μM), V_{max}= Maximum theoretical rate a substrate saturation (μM per sec), K_m= Michaelis–

Menten constant (μM), [I] = inhibitor concentration (μM), K_I = inhibition dissociation constant (μM).

3.4.4. Catalytic Stability

Catalytic stability of *DszB* was measured using fluorescent spectroscopy. The kinetic assay described in section 3.4.2 was used immediately following IMAC purification to establish a baseline *DszB* activity. The stability assay was then performed again on the same batch of *DszB* at 2, 6, 12, 24, and 48 hrs. While not in use, *DszB* batch solution was stored at 4 °C.

When investigating the impact of buffer additives on kinetic stability, *DszB* was still eluted from the column in the standard column elution buffer. Buffer modifications were made to the whole stock and occurred immediately after column elution prior to initial kinetic assay. To account for any immediate effect additives may have on activity, either by simple dilution or minor molecular interactions, kinetic stability was evaluated based on turn-over ratios. The turn-over number for each solution at time "t" was calculated ($k_{cat,t}$) and compared to its corresponding initial turn-over number observed immediately after purification ($k_{cat,0}$). The ratio between $k_{cat,t}$ and $k_{cat,0}$ was used as a measure of activity decline.

3.5. Differential Scanning Fluorimetry

Differential Scanning Fluorimetry (DSF) was used to evaluate the thermal stability of *DszB*. In this analytical technique, extrinsically fluorescent dye dependent on hydrophobic interactions to magnify emissions is combined with protein in an aqueous environment. Fluorescent response is monitored over an increase in temperature, with the protein melting point being revealed by a spike in fluorescence as the denatured protein's hydrophobic core becomes more dye accessible. Figure 3.3 provides a general example of DSF used as a protein stability measurement.

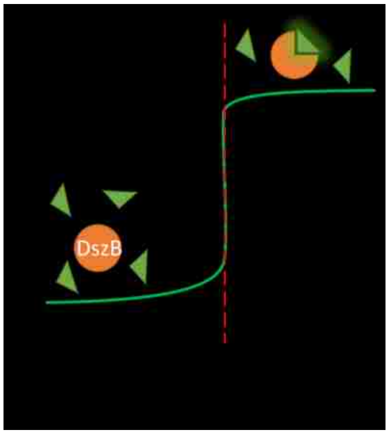

Figure 3.3. Differential Scanning Fluorimetry Example

Green triangles represent extrinsically fluorescent dye, full orange circle represents native protein, and the partial orange circle indicated denatured protein. Theoretical melting point (T_m) is indicated by the red dashed line.

IMAC-purified *DszB* from section 3.2 was combined with 5,000 x SYPRO™ Orange dye and diluted with phosphate buffer (0.3 M NaCl, 50 mM Na_2HPO_4) for a final concentration of approximately 0.15 mg/ml *DszB,* and 5 x SYPRO™ Orange dye. 10 μL of the *DszB*-Dye solution was added to each well in a 96-well optical-reading-compatible qPCR microplate and combined with 10 μL of desired buffer additives. The microplate was sealed and subject to a qPCR Thermofluor method monitoring fluorescent emission at 600 nm over a temperature range 20 °C to 90 °C. Temperature was changed in 1 °C increments with a 1 min equilibration time following each temperature change. Protein melting point was determined by locating inflection points between temperature and fluorescence. Inflection points were identified by locating relative extrema in the second derivative.

3.6. Molecular Dynamic Simulations

Molecular Dynamics (MD) simulations are a computational tool used to predict the physical motion of atoms over short time window. When applied to biological systems such as proteins and polysaccharides, MD simulations provide a powerful, complementary analysis tool for answering research questions regarding molecular-level phenomena. Beginning with initial atomic coordinates, often either provided by a crystal structure or constructed by homology and docking calculations, MD simulations follow a basic algorithm by dividing a desired simulation time window into small, iterated steps (usually 1-2 femtoseconds in length). In each time step, forces acting on atoms are predicted via Newtonian physics and, then, moved accordingly within in the limitations of boundary conditions set for the simulation. Metrics of interest are then calculated for the individual time step and the simulation progress to integrate the same process for the next time window.

3.6.1. Newtonian physics

Atom trajectory is determined by Newton's second law of motion, written in terms of the potential energy function in Equation 3.3. For a given atom "I," the force exerted on it is a function of its mass [m], and its acceleration (second derivative of position [r] with respects to time [t]), which in turn is a function of potential [U].

$$f_i(r_i) = m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = -\nabla_{r_i} U(r_1, r_2, r_3, \dots, r_N)$$

Equation 3.3. Newton's Second Law of Motion

Framed in this context, the trajectory of each atom in the system is a function of the position of every other atom in the system, requiring simultaneous solution of a large system of complex differential equations to calculate. The intricacy of this system far exceeds the possibility of analytical solution, requiring the use of computational integration methods to estimate trajectories. Within MD simulations in this dissertation the velocity Verlet algorithm was implemented in NAMD to integrate molecular systems over time. [74-75]

3.6.2. Potential Energy

The potential energy of the system is given by the summed total its bonded and non-bonded contributions. Equation 3.4 describes the former.

$$U_{bonded} = \sum_{bond} k_b(b - b_0)^2 + \sum_{angle} k_\theta(\theta - \theta_0)^2 + \sum_{dihedral} k_\varphi(1 + cos(n\varphi - \delta))^2 + \sum_{improper} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_{UB}(S - S_0)^2 + U_{cmap}(\varphi, \omega)$$

Equation 3.4. Bonded potential energy contributions

In each of the elements of bonded potential energy, “k” represents the force constant for the corresponding component. All subscripted “0” terms indicated equilibrium values for their corresponding variables (i.e. b = bond length, b_0= equilibrium bond length). The bonding contribution is calculated from bond stretching via bond length [b]. Angle contributions are determined by bending in bond angles [θ]. The dihedral element is calculated based on bond angle rotation [φ], period [n], and phase shift [δ]. Improper contributions come from improper angle bending [ω]. Finally, the Urey-Bradley element captures contributions from vibration [S] and backbone torsion [U_{cmap}].

Nonbonded interactions are described by the summed total of van der Waals and electrostatic contributions as shown in Equation 3.5.

$$U_{nonbonded} = \sum_{van\ der\ Waals} \varepsilon_{ij}^{min}\left[\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{6}\right] + \sum_{Electrostatic} k_c \frac{q_i q_j}{\varepsilon r_{ij}}$$

Equation 3.5. Non-bonded potential energy contributions

Van der Waals interactions are modeled according to a 12-6 potential describing a potential well of depth [ε_{ij}^{min}], reaching a minimum value at distance [R_{ij}^{min}]. The interatomic distance between atoms “i” and “j” is given by [r_{ij}]. Electrostatic interactions are modeled according to Coulomb’s law via Coulomb’s constant [k_c], the dielectric constant [ε], the charges of atoms “i” and “j” [q], and, again, the interatomic distance [r_{ij}].

3.6.3. Free energy summary

After solving potential energy equations, free energy calculations can be performed based on specific system designations. While a system is constrained to maintain a constant total number of atoms (via periodic boundary conditions), Helmholtz free energy can be calculated in a pressure-regulated isochoric and isothermal system, an ensemble referred to as the NVT or the Canonical ensemble. Alternatively, in an NPT ensemble, isobarically and isothermally-constrained, Gibb’s free energy can be calculated.

3.6.4. Simulation specificity

Molecular dynamic simulations were executed using CHARMM[76] and NAMD.[77] Visualization of MD simulation data was performed using Visual Molecular Dynamics (VMD) [78] and PyMOL software. Details regarding specific molecules simulated and the relevant parameters of such simulations conducted can be found in subsequent chapter-specific methods sections.

Chapter 4 – Inhibition mechanisms of Rhodococcus erythropolis 2'-hydroxybiphenyl-2-sulfinate desulfinase (*DszB*)

This chapter has been adapted from a publication currently in review by the *Journal of Physical Chemistry B*. All experimental work in this chapter was performed by the author of this dissertation. All MD simulations were performed by Dr. Yue Yu, co-first author. The author of this dissertation authored the discussion and conclusions of both the publication and this chapter.

4.1. Introduction

Liquid petroleum products, derived from sulfur-containing crude oil, comprise a majority of the fuel consumed in the transportation sector.[2, 11] Sulfur compounds naturally present in crude feedstocks pass through hydrodesulfurization processes to the refined fuel due to insufficient conversion of refractory thiophenic molecules. Upon combustion, the sulfur compounds form sulfur oxides (SOx). SOx 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).[3] Accordingly, government standards regulate the amount of sulfur allowed in finished fuels, and the standards are expected to become increasingly stringent.

Biodesulfurization via enzyme catalysis offers potential as a deep desulfurization process, where highly specific thiophenic desulfurization occurs at ambient temperature and pressure.[18,79] Sulfur fixation is both a common and essential ability shared by most bacteria, though the mechanism by which this occurs can vary significantly.[34] The mechanism by which microbial organisms access sulfur bonds in aromatic molecules appears to fall into one of two categories: ring-destructive or sulfur-specific.[9, 30, 80] In the former, enzymes catabolize the sulfur compounds in such a way that the aromatic ring is destroyed, easing access to the sulfur. In the latter, enzymes open the sulfur-containing rings breaking only bonds shared with the target sulfur atom.[39] As the “ring-destructive” approach cleaves the carbon-carbon bonds from which most of a fuel’s heating value is derived, industrial biodesulfurization processes have the most to gain from employing ring-opening enzymatic mechanisms that break as few carbon-carbon bonds as possible to achieve sulfur removal. Mild operating conditions, coupled with the ability to leave the fuel heating value virtually undisturbed, makes biodesulfurization an advantageous additive process, post hydrodesulfurization, from both an initial capital and operational cost standpoint.[20]

The 4S pathway, named for its 4-step sulfur removal mechanism, is shared by a variety of microorganisms and is capable of fixing sulfur from dibenzothiophene (DBT) and DBT derivatives via the “sulfur-specific” method.[18] DBTs account for 60% of sulfur emissions from liquid transportation fuels, making the 4S pathway an ideal biodesulfurization process.[18, 80-81] The 4S pathway employs four distinct enzymes to break DBT down to sulfite and 2-hydroxybiphenyl (HBP), beginning with oxidation of the ring-bound sulfur by the monooxygenase, *DszC* (this step requires indirect use of NADH cofactor). The second step involves further oxidation of the sulfur, again by enzyme *DszC*. Oxygen for both steps is supplied by free water molecules. Oxidation is followed by a ring-opening step catalyzed by a second monooxygenase, *DszA*, and two NADH cofactors, in which the first carbon-sulfur bond is broken. Both a hydroxyl group and sulfate group are left on opposite rings of phenylbenzene, forming 2’-hydroxybiphenyl-2-sulfinate (HBPS). Each of these first three steps are aided by an oxidoreductase, *DszD*. In the final step, catalyzed by desulfinase *DszB*, the sulfate group is cleaved from HBPS to form the pathway product, HBP.[17,39]

The 4S pathway is currently limited in its industrial utility given its relatively slow turnover rates. *DszB* is the rate-limiting enzyme in the biocatalytic process, as it exhibits both a low turnover and experiences product inhibition. [1, 35] In fact, the inhibitory effect of the presence of the HBP product on the activity of the 4S pathway as a whole has been documented. [36, 39] Recent studies report HBP inhibition of each of the pathway enzymes to varying degrees, but the type of inhibition and accompanying mechanistic relevance to the regulation of the 4S pathway remains largely unexplored, particularly with respect to the crucial rate-limiting step catalyzed by *DszB*.[40] Kinetic studies performed on purified *DszB* in solution indicate the possibility of a competitive form of product inhibition, with several HBP analogs of varying functional group attachments (Figure 4.1) also inhibiting *DszB* activity over a wide range of inhibition constants (K_I).[24] Molecules examined include HBP, 2’2-biphenol (BIPH), 1,8-naphthosultam (NTAM), 2-biphenyl carboxylic acid (BCA), and 1,8-naphthosultone (NAPO), all of which are found in crude oil.[82] Within the two-functional-group class, the most inhibitory of the tested analogs was BIPH. In the planar naphthenic class, the most strongly inhibiting compound was NTAM. While BCA and NAPO were non-inhibitory, they were also not reactive.[50]

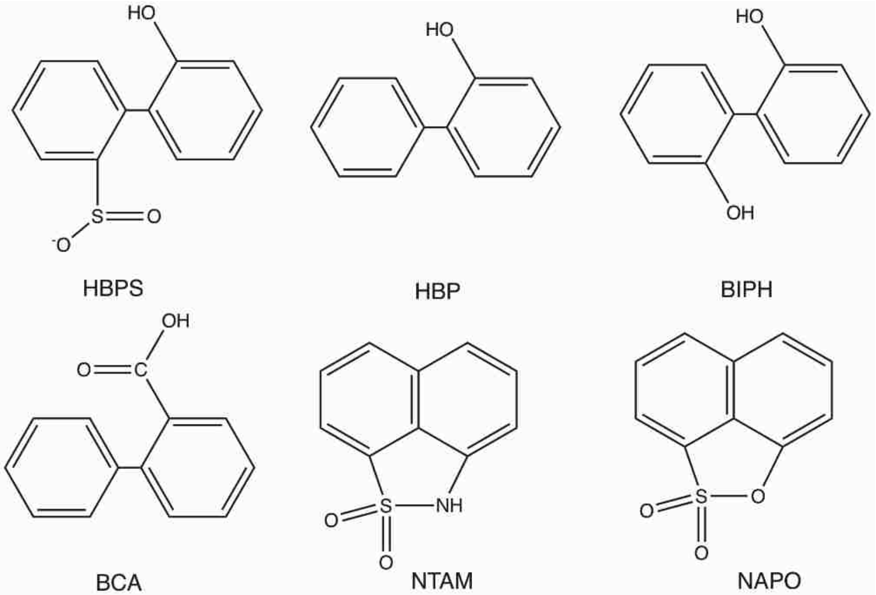

Figure 4.1. Aromatic HBPS analogs with varying inhibitory effects on DszB.

HBP is the reaction product and has a K_I of 0.5 µM under various experimental conditions; BCA and NAPO are non-inhibitory, yet not productive; and BIPH and NTAM inhibit DszB with K_I values of 17 µM and 1.8 µM, respectively.[50]

Thus far, structural characterization of isolated *DszB* beyond basic crystallographic studies is sparse, with little investigation into *DszB* structural changes upon binding to molecules other than its substrate, HBPS. [43] The two existing crystallographic studies from Lee *et al.* demonstrate a large-scale, substrate binding-induced conformational change, burying the otherwise surface-exposed H60 deep in the active site to participate in catalysis. Differences in the respective conformations of (1) the unbound *DszB* structure (PDB 2DE2), (2) the C27S catalytically inactive variant bound structure (*DszB*-HBPS, PDB 2DE3), (3) a C27S mutant structure bound to a known alternative substrate, 2-biphenyl sulfinic acid, (PDB 2DE4) and (4) an overlay of bound and unbound structures for conformation change comparison are shown in Figure 4.2.[43, 45] Notably, the 2DE3 and 2DE4 protein backbones adopt virtually identical conformations despite binding different substrates, suggesting the conformational change is not necessarily substrate-specific and may even apply to a wider group of ligands. Between the bound and unbound conformations (2DE2 and 2DE3), three distinct residue loops exhibit significant differences (Figure 2, loops shown in turquoise). Loop 1 (residues 50-60) changes from an extended structure to an α-helix upon substrate binding, introducing H60 to the active site. Loop 2 (residues 135-150) maintains helical structure but moves upward, and loop 3 (residues 180-200), similar to loop 1, changes from an extended structure to an α-helix. While one might expect comparable conformational responses to structurally similar inhibitors, structural data in support of this hypothesis does not yet exist. Molecular-level analysis of *DszB* inhibitor binding is, to the best of our knowledge, also unavailable, although the need for such characterization of inhibition mechanisms is essential to developing a complete understanding of enzyme function.

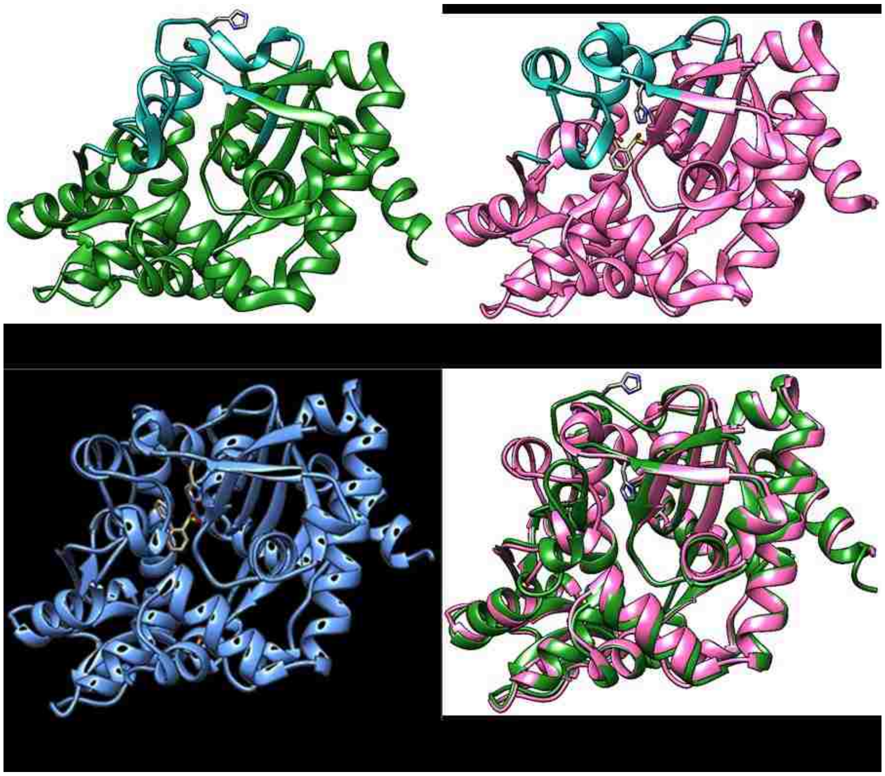

Figure 4.2. Binding-induced conformational changes in DszB.

PDB Entry 2DE2: Unbound DszB [green cartoon]; PDB Entry 2DE3: C27S DszB variant bound to substrate (HBPS) [pink cartoon]; PDB Entry 2DE4: C27S DszB variant bound to an alternate substrate (2-biphenyl sulfinic acid) [blue cartoon]. High mobility loop regions 1, 2, and 3 are shown in the top two panels in turquoise to illustrate regions of extreme conformation change. An overlay of unbound and bound structures at bottom left, with H60 shown in stick representation, further illustrates the conformational change upon substrate binding. In panels 2DE3 and 2DE4, the ligands are shown in stick representation.

To pave the way for future improvements to *DszB* function via structure-guided design, we seek to deepen molecular-level understanding of *DszB* inhibition mechanisms. Toward this goal, we examined the global changes in *DszB* conformation in response to a collection of substrate analogs and potential inhibitors (Figure 4.1). Using circular dichroism (CD) spectra and kinetic assays, we examined structural response and inhibition of *DszB* in the presence of HBP. We then developed a detailed molecular-level analysis of *DszB* small molecule binding using computational approaches. Using free energy perturbation with

Hamiltonian replica exchange molecular dynamics (FEP/λ-REMD), we compared binding free energy values of HBPS analogs with experimentally reported inhibition constants. Such comparison provided information as to how the thermodynamic signatures relate to enzyme function. Molecular dynamics (MD) simulations of *DszB* bound with the substrate and inhibitors provided insights into how the various functional groups of the selected ligands contribute to binding.

4.2. Experimental Methods

4.2.1. Materials

A HIS-Select® High Flow Nickel Affinity Gel (Sigma-Aldrich H0537) immobilized metal affinity chromatography (IMAC) column was used for protein purification. Polyacrylamide gel electrophoresis (PAGE) gels were purchased, precast, along with accompanying running buffer and loading dye from Sigma Aldrich. EZ-Run™ Protein Ladder (Fisher BP3602500) was used as a standard. HBPS was synthesized from biphenosultine as previously described;[83] the biphenosultine was a gift from the Texas A&M University National Products LINCHPIN Laboratory. All other chemicals were purchased from VWR (Radnor, PA).

4.2.2. Cell Culture and Protein Expression

Genomic DNA for *DszB* from *Rhodococcus erythropolis* IGTS8 was used as a template for restriction cloning and inserted into plasmid pTAC-MAT-Tag-2 (Sigma E5405, C-terminus MAT Tag). *Escherichia coli* BL21 was then transformed to include the resulting plasmid, in addition to the chaperone vector pG-KJE8 (Takara 3340), for expression of molecular chaperones *dnaK-dnaJ-grpE* and *groES-groEL*, as needed.

Transformed *E. coli* was cultured in Lysogeny Broth (LB) at 37 °C with ampicillin and chloramphenicol concentrations of 50 mg/ml each for strain selection. When cultures reached an optical density (OD600) of 0.8, they were chilled to ambient temperature, and protein expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a concentration of 1 mM. L-Arabinose was added at a concentration of 2 mg/ml for co-expression of *groEs-groEL* chaperone proteins.

4.2.3. DszB Purification

All purification procedures where performed at 4 °C. Cell lysis was achieved by suspending approximately 6 g of frozen cell pellet in 50 ml of chilled equilibration buffer and pulse sonicating on ice with a sonic dismembrator (Fisher Scientific Model 505). The resulting lysate was centrifuged at 64,000 x g for 30 min, after which the supernatant was filtered with a 0.22 µm syringe filter for removal of cell debris. The filtrate was immediately gravity fed through an IMAC column previously equilibrated with 5 column volumes of equilibrium buffer (50 mM sodium phosphate, 0.3 M sodium chloride, and 10 mM imidazole). The IMAC column was then washed with an additional 5 column volumes of equilibration buffer. Purified *DszB* was eluted from the IMAC column with imidazole containing elution buffer (equilibrium buffer with 100 mM imidazole) at a concentration of 0.2 mg/ml. Purity was assessed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 0.2).

4.2.4. Circular Dichroism

IMAC-purified *DszB* was dialyzed against a low salt-content, phosphate-buffered saline solution to remove remaining residual imidazole and salts. Serial dilutions of HBP were prepared in five-fold steps using waste dialysate with the addition of 1 % ethanol for solvation of the hydrophobic molecule, spanning a final concentration range of 125 nM to 80 µM HBP. HBP solutions were then combined with the dialyzed *DszB* solution in equal volumes for circular dichroism (CD) analysis. An HBP-free *DszB* control sample was prepared by mixing 1 % ethanol-containing dialysate with dialyzed *DszB* solution to ensure the solvation environment of all CD samples were identical. Samples were then analyzed using a Jasco J-810 CD spectrophotometer at 25 °C over wavelengths spanning 200 nm to 250 nm. [84,85]

4.2.5. Fluorescent Kinetic Assay

IMAC-purified *DszB* was added to well plates containing 2-fold serial dilution of HBPS in 50 mM sodium phosphate buffer (with 0.3 M sodium chloride) spanning a concentration range of 625 nM to 20 µM HBPS. *DszB* was also added to wells containing only buffer to serve as controls. Samples were incubated at 20 °C and monitored for fluorescence emission at a 414 nm wavelength using a BioTek Synergy plate reader with an excitation wavelength of 288 nm. [25] Using a calibration curve generated from the linear relationship between HBP concentration and fluorescence intensity, initial reaction velocities were calculated to generate a Michaelis-Menten saturation curve. Apparent Michaelis-Menten parameters were determined numerically using the Generalized Reduced Gradient (GRG) algorithm. This process was then repeated with varying initial concentrations of HBP, spanning a concentration range of 625 nM to 10 µM HBP; further increases in HBP concentration generated too high a background to observe initial reaction rates.

4.2.6. Molecular Dynamic Simulations

MD simulations were used to examine the dynamic behavior of the set of aromatic inhibitors and substrates (Figure 4.1) bound in the active site of *DszB*. The *DszB*-HBPS complex model was constructed based on crystallographic data from Protein Data Bank (PDB) structure 2DE3; the C27S mutant was reversed to the wild-type cysteine.[86] For the *DszB*-HBP complex, construction started from the 2DE3 structure (reverted to wild-type); the sulfinate group was deleted from the HBPS substrate, resulting in the HBP product. For the remaining molecules, the bound HBPS substrate was manually removed from the structure, and the aromatic analogs, BIPH, BCA, NAPO, and NTAM, were placed at the initial HBPS position as initial coordinates for docking.

Ligand docking calculations were performed using standard affinity-based methods in AutoDock (version 4.2) to compute and cluster the grid-based free energies of protein-ligand complexes.[87] Initial conformations of BIPH, BCA, NAPO, and NTAM were obtained from quantum mechanical geometry optimizations, as reported in our previous study on force field parameterization of HBPS analogs.[83] Ligand conformations were allowed to be flexible, including torsion angles, during docking calculations to obtain the optimum binding position inside the *DszB* binding pocket. Autogrid4 was used for setting up grid maps centered on selected ligands, with grid dimensions of 66 Å × 66 Å × 66 Å for all molecules and a grid spacing parameter of 0.375 Å. The Genetic Algorithm (GA) was applied to search for docking positions of each molecule with a ranked cluster analysis on 50 independent GA runs. Parameters for the calculations were a rate of gene mutation of 0.02, a rate of crossover 0.8, a population size 150, and the maximum number of evaluations capped at 25 M.

The *DszB* bound systems were solvated in a 90 Å × 90 Å × 90 Å pre-equilibrated cubic box filled with TIP3P water molecules. [88-89] Prior to solvation, the systems were minimized in vacuum for 500 steps using the steepest descent (SD) method followed by 1000 steps using the adopted basis Newton-Raphson (ABNR) method. The CHARMM36 force field with CMAP correction was applied to model the protein.[90-92] Optimized CHARMM general force field (CGenFF)-based parameters were used to model the HBPS analogs.[93,94] Protonation states of all the titratable residues were calculated using the H++ web server at the optimal pH of 7.4.[95, 50] Additionally, 0.3 M NaCl and 50 mM HPO_4^{2-} were included in the solvent to mimic experimental media, and extra Na^+ ions were added to neutralize the system charge (*DszB* is $17e^-$, and HBPS is $1e^-$). Periodic boundary conditions were applied, and Lennard-Jones interactions were truncated at 12 Å. Electrostatic interactions were calculated using the particle mesh Ewald method (PME)[96] with 1 Å grid spacing and 6th order spline. Hydrogen-heavy atom bond lengths were constrained by SHAKE.[97] The solvated systems were minimized and equilibrated prior to production MD simulations (i.e., data collection). Each system was minimized in CHARMM with 2000 steps of the SD algorithm holding the protein and ligand fixed via harmonic restraints. The restraint on the ligand was then removed, and 2000 additional SD steps were taken. A final 5000 ABNR minimization steps were taken following the removal of all restraints. The minimized systems were equilibrated by slowly heating the system from 90 K to 300 K in 50 K increments over 100 ps 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 1 atm to equilibrate system densities.[98-100] The density-equilibrated systems were simulated in NAMD for 200 ns.[77] MD production simulations were conducted in the

NVT ensemble using a 2-fs time step at 300 K. Temperature was controlled using the Langevin thermostat in NAMD.

4.2.7. Free Energy Calculations

For quantitative examination of the thermodynamic factors contributing to ligand binding in *DszB*, we performed absolute binding free energy calculations using free energy perturbation with Hamiltonian replica-exchange molecular dynamics (FEP/ λ-REMD), an enhanced sampling free energy calculation method.[101,102] The aromatic analog absolute binding free energies were calculated for direct comparison with experimental inhibition constants, K_I.

Ligand insertion into the binding pocket was split into two thermodynamic paths (Figure 4.3). The two paths were: (1) decoupling the bound ligand from *DszB* from solvated conditions into a vacuum, ΔG_1; and (2) decoupling the solvated ligand from solution into a vacuum, ΔG_2. The difference between the two changes in free energy, ΔG_b°, is the absolute ligand binding free energy of enzyme-ligand complex. For each free energy calculation, the potential energy, U, can be expressed in terms of four different thermodynamic coupling parameters, λ_{repu}, λ_{disp}, and λ_{elec} to control the nonbonded repulsive, dispersive, and electrostatic interactions, respectively, of molecule with the surrounding environment (Equation 4.1). An additional parameter, λ_{rstr}, was used to control the translational and orientational restraints.

$$U(\lambda_{repu}, \lambda_{disp}, \lambda_{elec}, \lambda_{rstr}) = U_0 + U_{repu}\lambda_{repu} + U_{disp}\lambda_{disp} + U_{elec}\lambda_{elec} + U_{rstr}\lambda_{rstr}$$

Equation 4.1. Potential energy equation governing free energy calculations

U_0 is the potential energy of the system with a non-interactive ligand; U_{repu} and U_{disp} are the shifted Weeks-Chandler Anderson repulsive and dispersive components of the Lennard-Jones potential; U_{elec} is the electrostatic contribution; and U_{rstr} is the additional restraining potential energy. Orientational and translational restrains were not used in the determination of the ligand solvation free energy.

$$(\text{DszB-Ligand})_{solv} \xrightarrow{\Delta G_1} (\text{DszB})_{solv} + (\text{Ligand})_{vac}$$

$$(\text{Ligand})_{solv} \xrightarrow{\Delta G_2} (\text{Ligand})_{vac}$$

$$(\text{DszB})_{solv} + (\text{Ligand})_{solv} \xrightarrow{\Delta G_b^\circ = \Delta G_1 - \Delta G_2} (\text{DszB-Ligand})_{solv}$$

Figure 4.3. Thermodynamic cycle to determine the free energy of binding small molecules to DszB using FEP/ λ-REMD.

"Ligand" represents the bound small molecule in a given set of simulations. The subscript "Solv" represents solvated condition, and "Vac" refers to vacuum conditions.

Free energy calculations were performed for the neutral analogs only (HBP, BIPH, NTAM, and NAPO), starting from the final timestep of the 200-ns MD simulations. The negatively charged HBPS was not considered in the free energy calculations due to finite-size effects, which gives the most significant discrepancy when protein and ligand are both charged.[103-107, 51] The free energy of binding BCA was also not determined, as the ligand did not remain bound in the active site over the course of MD simulations. The free energies were determined from 3-ns simulations conducted using 30 consecutive, 0.1-ns windows for both thermodynamic paths. The simulations employed a set of 128 replicas (72 repulsive, 24 dispersive, and 32 electrostatic) with an exchange frequency of every 0.1 ps. In determining the first thermodynamic step, ΔG_1, a positional restraint on the distance between the center of mass of the ligand and the protein was applied to aid in convergence; the bias imposed as a result of this restraint was determined numerically and accounted for in calculation of ΔG_b°.[102] Calculation of ΔG_1 for BIPH and NAPO required application of a second harmonic rotational restraint given the observed flexibility in the binding pocket. This additional restraint was applied to the root mean square deviation (RMSD) values of the aromatic carbons for each molecule to improve convergence of the calculations. Again, the bias of the additional restraint was determined by performing 5-ns umbrella sampling MD simulations, then, slowly removing the harmonic restraint on the RMSD of the aromatic carbons, and using the Multistate Bennett Acceptance Ratio (MBAR) method to determine the contributed potential.[108]

Individual repulsive, dispersive, and electrostatic energy contributions and statistical uncertainties were determined based on the output energies from the last 1 ns (the converged region) of the replica exchange simulations using MBAR for both thermodynamic paths.[108] The energies were summed for each path, including the additional restraint potentials, to obtain the total free energy change of each path. Convergence was determined by monitoring the free energy calculations of the two thermodynamic paths over time (Figure 0.3). Experimental K_I values were converted to $\Delta G_b°$ based on Equation 4.2 where $\Delta G_b°$ is the binding free energy, R is the ideal gas constant, K_d is the dissociation constant (equivalent to K_I), and c is the standard reference with unit μM.

$$\Delta G_b^\circ = RTln\left(\frac{K_d}{c}\right)$$

Equation 4.2 Relationship between binding free energy and dissociation constnat

4.3. Results

4.3.1. HBP-Induced Conformation Change

The mean residue ellipticity of *DszB* in response to the presence of HBP was measured for concentrations ranging from 64 nM to 40 μM (Figure 0.4), revealing an increase in signal intensity with increasing ligand concentration. While any change in the mean residue ellipticity would indicate a shift in protein conformation, the preservation of the spectra's overall shape with an increase in signal intensity over the 200-250 nm wavenumber is characteristic of a protein adopting a more α-helical secondary structure. Mean residue ellipticity over wavenumber for each concentration of HBP was numerically integrated (Figure 4.4), taking the increase in signal intensity to be directly proportional to an increase in helical content. The resulting "binding curve" indicates increasing amounts of HBP elicited increased conformation change up to approximately 2 μM, where α-helicity of *DszB* saturated, and the conformation change became static.

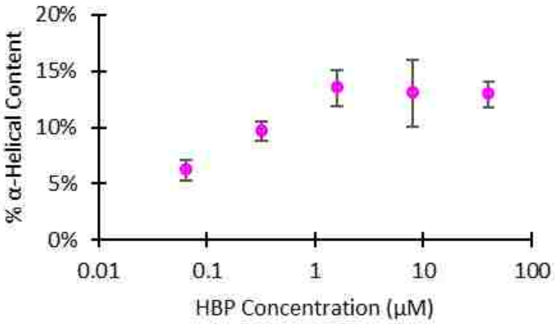

Figure 4.4. Percent α-helical content of DszB as a function of HBP concentration.

Helical content increased with increasing concentration of HBP, as determined from comparison of integrated CD spectra (individual CD spectra available in supplemental material, Figure 0.4). Error bars depict standard error with N=3.

4.3.2. Product Inhibition

The effects of conformation change on *DszB* activity were evaluated by measuring impact on initial velocity, revealing the product, HBP, competitively inhibited *DszB*. Michaelis-Menten saturation curves for purified *DszB* and *DszB* with varied HBP concentrations were obtained from fluorescent kinetic assays (Figure 4.5). The reaction rate of *DszB* was significantly reduced by the presence of HBP at substrate concentrations below 8 μM. At high substrate concentrations, *DszB* reaction rates were relatively unaffected by HBP. Apparent Michaelis-Menten parameters for each concentration of HBP were determined numerically using a Generalized Reduced Gradient (GRG) algorithm. Each parameter was then considered as a function of HBP concentration. The theoretical maximum rate of reaction (V_{max}) appeared to be independent of HBP concentration, while the Michaelis constant (K_m) increased with increasing concentrations of HBP (Figure 0.5). This is consistent with a classic competitive inhibition model.

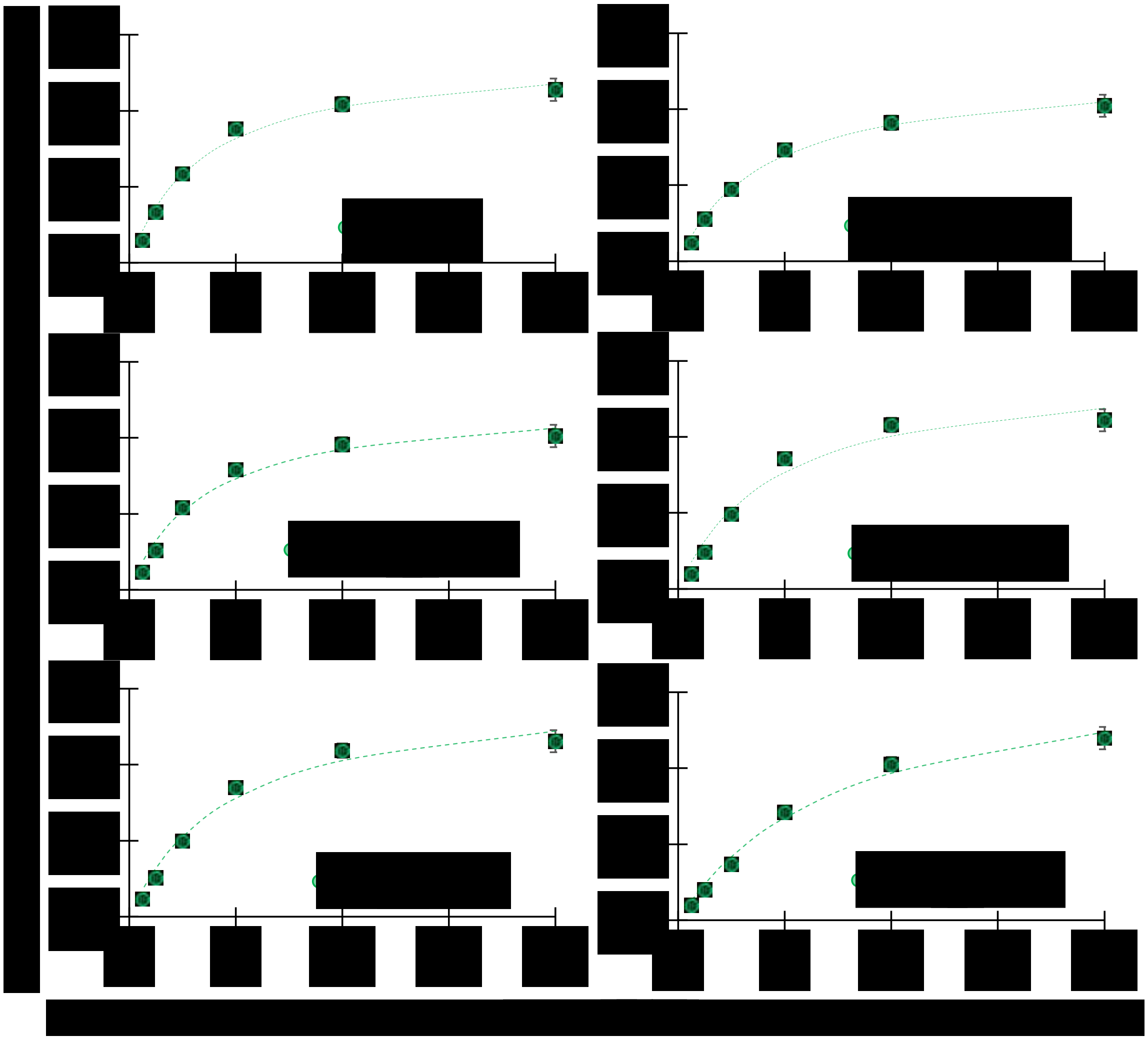

Figure 4.5. Michaelis-Menten saturation curves for purified DszB (w/o HBP) and DszB with varied concentrations of HBP (w/ 1.25, 2.5, 5, 10, and 625 µM).

Dashed lines indicate curve fitting results to the standard Michaelis-Menten equation. Error bars denote standard error with N=3.

4.3.3. Aromatic Molecule Binding Poses

DszB is thought to implement an electrophilic aromatic substitution mechanism to cleave the carbon-sulfur bond of the HBPS substrate, involving several active site residues in the process.[10] Specifically, the sulfinate group of HBPS is directly substituted by the C27 proton without formation of an arenium ion (σ-complex); H60, R70, and G73 in the active site of *DszB* play important roles in stabilizing transition states.[39, 46] As a consequence, these catalytic residues were in close proximity to HBPS in the 2DE3 structure (Figure 4.6A). The C27 sulfur (denoted at C27-S) is 2.8 Å away from the HBPS aromatic carbon directly participating in the electrophilic aromatic substitution to cleave the sulfinate group. The protonated N atom of the H60 imidazole ring (H60-N_τ) measures 3.0 Å from the sulfinate O; $N_{\eta 1}$ of R70 is 2.7 and 3.5 Å away from the O atoms of the HBPS sulfinate group; and the main chain N atom of G73 (denoted as G73-N) is 3.5 Å from the sulfinate O (Figure 4.6A). The HBP binding position was adopted from the HBPS position (without docking calculations) by manually cleaving the sulfinate group (Figure 4.6B).

AutoDock was used to model the binding positions of the remaining aromatic analogs for which there were no crystallographic structures to guide model construction, including BIPH, BCA, NTAM, and NAPO. The AutoDock calculations considered only binding within the substrate-binding pocket of *DszB* (i.e., no secondary binding sites), as determined from the HBPS-bound structure PDB 2DE3.[43] The ranked clustering analysis from AutoDock revealed preferred binding positions (Figure 4.6C-F). As expected, all binding poses for the aromatic analogs resembled the HBPS binding pose captured in PDB 2DE3.

Docking calculations placed the aromatic analogs, BIPH, BCA, NTAM, and NAPO, in similar proximity to C27, H60, R70, and G73 as the HBPS substrate (Figure 4.6C-F). In the case of NTAM, the most energetically favorable binding position put the C27-S 3.4 Å away from the N of the sultam; H60-N_τ and R70-$N_{\eta 1}$ were 3.1 and 2.6 Å away from the O atom of the sulfone, respectively (Figure 4.6C). BIPH is a comparatively less inhibitory molecule than NTAM; however, the calculated binding position of BIPH reflected a similar proximity to C27 and H60, with distances between C27-S and H60-N_τ and the hydroxyl O atom of BIPH equal to 3.0 and 2.9 Å, respectively (Figure 4.6D). NAPO is a non-inhibitory molecule, and docking calculations suggested a preferred binding position somewhat farther from catalytic residues. C27-S was 3.6 Å from N of the sultone, and H60-N_τ and R70-$N_{\eta 1}$ were 3.5 and 3.0 Å away from the O atoms of the NAPO sulfone, respectively (Figure 4.6E). BCA is another non-inhibitory molecule, and similar to NAPO, it preferentially bound near catalytic residues, although farther than the HBPS substrate. In the case of BCA, C27-S and H60-N_τ were 3.5 and 2.8 Å away from the hydroxyl O atom of the BCA carboxyl, respectively; R70-$N_{\eta 1}$ was close to both the hydroxyl and carbonyl O atoms, with distances equal to 3.2 and 3.9 Å, respectively (Figure 4.6F).

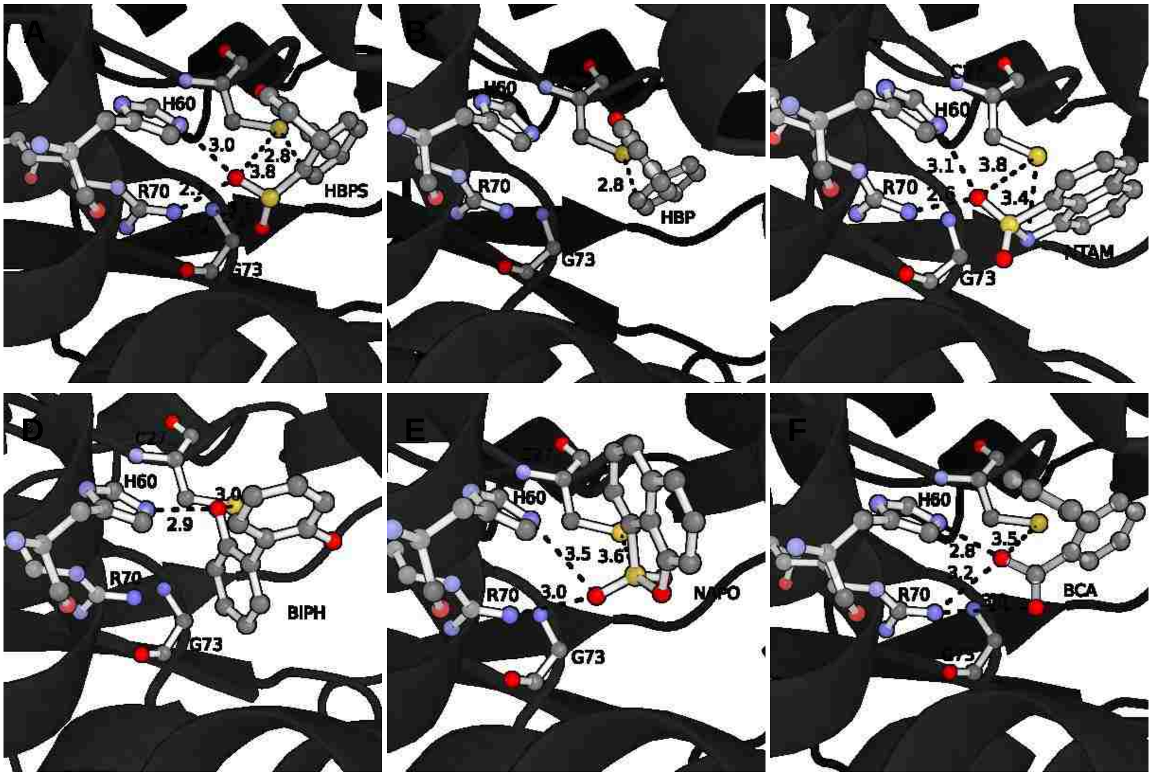

Figure 4.6. Binding poses of aromatic analogs inside the DszB binding pocket.

(A) Substrate, HBPS, bound to DszB (PDB code: 2DE3); (B) HBP binding position inside DszB, generated by manually removing the HBPS sulfinate group, as would occur in catalysis; (C-F) NTAM, BIPH, BCA, and NAPO binding positions, respectively, as predicted from docking calculations. Catalytic residues, C27, H60, R70, and G73, and bound ligands are illustrated in ball-sticks representation. All C atoms are colored in gray; O atoms in red; N atoms in blue; and S atoms in yellow. Protein structures are shown in the cartoon representation in gray. Distances between ligands and catalytic residues that were below 4 Å are labeled as dashed lines. The figure was generated using PyMOL.[109]

4.3.4. Molecular-level Binding Interactions

Molecular dynamics simulations of each *DszB*-ligand complex were performed, snapshots from which are shown in Figure 4.7. Structure coloration in Figure 4.7, from blue to red, represents the MD snapshot from initial to final position. The substrate, HBPS, remained stable within the binding site over the entire 200-ns simulation. The inhibitory product, HBP, on the other hand, moved away from the substrate binding site starting at 40 ns toward the region that undergoes a distinct conformational change upon substrate binding (superimposition of apo 2DE2 and holo 2DE3, Figure 4.2). This “gap region” is defined by three loops: loop 1 (residues 50-60), loop 2 (residues 135- 150), and loop 3 (residues 180-200). The other inhibitory molecules, NTAM and BIPH, remained in the binding site but re-positioned to interact more closely with the catalytic residues C27, H60, R70, and G73. Non-inhibitory NAPO was highly mobile in the interior of the binding pocket, and BCA completely exited the binding pocket, migrating out through the path labeled by the arrow in Figure 4.7. The exit path of BCA traversed through the three loops forming the gap region.

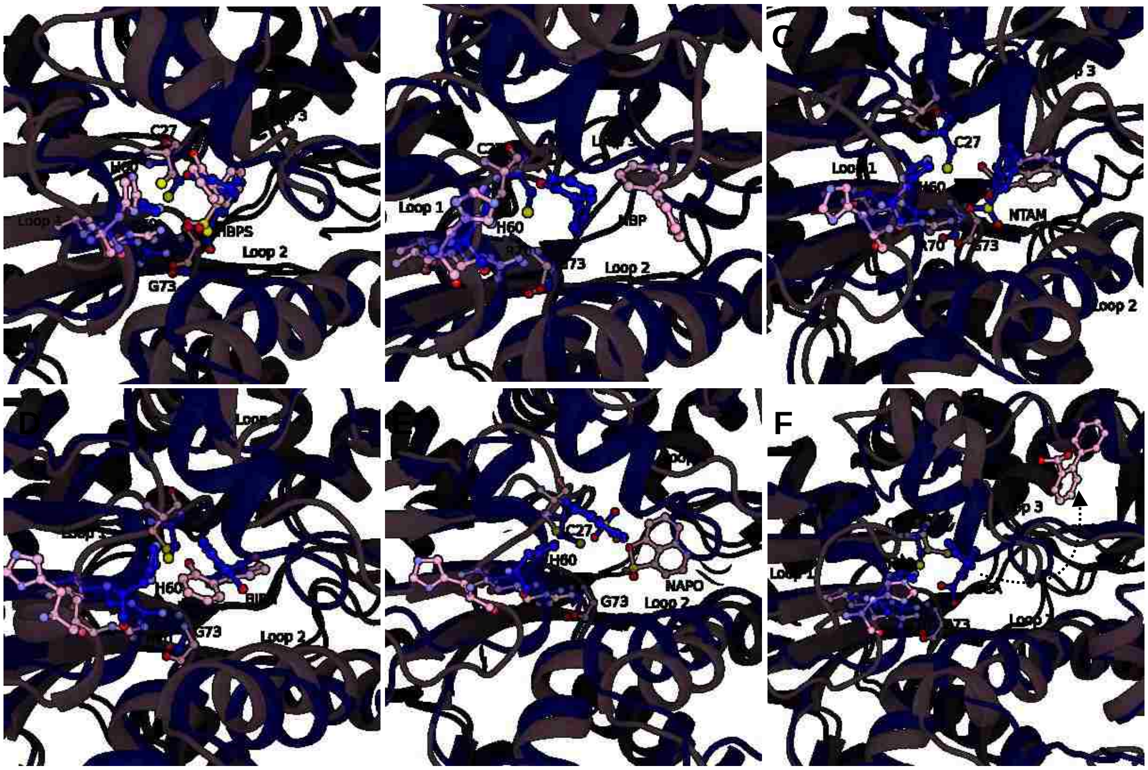

Figure 4.7. Snapshots from the MD simulations of DszB

DszB bound with (A) HBPS, (B) HBP, (C) NTAM, (D) BIPH, (E) NAPO, and (F) BCA along the course of the 200-ns simulations. The color transition from blue to red represents MD simulations from start to finish. The panel order, exclusive of HBPS, from top to bottom and left to right, corresponds to experimentally determined inhibitory effect on DszB, from most inhibitory to least inhibitory. In other words, HBP > NTAM > BIPH > NAPO > BCA in terms of inhibitory effects, with HBP as the most inhibitory and BCA as *non-inhibitory. BCA exited the binding pocket by the end of the MD simulation, with the exit path labeled in dashed line.*

From the simulation trajectories, we obtained information about the dynamic behavior of *DszB* in response to ligand binding. We examined four quantitative measures of dynamic behavior: (1) the distance between the center of mass (COM) of the protein from the COM of the bound ligand; (2) the root mean square fluctuation of the protein backbones in each simulation; (3) hydrogen bonding propensity; and (4) nonbonded interaction energies between the protein and ligands.

The distance between the COM of the protein and ligands was calculated for each time frame of the simulation and binned into histograms to illustrate relative mobility in the binding site (Figure 4.8). Non-inhibitory molecules appeared to have greater mobility inside the binding pocket than inhibitory molecules aside from the product HBP. Inhibitory molecules other than HBP remained closely engaged with catalytic residues C27, H60, R70, and G73. NTAM and BIPH, as shown in the snapshot above, were relatively stable in the binding pocket with COM distance distributions centered around 7.6 and 6.7 Å, respectively (Figure 4.8, top panel). HBP exhibited high initial stability in the binding pocket, with a COM distance of 6.6 Å; however, after 40 ns, HBP migrated from the interior of the protein to the gap region, as illustrated by the increase in COM distance to ~10.4 Å. NAPO exhibited low overall stability in the *DszB* binding pocket. BCA exited the binding pocket after 180 ns through the gap region formed between loops 1, 2, and 3 (Figure 4.8 inset). After BCA left the hydrophobic core of the protein, it migrated to surface of protein and stayed outside of loop 2, where it remained for the final 5 ns of the simulation.

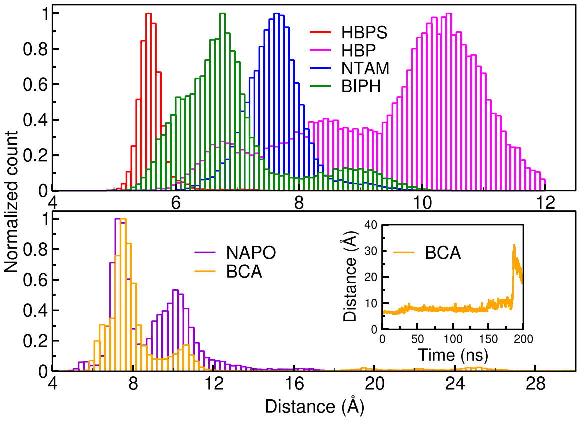

Figure 4.8. Center of mass (COM) distance analysis.

Top panel: Distance distributions between the protein and ligand COM, including substrate HBPS (red) and inhibitory molecules HBP (magenta), NTAM (blue), and BIPH (green). Bottom panel: COM distances distribution between protein and ligand COM of non-inhibitory molecules, including NAPO (purple) and BCA (orange). The inset plot in the bottom panel shows the time series distance between the BCA and protein COM.

The root mean square fluctuation (RMSF), expressed in distance, represents average movement in the protein backbone at each residue position over the duration of the simulation, with higher intensity reading corresponding to more motion in the protein which generally indicates instability in the current fold. RMSF of the protein backbones over the course of the simulations reveals that the gap region (especially loops 1 and 3) fluctuates the most, regardless of the bound ligand (Figure 4.9). RMSF on a per-residue basis is an illustration of how the backbone deviates/fluctuates about the average simulation structure. In general, the larger RMSF of the loop 1 and loop 3 corresponds to greater fluctuations, reaching RMSF values over 2 Å in all cases. Even in the apo *DszB* scenario, the loop regions exhibited high fluctuations compared to other regions of protein, supporting a potential role in active site ligand ingress/egress.

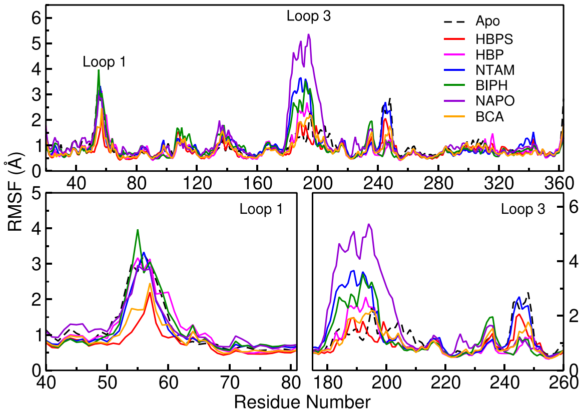

Figure 4.9. Root mean square fluctuation (RMSF) of DszB protein backbones over MD simulations.

The apo RMSF is shown as black dashed line for comparison. The bottom panels are enlargements of the loop 1 and 3 regions.

Hydrogen-bond analysis reveals that the substrate, HBPS, which is a negatively charged molecule, formed twice as many hydrogen bonds with *DszB* than any of the other molecules remaining in the binding site. For the sake of this analysis, we defined a hydrogen bond as having a donor-acceptor distance of less than 3.0 Å and a donor-hydrogen-acceptor angle of less than 20°. The primary hydrogen bonds formed between HBPS and the protein were between R70-$N_{\eta 1}$ and main chain G73-N with a sulfinate O from HBPS with 77.3% and 49.1% occupancy, respectively (Table 0.1). Note, hydrogen bond occupancy refers to the percent of the simulation in which the hydrogen bond was formed. Inhibitory molecules NTAM and BIPH intermittently formed hydrogen bonds with binding site residue G73, having 7.7% and 5.5% occupancies, respectively. HBP moved toward loops 1 and 2 after 40 ns, increasing the number of hydrogen bonds formed outside the binding site. HBP hydrogen bonded mainly with H60-N_{τ} (belongs to loop 1) with 2.5% occupancy. BCA exhibited a somewhat higher number of hydrogen bonds formed within loop 1 with H60 with 5.5% occupancy. After BCA exited the binding pocket, it hydrogen bonded with loop 2 through E189. NAPO did not appear to form any long-lived hydrogen bonds with *DszB*.

Nonbonded interactions played an important role in maintaining the stability of NTAM in the binding site. Average nonbonded interaction energies between the protein and ligands were determined from the MD simulations on a per-residue basis (Figure 4.10A). HBP did not form strong nonbonded interactions within the binding site. NTAM, on the other hand, formed strong electrostatic interactions with catalytic residue R70 and strong van der Waals (VDW) interactions with aromatic residues W155 and W255. The aromatic stacking with W155 and W255 appears to facilitate NTAM stability inside the binding site (Figure 4.10B). BIPH, in general, showed weaker nonbonded interactions compared to NTAM, but the VDW interactions between BIPH and catalytic residues C27 and G73 kept BIPH stable in the binding site.

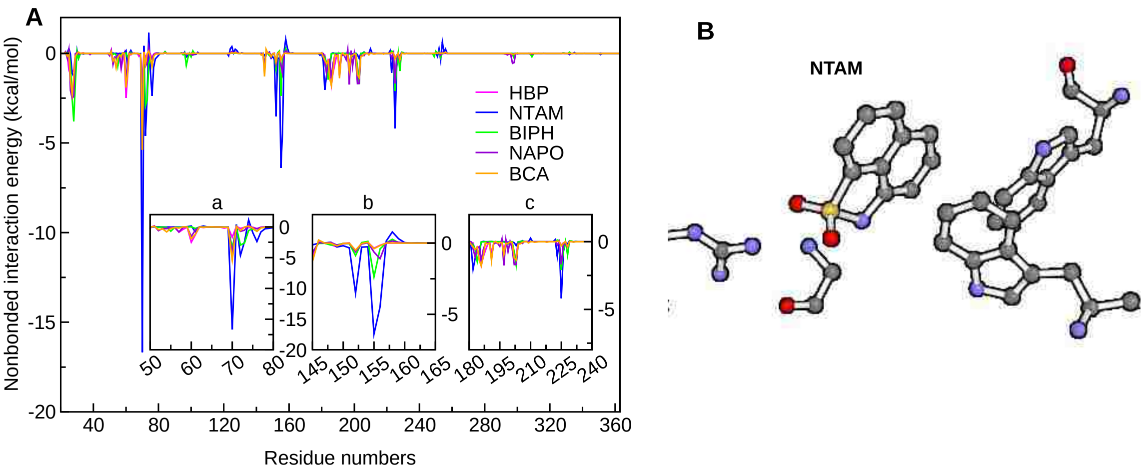

Figure 4.10. Nonbonded interactions between ligands and DszB from MD simulations.

A) Average nonbonded interactions (sum of electrostatic and van der Waals components) between aromatic analogs and DszB on a per-residue basis. Insets (a,b,c) represent the enlarged region where selected molecules have the most significant nonbonded interaction with DszB. B) Active site residue positions with bound NTAM (the initial configuration before MD production run); residues R70 and G73 are catalytic residues, and W155 and

W225 are neighboring aromatic residues that were observed as having strong VDW interactions with NTAM; color representations for the atoms are carbon in gray, nitrogen in blue, sulfur in yellow, and oxygen in red.

Calculated binding free energies aligned with experimentally-determined inhibitory effects amongst the selected molecules (Table 4.1). The calculated binding free energies for HBP, NTAM, BIPH, and NAPO were -9.4 ± 0.7, -6.9 ± 0.5, -5.8 ± 0.8, and -4.4 ± 0.4 kcal/mol, respectively. Experimental binding free energies, estimated from K_I values according to Equation 2, were -8.6, -7.9, -6.5, and N/D (not inhibitory), respectively.[50] Computationally-determined values were within ~1 kcal/mol of available experimental values and reflected the inhibitory trend of the selected molecules. The binding free energy of BCA is not reported, as it exited the binding pocket during the MD simulation. Although these values trend with experimental results, their respective inhibition potential initially appears to be independent of any shared physical properties such as hydrophobicity, indicating the chemical structure of each molecule plays at least a partial role in its effectiveness. HBP, with its high affinity for the gap region, exhibited large dispersive interactions (-6.4 kcal/mol), with weaker, favorable repulsive and electrostatic energies (-2.1 and -0.7 kcal/mol, respectively). NTAM affinity to *DszB* was dominated by favorable dispersion and electrostatic interactions (-10.3 and -2.5 kcal/mol, respectively), but binding affinity was weakened by unfavorable repulsive interactions (5.7 kcal/mol). For non-inhibitory molecule NAPO, the binding affinity was weakened primarily by unfavorable electrostatics (4.7 kcal/mol). The binding free energy as a function of time is provided in Figure 0.3 and illustrates convergence.

Table 4.1. Free energies of binding HBP, NTAM, BIPH, and NAPO to DszB.

Binding free energies calculated using FEP/λ-REMD (ΔG_b°) at 300 K are compared with experimental values (ΔG_b^{exp}), obtained from converting K_I values from Watkins et. al.[50]

Errors associated with each calculated free energy components represent one standard deviation over the last 1 ns of collected data. Error was propagated 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). All values are in kcal/mol.

	$\Delta G_b^{\circ *}$	$\Delta G_b^{exp *}$	ΔG_1	ΔG_2	ΔG_{repu}	ΔG_{disp}	ΔG_{elec}	ΔG_{rstr}
DszB-HBP	-9.4 ± 0.7	-8.6	-16.0 ± 0.6	-	20.5 ± 0.5	-27.7 ± 0.2	-8.6 ± 0.2	-0.2
HBP	-	-	-	-6.6 ± 0.2	22.6 ± 0.2	-21.3 ± 0.04	-7.9 ± 0.1	-
DszB-NTAM	-6.9 ± 0.5	-7.9	-23.2 ± 0.4	-	27.5 ± 0.3	-34.8 ± 0.2	-16.1 ± 0.2	0.2
NTAM	-	-	-	-16.3 ± 0.1	21.8 ± 0.1	-24.5 ± 0.04	-13.6 ± 0.04	-
DszB-BIPH	-5.8 ± 0.8	-6.5	-11.3 ± 0.5	-	20.7 ± 0.4	-27.9 ± 0.09	-6.7 ± 0.2	2.6
BIPH	-	-	-	-5.5 ± 0.1	23.3 ± 0.1	-22.2 ± 0.04	-6.6 ± 0.1	-
DszB-NAPO	-4.4 ± 0.4	[a]N/D	-18.5 ± 0.5	-	19.8 ± 0.4	-32.4 ± 0.2	-7.2 ± 0.2	1.3
NAPO	-	-	-	-14.1 ± 0.1	21.5 ± 0.1	-23.7 ± 0.04	-11.9 ± 0.03	-

[a] experimentally determined to be non-inhibitory

4.4. Discussion

4.4.1. HBP-Induced Conformation Change

Fluorescent kinetic assays demonstrated the subtle inhibitive effect of HBP on *DszB* activity. Moreover, inhibition studies performed over a range of HBP concentrations indicate the inhibition to be competitive in nature. This agrees with previous experimental observations and strongly supports HBP's presence in the binding pocket of *DszB* as part of the inhibition mechanism.[24, 40]

The inhibitory effect of HBP also appears to be linked, in part, to a structural response by *DszB* to the molecule. CD studies revealed increased concentrations of HBP elicited an increase in α-helical content. The structuring of the protein upon small molecule binding is consistent with a prior crystallographic study depicting a conformation change in loop 1 upon HBPS substrate binding.[43, 45] Notably, loop 1 contains residue H60, which is known to play an essential role in catalysis and resides at the protein surface in the unbound *DszB* conformation (2DE2).[43, 45] Hydrophobic molecule binding in *DszB* appears to elicit a conformational change that may be part of the catalytic mechanism, as formation of α-helical content in loop 1 repositions H60 into the active site. The relative flexibility of loop

1 was reflected in the MD simulation RMSF analysis, along with similarly high mobility in loops 2 and 3, which collectively form a gap region. The movement of the gap region loops toward the closed conformation provide weak but plentiful opportunities for hydrophobic interactions with potential hydrophobic inhibitors, in addition to positioning key catalytic residues for substrate conversion. Accordingly, we suggest that HBP, along with structurally similar inhibitors, may reduce *DszB* activity via a mechanism that traps the protein in a closed, inactive conformation similar to its substrate-bound form until the inhibitor can be displaced.

4.4.2. Active Site Interactions

The inhibitory compounds NTAM and BIPH remained highly stable inside the substrate binding site compared to HBP, as illustrated by the COM distance analysis. NTAM slightly re-oriented over the course of the MD simulation so as to more effectively interact with binding site residues R70 and G73 but remained firmly within the binding site (Figure 4.7C). NTAM stability can be attributed to hydrogen bonding with the G73-N as well as general nonbonded interactions, particularly the aromatic stacking with W155 and W225 in the protein core. BIPH stability in the binding site was facilitated primarily through VDW interactions with G73 an catalytic residue C27 in the protein core, in addition to some hydrogen bonding with G73-N. We suggest BIPH and NTAM may bind with *DszB* from solution and compete with HBPS for the binding site via a traditional competitive binding mechanism, where both can occupy the primary binding site (Figure 4.11).

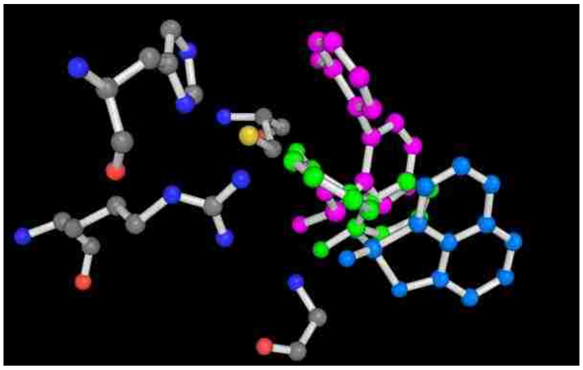

Figure 4.11. Superimposition of inhibitory molecules NTAM and BIPH with HBPS substrate.

Catalytic residues C27, H60, R70, and G73 are shown in ball and sticks representations with C atoms in gray, N atoms in blue, the S atom in yellow, and O atoms in red. HBPS, NTAM, and BIPH are shown in magenta, marine, and green, respectively.

HBP exhibited flexibility at the binding position relative to HBPS (i.e., post-catalytic cleavage position). HBP subsequently dislodged from the active site and favorably interacted with the loop 1 and 3 regions. As mentioned above, we expect the inhibitory nature of HBP may be facilitated by hydrophobic interactions upon migrating towards loop 1 and 3 region (Figure 4.7B). The high affinity of HBP for the gap region (-9.4 ± 0.7 kcal/mol) would clearly impede egress from the hydrophobic core through the 'gate' of loops 1, 2, and 3.

The non-inhibitory molecules, NAPO and BCA,[50] displayed high flexibility within the binding site, with BCA actually exiting the binding pocket after 180 ns. For BCA, a hydrogen bond with R70 existed in the first 20 ns (1.5% occupancy), and the molecule exhibited strong nonbonded interactions with R70 within the same time range shown (Figure 0.6). The interactions ceased after 20 ns when BCA migrated to the loop 1 region and formed a new hydrogen bond with H60 with a higher occupancy of 5.5%. After 170 ns, the interaction between BCA and H60 ceased, and the ligand exited the binding pocket due to its hydrophilic preference for the solvent. After BCA exited, loop 2 hydrophilic residue E189 formed a hydrogen bond with BCA trapping the molecule near the protein surface outside the loop 2 region.

There were no long-lived hydrogen bonds formed between NAPO and residues of *DszB*. Compared with NTAM, the two molecules have very similar structures (C-SO_2-O-C) to (C-SO_2-NH-C). By substituting NH to O, the O atoms surrounded by high electronegative environment (SO_2) in NAPO exhibited less electronegativity compare to the N attached to a more electropositive H in NTAM (the magnitude of the oxygen charge decreased by 0.1 e compared to N from NTAM). NAPO, therefore, formed fewer hydrogen bonds and weaker electrostatic interactions than NTAM, corresponding with the observed high flexibility of NAPO inside the *DszB* binding site. These observations are consistent with the non-inhibitory behavior of NAPO.[50]

4.4.3. Relating Binding Free Energies to Inhibition Mechanisms

The calculated binding free energies were in excellent agreement with reported values of the inhibition constants. Watkins *et. al.* determined inhibition constants for HBP, NTAM, and BIPH of 0.5 μM, 1.8 μM, and 17 μM, respectively. NAPO and BCA were reported as non-inhibitory at the examined concentrations (1 mM). From these values, binding free energies were calculated through Equation 2, ΔG_b^{exp}, for comparison to calculated binding free energy, $\Delta G_b°$ (Table 4.1). The calculated binding free energies matched experimental values well, within ~1 kcal/mol, and predicted the correct inhibitory order among the examined molecules. We expect calculations can be further extended to investigation of the many additional putative *DszB* inhibitors present in crude oil-derived feedstocks.

FEP calculations, when compared with the non-bonded interaction analysis, indicate molecule hydrophobicity significantly impacts the effectiveness of each inhibitor. All inhibitors studied here exhibited a similar trend in the individual elements of binding free energy, with the dispersive component dominating the favorable contribution and the electrostatic and repulsion components accounting for weaker or even unfavorable binding elements. Despite NTAM's greater stability within the *DszB* active site, as evidenced from the COM distance analysis and nonbonded interactions, HBP's high affinity for the gap region, along with its aversion to hydrophilic solutions, proves a significant enough barrier to account for the weaker active site interactions it displays. While the efficacy of each inhibitor does not trend with hydrophobicity, much of their binding affinity is likely derived from the entropic effects of burying these hydrophobic molecules within protein core away from the exterior aqueous environment. Individual differences in the structure of these inhibitors plays a role in their total binding affinity. In the case of NTAM, these interactions are significant enough to produce more favorable binding than the more hydrophobic BIPH but still fail to overcome the considerably more hydrophobic HBP.

4.5. Conclusions

Experimental studies of crystal structures and product inhibition motivated a desire to understand the underlying molecular-level mechanisms. Florescent kinetic assays of *DszB*-catalyzed conversion of HBPS confirmed the resulting product's effectiveness as a competitive inhibitor, showing an affinity for the active site comparable to that of HBPS. Basic structural characterization of *DszB*'s response to HBP via circular dichroism then indicated a shift in protein helical content indicative of a conformation change, which was consistent with observed substrate-induced binding from Lee et al. Based on this, we hypothesized that HBP and structurally similar competitive inhibitors may elicit a similar, if not identical, conformation change to that of HBPS.

All inhibitors studied here, but especially HBP, share enough structural similarity with substrate HBPS to occupy the active site; however, once bound, their inhibitive nature is derived less from the expected specific interactions with key catalytic residues and more from a collection of small but numerous hydrophobic interactions. Docking calculations and subsequent COM distance analysis indicated all of the inhibitors favored a binding position. NTAM and BIPH active site binding would compete directly with HBPS, whereas HBP favored a binding position slightly away from active site, nearer to all three loops of the gate region. Analysis of nonbonded interactions from MD simulations confirmed that the inhibitors exhibit varying degrees of nonspecific interactions with the hydrophobic residues either in the active site or around the proposed loop regions. Finally, FEP/λ-REMD calculations defined the binding free energies of the inhibitors and indicated the effectiveness of each was derived from a balance between the more dominant dispersive elements of free energy contrasted with more subtle shifts in the repulsive and electrostatic components. Collectively, this demonstrated each of the inhibitors studied were entropically driven to remain in the active site by their hydrophobic nature and that their structural differences play a smaller role in their efficacy.

Overall, decreasing experimentally observed binding affinity of ligands (HBP> NTAM > BIPH > NAPO > BCA), from inhibition constants, corresponded with less favorable calculated binding free energies. Accordingly, we have accurately simulated inhibition effects (on a macroscopic scale) and used the molecular-level information to provide insights into the mechanisms by which these inhibitors deactivate *DszB*. Product inhibition of *DszB* by HBP remains one of the greatest barriers not only to *DszB* functionality, but to the industrial effectiveness of the 4S pathway as a whole. Given the unavoidable presence of HBP as the primary product of the 4S pathway, future research in the field will need to focus on methods to alleviate product and pathway inhibition. The unique mechanism of HBP inhibition, which we propose results from ligand entanglement at the gap region, serves as a starting point for future rational engineering approaches. Future *DszB* improvements via site-directed mutagenesis should consider all residues in the gap regions as potential targets for mutation to generate a desulfinase less impeded by product concentration. While not an exhaustive list, HBP stability around loop 1 (residues 50-60) should be considered high priorities for substitution. While this study does not consider the impact such mutations may have on stability, substitutions to residues within the loop 1 that reduce the size of attached functional groups may serve to widen the gap region, thereby reducing the number of interactions HBP may engage in while exiting the active site and, in turn, its effectiveness as an inhibitor. Alternatively, exploring changes in solvation environment for *DszB*, attempting to introduce hydrophobic or surfactant molecules into its reaction buffer, may help to decrease inhibitor aversion toward free solution.

Chapter 5 – The Impact of Solvation Environment on the Stability of Desulfurization Enzyme 2'-hydroxybiphenyl-2-sulfinate desulfinase (*DszB*)

This chapter has been adapted from a publication currently in draft. All work in this chapter was performed by the author of this dissertation.

5.1. Introduction

Liquid transportation fuels derived from petroleum provide the majority of energy used in the transportation sector. [11, 14] Current processes for refining crude oils into these petroleum products allows some sulfur-containing molecules to pass through from oil to fuel, where, upon combustion, the sulfur compounds oxidize to sulfur oxides (SOx). [82] Acutely toxic under both the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), SOx represents both an environmental threat, encouraging acid deposition (acid rain), and a human health concern, causing respiratory irritation. [110]

Hydrodesulfurization, the current industry standard for sulfur removal in crude oil refining, utilizes 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. [5, 10, 12] 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 remain intact; these DBT derivatives now account for 60% of sulfur emissions from liquid transportation fuels. [10, 42]

Biodesulfurization (enzymatic sulfur removal) offers highly specific, rapid thiophenic desulfurization at ambient temperature and pressure, exceeding traditional catalyst performance in nearly every metric. [6, 18, 111] For most microorganisms, sulfur fixation pathways exist as a necessary mechanism for survival, although the mechanism by which it is achieved can differ greatly. [6, 112-113] When considering aromatic sulfur-containing molecules such as DBT, these mechanistic approaches can be broken down into two categories: ring destruction and sulfur specific. [9, 17, 30] In “ring destruction” approaches, the sulfur containing ring is, as implied, first destroyed (via one if not multiple bond disruptions) to access the sulfur more directly. In sulfur-specific mechanisms, the ring is opened by breaking only bonds shared with the sulfur targeted for fixation. [39] Given that the “ring destruction” approach disturbs carbon-carbon bonds essential to a fuel’s heating value, biodesulfurization mechanisms of industrial interest are limited to the ring opening approaches that break a minimum number of bonds to release the target sulfur. [20]

One such naturally occurring mechanism, shared by numerous organisms in sulfur limiting conditions, is a 4-step catabolic pathway (the 4S pathway) that converts DBT and many of its derivatives into 2-hydroxybiphenyl (2-HBP) without disruption of the carbon-carbon bonds. [17-18, 37] While present in many bacteria, this pathway is most studied in the soil bacteria *Rhodococcus erythropolis*, strain IGTS8, in which the pathway was first isolated. The focus on *R. erythropolis* can be attributed to fact that, in most other organisms, diffusion of substrate into the cell is rate limiting, a complication not shared by *R. erythropolis*. [1, 26, 38]

With the aid of oxidoreductase, *DszD*, and several NADH cofactors in its first 3 steps, the 4S pathway begins with catalytic oxidation of the ring-bound sulfur via the monooxygenase, *DszC*. The second step again utilizes *DszC* to further oxidize the sulfur. In both cases, free water molecules supply oxygen. Following oxidation, the sulfur-containing ring is opened via disruption of one of the C-S bonds via a second monooxygenase, *DszA*, leaving both a hydroxyl group and sulfinate group on opposite rings of phenylbenzene, forming 2’-hydroxybiphenyl-2-sulfinate (HBPS). Desulfurization occurs in the final step of the pathway catalyzed by the desulfinase *DszB*, which cleaves the sulfate group from HBPS to form the product, 2-hydroxybiphenyl (HBP). [17, 25, 32, 39]

One major limitation of the 4S pathway is 2’-hydroxybiphenyl-2-sulfinate desulfinase (*DszB*), the rate-limiting enzyme in this biocatalytic process, as it exhibits both a low turnover and experiences product inhibition. [41-42] Investigations have been performed evaluating the activity of *DszB* many even probing the concerns with inhibition. [24, 35, 40] Recent studies in product inhibition indicate the hydrophobicity of the HBP product and several other inhibitory molecules, play a role in their effectiveness as inhibitors, with their aversion to aqueous environments driving them to remain buried in the protein active site. This particular barrier in *DszB* functionality suggests changes in *DszB*’s solvation environment may facilitate egress of the hydrophobic product from the protein core, reducing product inhibition effects.

The bulk of available studies related to *DszB*, however, focus on directed evolution of the entire 4S pathway and host organism rather than characterizing an isolated protein. Thus, the possibility of free enzyme cocktails as a means of desulfurization is overlooked, and research fails to prioritize one of the most essential features required in an industrial catalyst, stability. To better characterize *DszB* stability and gain insights into ways by which its life time may be extended we seek to evaluate the basic thermal and kinetic stability of *DszB* in a variety of solvation environments. [114] Such research also serves a secondary purpose, paving the way for future investigations in inhibition regulation, as recent studies emphasis the importance of reaction medium on basic activity. In this study we examined thermal stability of *DszB* in a wide range of different commercially available buffer additives using differential scanning fluorimetry (DSF) as a means to quickly identify favorable changes in protein melting point. [115] Additionally, using a fluorescent kinetic assay we investigated *DszB* reaction rate over a 48hr time scale in a more focused group of buffer environments to assess the life time of enzyme activity.

Buffer additives tested were chosen from a Hampton solution and solubility screen and include components that are amino acid derivatives, chaotropes, chelators, cyclodextrins, ionic liquids, linkers, metals, non-detergents, sulfobetaines, organic acids, osmolytes, peptides, polyamines, polymers, polyols, and salts. Solution additives share a collective non-specific affinity for proteins in general that work to stabilize or better solubilize. The precise mechanisms by which each category achieves such effects can vary greatly, but, generally speaking, each works to either shift equilibrium in favor a particular protein fold, or to modify properties of the bulk solution (such as ionic strength) to increase solubility.[116]

5.2. Methods

5.2.1. Materials

All chemicals and materials without specifically stated origins were purchased through VWR (Radnor, PA). Protein purification supplies including immobilized metal affinity chromatography (IMAC) column (HIS-Select® HF Nickel Affinity Gel), Polyacrylamide Gel Electrophoresis (PAGE) gels, and associated Sodium Dodecyl sulfate running buffer and loading dye were purchased from Millipore Sigma. An EZ-Run™ Protein Ladder standard was purchased from ThermoFisher. Substrate HBPS was synthesized from biphenosultine according to the methods discussed in Yue et. al.; Texas A&M University National Products LINCHPIN Laboratory provided biphenosultine as gift.[94]

5.2.2. Cell culture

All protein was produced via transformed *Escherichia coli* BL21 containing *pTAC-MAT-Tag-2* (Sigma E5405, C-terminus MAT Tag) plasmid vector with the *DszB* gene derived from *Rhodococcus erythropolis* IGTS8 inserted, as well as molecular chaperone producing vector *pG-KJE8* (Takara 3340). All *E. Coli* cultures were prepared in Lysogeny Broth (LB) containing 100 mg/ml ampicillin and 50 mg/ml chloramphenicol for *pTAC-MAT-Tag-2* and *pG-KJE8* selection, respectively. All cultures were maintained under constant agitation and temperature control via an orbital shaker table. A 10 ml seed culture was grown in LB at 37 °C overnight. The seed culture was then transferred to a large culture flask containing 2 L of LB and left to grow to OD600 of 0.8, at which point culture temperature reduced to 25 °C for protein expression. *DszB* expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a concentration of 1 mM, while molecular chaperones *GroEL–GroES* were co-expressed for increased folding and stability via L-arabinose at a concentration of 2 mg/ml. Cultures were pelleted via centrifugation at 5,000 x g for further use.

5.2.3. Protein Purification

Cell pellets from cell culture were suspended in 50 ml of chilled (4 °C) equilibration buffer (50 mM sodium phosphate, 0.3 M sodium chloride, and 10 mM imidazole) and lysed via pulse sonicating on ice with a sonic dismembrator (Fisher Scientific model 505). Liquid cell lysate was isolated from insoluble cell debris via centrifugation (64,000 xg for 30 min at 4 °C). Supernatant was then filtered with a 0.22 µm syringe filter for removal of small residual cell debris. Protein purification was achieved via IMAC using a chilled HIS-Select® HF Nickel Affinity Gel pre-equilibrated with 5 column volumes of equilibrium buffer. [117] Filtered cell lysate was gravity fed into the IMAC and then flushed out with an additional 5 column volumes of equilibration buffer. Protein was eluted from the IMAC column using column elution buffer (50 mM sodium phosphate, 0.3 M sodium chloride, and 100 mM imidazole). Purity was evaluated via sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS PAGE) on a 12% polyacrylamide gel (Supplemental 1). Bands for molecular chaperones *GroEL* and *GroES* were observed on the SDS PAGE in purified product. [23]

5.2.4. Differential Scanning Fluorimetry (DSF)

DszB stock solution, freshly purified from IMAC, and suspended in elution buffer at a concentration of approximately 0.15 mg/ml was used to dilute Sypro Orange dye concentrate 1:1000 to form a *DszB*-Sypro solution. In a 96-well clear, non-skirted, low profile microplate, 10 µl from each solution in the “Hampton Solubility and Stability Screen”(HR2-072) collection was transferred to an equivalent well position. [114-115] 10 µl of *DszB*-Sypro solution was then added to each well, and the plate was sealed and centrifuged for 1 min at 500 x g to settle each solution. A real time quantitative PCR instrument (Bio-Rad CFX96 Touch) (qPCR) was then employed, starting all samples at 20 °C and raising the temperature in 1 °C increments every 1 min up to a final temperature of 90 °C. Emission around 600 nm from were monitored at each temperature increment. Inflection points observed in florescent intensity over the rise in temperature of each sample were quantified by identifying the corresponding temperatures to peaks in the second derivative and recorded as *DszB*'s melting point. The top 4 buffers yielding the most positive impact on thermal stability were selected for further testing.

5.2.5. Fluorescent spectroscopy

IMAC-purified *DszB* suspended in buffers of interest (as determined by DSF) at a concentration of 0.2 mg/ml was added to well plates containing 2-fold serial dilution of HBPS in imidazole-free elution buffer spanning a concentration range of 1.25 μM to 40 μM HBPS. *DszB* solution was also added to wells containing only buffer to serve a control. Samples were incubated at 20 °C and monitored for fluorescence emission at a 414 nm wavelength using a BioTek Synergy plate reader with an excitation wavelength of 288 nm. Using a calibration curve generated from the linear relationship between HBP concentration and fluorescence intensity, initial reaction velocities were calculated to generate a Michaelis-Menten saturation curve. [24-25] Apparent Michaelis-Menten parameters were determined numerically using a Generalized Reduced Gradient (GRG) algorithm.

For initial time lapse kinetics, the *DszB* solution buffer of interest was column elution buffer. The above process was executed for *DszB* immediately after purification and again at 2, 6, 12, 24, and 48 hrs, with the *DszB* solution stored at 4 °C while not in use. For expanded kinetic stability trials, this method was repeated for several different *DszB* solutions, again in column elution buffer, but with the 4 additives from the Hampton solubility and stability screen collection that yielded the most positive impact on thermal stability as determined by DSF. Additionally, to evaluate the impact of potential anti-inhibitor surfactants, a 1 % TWEEN20 elution buffer was also tested. The turn-over number ($k_{\mathrm{cat,t}}$) for each solution and time point was determined and compared to its corresponding initial turn-over number observed immediately after purification ($k_{cat,0}$). The ratio between $k_{cat,t}$ and $k_{cat,0}$ was used as a measure of activity decline.

5.3. Results

5.3.1. Initial Time Lapse Kinetics

The catalytic stability of *DszB* was evaluated by measuring the impact of time on initial reaction velocity in a standard column elution buffer, revealing an extremely low enzyme lifetime. Michaelis–Menten saturation curves were generated for *DszB* catalyzed conversion of HBPS in elution buffer immediately after purification (0 hrs) and at 2, 6, 12, 24, and 48 hrs thereafter, with aliquots remaining in 4 °C cold storage while not in use (Figure 5.1). Data indicates an extremely rapid decrease in *DszB* activity observably even in the first 2 hrs post purification. At the 12 hrs, *DszB*-catalyzed reaction rates approach the limit of detection. Such a short enzyme lifetime marks *DszB* as a very unstable protein.

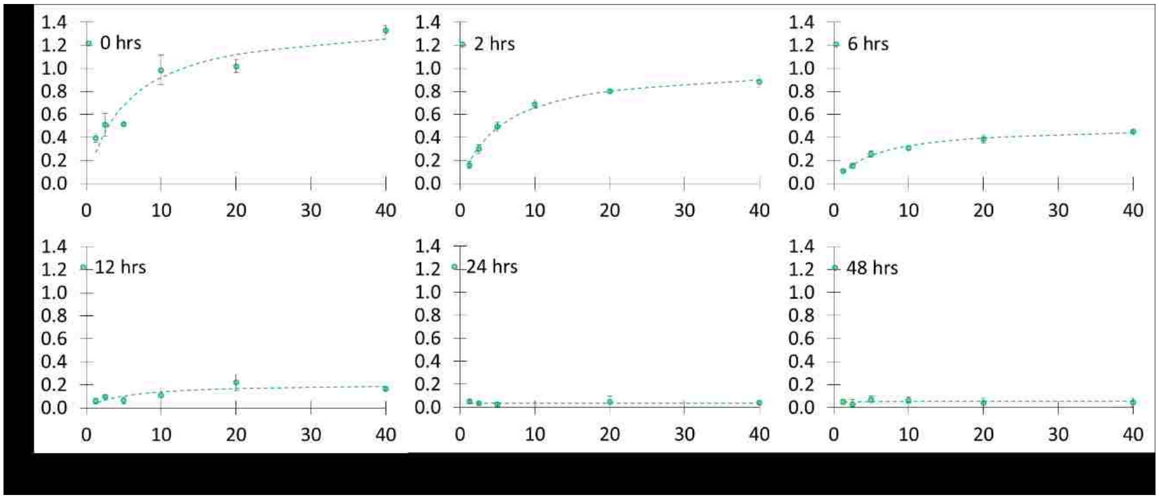

Figure 5.1. Time lapse of Michaelis–Menten kinetic plots

Michaelis–Menten kinetic plots for DszB stored at 4 °C in column elution buffer. Time zero indicates activity measurement taken immediately following IMAC purification. Error bars represent standard deviation with n=3.

5.3.2. Thermal Stability Screening

The thermal stability of *DszB* was assessed by measuring protein melting point via DSF, indicating that few general classes of stability buffers have a consistently positive effect on *DszB* stability. Samples from freshly purified *DszB* in elution buffer combined in equal volumes with the 96 solutions from the Hampton Solubility and Stability Screen (HR2-072) were subjected to DSF. The melting points of each solution additive with resolvable spectra that did not precipitate or otherwise immediately denature were determined (Table 0.2) and compared to the water control. The average difference between each group of solution additives from the water control are shown in Figure 5.2. Average DszB protein melting point temperature differentials (°C); a positive difference indicates the elution buffer plus additive group's melting point was higher than that of elution buffer plus water (melting at 45 °C) and, thus, had a positive impact on thermal stability.

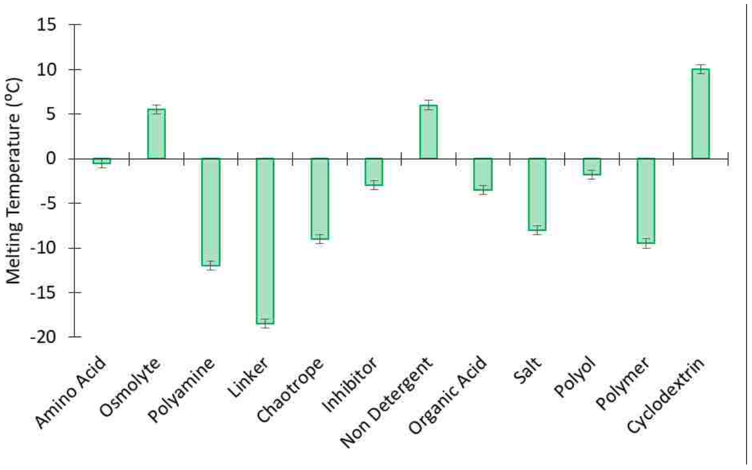

Figure 5.2. Average DszB protein melting point temperature differentials (°C)

Melting point for various buffer additives categories. Melting points differentials are expressed the as the difference between buffer + additive from buffer + water control. Concentrations and specific additives used reflect what is available via Hampton Solubility and Stability Screen. Error bars indicate the half degree temperature span sampled in DSF measurement.

Solutions from the Hampton screen in the peptide, reducing agent, chelator, metal, and ionic liquid groups universally caused either precipitation or yielded spectra too unstable to obtain a melting point from and, thus, are not displayed. Results indicate osmolytes, non detergents, and cyclodextrins were effective at improving *DszB* stability. Other additives showed, on average, deleterious effects on protein melting point. The four best preforming additives (additives who yielded the highest positive shift in *DszB* melting temperature) were selected for further investigation: (1) 50 mM methyl-β-cyclodextrin (MBCD) a cyclodextrin at 55 °C, (2) 2 M xylitol an osmolyte at 52 °C, (3) 2 M sucrose an additional osmolyte also at 52 °C, and (4) sulfobetaine 201 (SB201) a non-detergent at 51 °C (temperatures indicate melting point of *DszB* with buffer containing additive). These high performing additives were from a variety of groups rather than from a single superior category; some categories that were, on average, deleterious still contained individual additives that were moderately favorable or at least neutral. *DszB* stability appeared to be independent of any predefined grouping in the solution screen, apart from osmolytes proving universally beneficial to stability.

5.3.3. Catalytic Activity Response to Solvation Environment

The influence of buffer additives on the catalytic stability of *DszB* was determined using a comparative ratio between the initial and time-resolved turnover numbers, revealing a significant positive impact resulting from several of the buffer additives. Michaelis-Menten parameters were calculated from initial reaction velocities for several of the higher performing buffer additives identified in the thermal stability screen (MBCD, xylitol, sucrose, and SB201). We also examined the effect of adding TWEEN20, a surfactant, to the buffer on *DszB* kinetics; the Hampton stability screen did not include a surfactant grouping, and the nature of such molecules is of interest for future investigations into product inhibition. A ratio between the observed turnover *(k*$_{cat,t}$) at a given time point and the initial turnover number *(k*$_{cat,0}$) at the time of purification was determined for time points 0, 2, 6, 12, 24, and 48 hrs along (Figure 5.1).

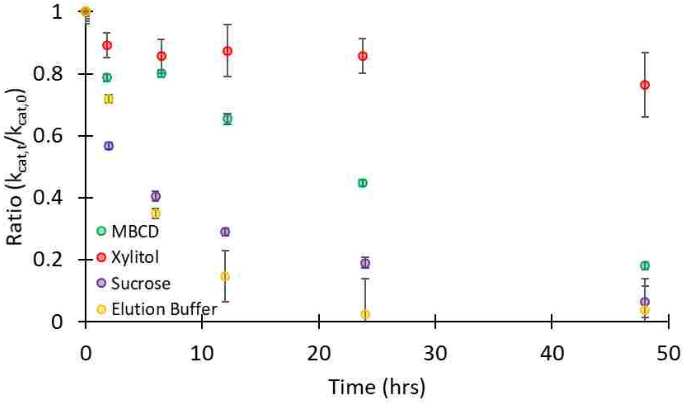

Figure 5.3. Change in turnover number over time for stability-enhancing buffer additives.

The y-*axis is a ratio between the apparent turnover number observed after time stored at 4 °C and the turnover number determined at time zero (*$k_{cat,t}/k_{cat,0}$*). Error bars standard error.*

TWEEN20 and non-detergent SB201 showed no initial activity, indicating either the molecules served as strong inhibitors themselves or, more likely, denatured *DszB* entirely. MBCD, xylitol, and sucrose all had a positive impact on protein stability, although xylitol maintained stability consistently over the entire time course. Unfortunately, most improvement is isolated to the first 24 hrs, with *DszB* activity significantly declining by the 48 hr mark, even in xylitol, the best preforming additive tested, activity falls below 80%. While this xylitol performance increase stands considerably above the rest and offers a solvation environment better suited for extended bench top experimentation, ultimately it is still provides only a short term increase in DszB stability.

5.4. Discussion

Initial time-lapse studies of *DszB* activity showed that, even stored at refrigerated temperatures (4 °C), enzyme lifetime rapidly deteriorated in less than a day. The swift loss of activity is a clear indication of poor protein stability. The buffering capacity of a standard phosphate-based column elution buffer proves sufficient for regulating the pH of our protein sample but does little else to stabilize *DszB*'s fold. From a practical standpoint our interests are focused on the kinetic stability of *DszB*, prolonging the among of time *DszB* remains active. However, stability when viewed in thermal context through the shift in melting point of *DszB* can prove predictive of the impact a change may have on the kinetic stability of *DszB*. Thermal shift allows evaluating a much larger pool potential solvation environments with much smaller sample sizes and any solution thermally stabilizing *DszB* likely also prolongs its activity.

While 94 additives were initially investigated in the thermal stability screen, only a fraction of those tested had clear enough spectra to resolve a melting point. While a large amount of this interference can be attributed to expected phenomenon, including protein precipitation in response to destabilizing solutions or such unfavorable solvation environments so as to immediately denature *DszB*, it is likely some spectra noise was derived from the presence of molecular chaperones lingering from purification. *DszB* has been shown to retain some affinity for chaperones groES and groEL and folds poorly when expressed without their aid. Trace amounts of such folding proteins can be seen in our own samples (Figure 0.2), and their response to DSF dye across the temperature gradient likely provides some interference that, in select cases, elevates the background fluorescence to obscure melting point detection. [34] Thus, some solution groups such as peptide, reducing agent, chelator, metal, and ionic liquid groups were unable to be measured. Of the remaining groups, few showed any real trend among their effectiveness or lack thereof at stabilizing *DszB*.

While the cyclodextrin MBCD provided *DszB* with enhanced thermal stability, and even extended catalytic stability considerably in the short term, it was the only solution of its group to resolve, preventing us from drawing more general conclusions about cyclodextrins. MBCD did not significantly enhance long-term stability of *DszB*. Similarly, SB201 was the only “non-detergent” in the screen providing any measure of thermal stability, although SB201 also completely deactivated *DszB*. *DszB* is known to bind alternate substrates beyond HBPS, including a range of inhibitors. Most binding events stabilize a protein and can be observed via a shift in thermal stability. [43-44] It is possible that the increase in *DszB* thermal stability resulted from binding the sulfur-containing SB201 in an inhibitory manor. Yet another, and perhaps more likely possibility given the general sulfobetaine focus on solubility rather than fold stability, is that SB201 denatures a particular domain of *DszB*, and the DSF data reflects a separate *DszB* domain or one of the trace molecular chaperones melting at a higher temperature. [118]

Osmolytes were the one molecule class that showed a wholly positive impact on thermal stability, with all multiple solutions having resolvable melting points and two (xylitol and sucrose) ranking among the best performing. Osmolytes stabilize proteins by competing for solvation water. Since the unfolded state of a protein generally exposes it to more solvent interactions the presence of a competitor such as an osmolyte bias equilibrium toward a folded protein requiring less solvent interactions.[119] Thus, osmolytes drive the equilibrium between folded and unfolded protein toward the folded state by more heavily destabilizing the non-native conformation than the native one. [120] This results in a comparative shift in equilibrium toward the protein's native state. *DszB* relies on a relatively large-scale conformation change to achieve catalysis, however, and may find an osmolyte-based solvation environment biasing the more closed of the two conformations, potentially interfering with activity. [43, 45] Notably, initial *DszB* activity in the presence of xylitol and sucrose was similar to the unmodified elution buffer standard. Accordingly, osmolyte effects dampening *DszB* conformational change appear to be minimal.

The hydrophobicity of 4S pathway product HBP along with other common crude oil contaminates, was shown in recent studies to enhance their effectiveness as inhibitors. The apprehension towards entering an aqueous environment drives these molecules to remain buried in the more hydrophobic protein. While the presence of any amphiphile would almost certainly destabilize *DszB* to a degree, ideally, in small concentrations, such a molecule may alleviate *DszB* product inhibition by encouraging the hydrophobic HBP product to leave the protein core. To this end TWEEN20, a common surfactant occasionally included in IMAC wash buffers, was evaluated alongside other solution additives to assess the general impact emulsifiers may have on *DszB* kinetic stability. While TWEEN20 did seem to immediately denature *DszB* even at a low concentration, this is by no means conclusive for surfactants, or even emulsifiers as a whole. Several other additives tested in the screen possessed amphiphilic qualities but failed to resolve or show a significant positive impact on *DszB* melting point. With the necessary hydrophobic elements of these molecules inherently destabilizing *DszB* in order to better sequester product HBP, its clear testing their addition in isolation from other additives as preformed in these studies is insufficient. A more focused study testing a broader range of such additives at multiple different concentrations is required to appropriately address the effectiveness of amphiphiles in regulating product inhibition.

Osmolyte sugars (sucrose and xylitol included) develop the three-dimensional hydrogen-bonded structure of water, which strengthens hydrophobic interactions within the aqueous environment. [121-122] This phenomenon is also attributed to their positive impact on thermal stability but has the added benefit of reducing the free energy required to transfer a hydrophobic molecule in an aqueous environment to a surfactant micelle's interior. [121] Such an effect implies that a mixed system, containing both xylitol and surfactant, may provide the desired protein stabilizing effect from osmolyte, as well as reducing inhibition by strengthening the effectiveness of the surfactant at stabilizing hydrophobic product HBP unbound in solution. While this study failed to identify a suitable surfactant to pair with xylitol, future investigations into a more diverse set of emulsifiers at more variable concentrations may yield a solution to product inhibition in addition to prolonging *DszB* stability.

5.5. Conclusion

Initial kinetic fluorescent assays revealed an extreme lack of stability, even short-term, in *DszB,* with activity in a standard phosphate elution buffer declining to near undetectable levels after just 12 hours of refrigerated storage. Evaluation of a wide variety of buffer additives from numerous different molecular groupings revealed osmolytes to be effective at stabilizing *DszB,* and indicated several buffer components that may have a strongly positive effect on the lifetime of *DszB* activity. Investigation into the effects of the thermal stability-enhancing buffer systems, MBCD, xylitol, sucrose, and SB201, on activity revealed that while SB201 demonstrated a deleterious effect on even initial *DszB* reaction rates, the remaining three additives had a highly positive impact on short-term protein stability, with 2 M xylitol performing the best. Kinetic stability investigations also extended to TWEEN20 due to its potential for reducing product inhibition, however much like SB201, TWEEN20 deactivated *DszB* completely.

Of the solvation environments tested, osmolytes provided the most consistent improvement in *DszB* stability, with their competition for solvation water driving equilibrium away from the more solvent exposed denatured protein conformation, with the xylitol additive yielding the best kinetic performance from those tested. While a step toward increased stability, limitations of *DszB* exceed simple correction by solvation environment alone, necessitating further investigations into more direct protein alterations such as site directed mutagenesis or post-translation modification.

Chapter 6 – The hydrophobicity and conformations of common glycosylation motifs across the kingdoms of life

This chapter has been adapted from a publication currently in draft. All data processing and analysis in this chapter was performed by the author of this dissertation. All MD simulations were performed Dr. Christina M. Payne, the director of this dissertation. The author of this dissertation authored the entirety of this chapter and corresponding publication.

6.1. Introduction

Post-translational modification of proteins is an essential phenomenon occurring in all living organisms. The modifications enable critical biological processes, such as but not limited to: physically attaching FMN to an oxidoreductase to impart catalytic function, and acetylating residues to introduce structural uniqueness essential in signaling pathways. Glycosylation is one of the more commonly employed post-translational amendments, occurring in every kingdom of life to some degree. Glycosylation, usually beginning in the endoplasmic reticulum, involves the covalent attachment of a saccharide molecule to the resulting glycoprotein. The glycans can be appended to proteins through O-linkages at a serine or threonine residue or N-linkages at an asparagine residue.[51-53] Unlike proteins, which are formed from a pool of 20 amino acid building blocks and linked linearly through a consistent peptide bond, glycan structures have the potential to be considerably more diverse. Drawing from a much wider selection of monomeric building blocks, potentially including any bioavailable sugar, glycans appear in a vast assortment of lengths, ranging from simple monomeric attachments to chains hundreds of monomers long. Further diversity arises from the variety of ways in which sugars can be covalently linked, using either an α or β conformation at one of several available carbons on a given saccharide monomer. In many cases, the linkages form complex branched structures. The resulting size, chemical and structural variability, and generally dynamic quality of most glycans makes structural studies of glycoproteins difficult.

Glycans serve a variety of functions in nature including signaling, recognition, folding, and stability. While glycosylation enables these functions across all kingdoms of life, the types of glycans employed, in terms of preferred monomers and linkage types, vary considerably between organisms.[60-61] For example, bacteria and archaea typically employ simple forms of O-glycosylation, favoring mostly single monomer glycans of common saccharides such as glucose (Glc), galactose (Gal), and N-acetyl-galactosamine (GalNAc).[123-124] Some chain elongation can also occur, especially in bacteria.[51]

Eukaryotic organisms generally exhibit more complex glycosylation patterns. Plantae typically produce N-acetyl-glucosamine (GlcNAc), GalNAc, and Fuc-based glycans, with chain elongation generally occurring with Gal additions; there are no conserved O-glycan motifs present across the entire Plantae kingdom.[51] And while plants exhibit more diverse, organism specific glycosylation trends, this is not the case for other eukaryotes. Fungi, especially yeast, favor mannose (Man)-based glycans almost exclusively.[65-67] Linear Man chains are a commonly occurring motif across fungi, with occasional further functionalization via the addition of alternative saccharides at branch points and end caps such as xylose (Xyl), Gal, and Glc. In general, Man is a very versatile glycan building block employed in nature, often starting a branching point in more complex motifs. Man is even employed in the animal kingdom in cell adhesion proteins such as cadherin and dystroglycan.[125]

In the Animalia kingdom, a wide variety of O-linked glycans are employed. Fuc, Glc and GlcNac are used in the glycosylation of epidermal growth factor (EGF)-like proteins to separate and stabilize two domains that would denature in the absence of glycans. [51, 68] Xyl is employed as an initiator for the creation of proteoglycans (heavily glycosylated glycoproteins), where further chain elongation with Gal, GalNAc, and glucuronic acid additions build more complex glycans. In extracellular matrix proteins such as collagen, glycosylation with Gal with Glc additions are employed.[17]

In the case of N-linked glycans, the fundamental initiation motif is highly conserved across most kingdoms, with the majority of such glycans falling into one of three primary categories: a branched predominantly mannose structure ($Man_{2-6}Man_3(GlcNAc)_2$), a complex multi-component structure ($(GlcNAc)_2Man_3(GlcNAc)_2$), and a hybrid structure combining elements of both mannose branching and complex glycan diversity ($Man_2GlcNAcMan_3GlcNAc_2$). All three motifs are built upon the conserved $Man_3GlcNAc_2$ structure as shown in Figure 6.1.[62, 69-72]

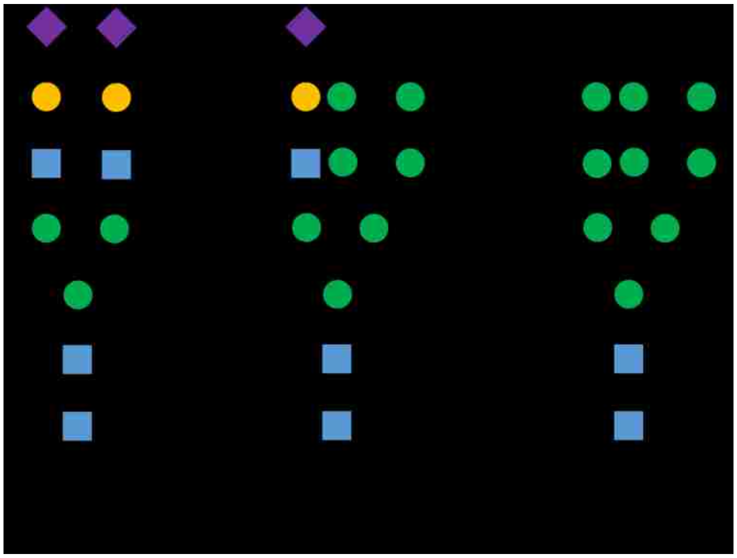

Figure 6.1. Common N-glycan motifs in nature

With so many glycosylation motifs observed in nature and many more theoretically feasible, there is much to gain through systematic study of the purpose of each naturally occurring glycan and the extent to which it modifies the chemical properties of its host protein. To this end, we set about to examine the role of several common O- and N-linked glycan motifs in stabilizing protein structures.[63] As many proteins function in primarily aqueous environments, the effect of glycosylation on solvation free energy provides a simple and effective metric for assessing the impact of a particular glycan on general protein stability.[126-129] By extension this focus on solvation free energy as a lens for stability provides a more direct measure of solubility, with favorable increases in solvation free energy correlating to increases in aqueous protein solubility. [130-131]

Focusing on simple and single glycoforms builds fundamental understanding of each glycan building block, linkage type, an linkage location, which can be extended to more complex glycoforms. Using thermodynamic integration, we determined the solvation free energies of a variety of glycoforms, both monomeric and short chained and linked in multiple conformations and locations, to all three common glycosylation residues serine, threonine, and asparagine. The chemical structure, names, and shorthand notation for the collection of common monosaccharides examined in this study and from which many naturally occurring glycans are built are given in Figure 6.2. Our results form a library of fundamental glycan building blocks to enhance understanding of more complex protein-carbohydrate structures moving forward.

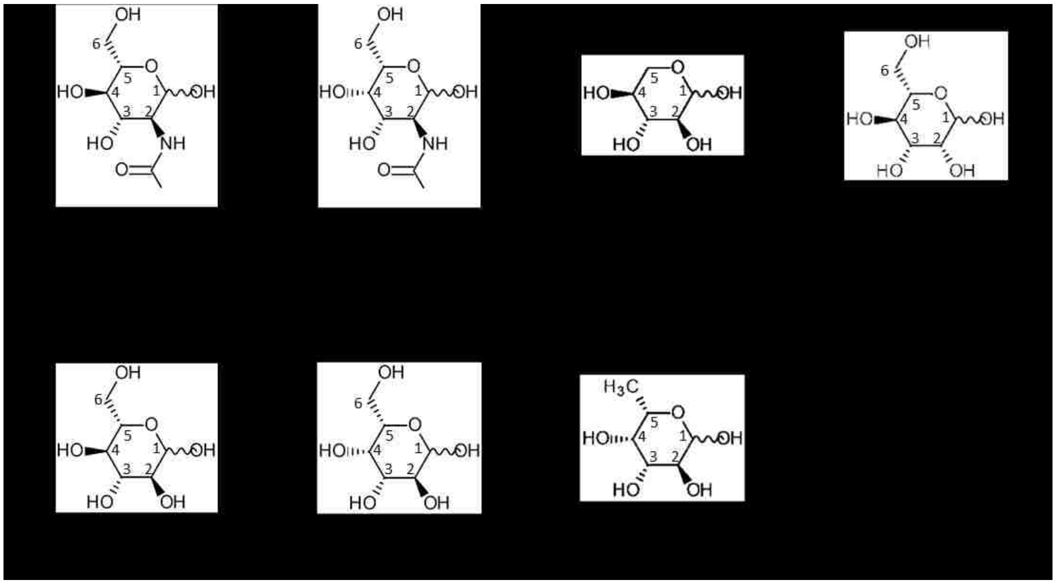

Figure 6.2. Short form names and line-angle formula depicting structures of commonly occurring glycan monomers in nature

6.2. Methods

To explore the role of glycosylation in protein stability, thermodynamic integration (TI) calculations were performed to determine the change in the free energy of solvation, $\Delta G_{solvation}$, resulting from glycosylation of aspargine, serine, and threonine with a variety of glycans and glycan linkage conformations (Figure 6.2). Model compounds – acetamide, methanol, and ethanol – were used as side chain analogs in place of asparagine, serine, and threonine, respectively, for all simulations to reduce the system to a "minimal protein contribution," reducing error introduced by model proteins or uncapped amino acids. [76, 132-133] A total of 34 glycoform models were constructed, symbolically represented in Figure 6.3 using standard nomenclature defined in Essentials of Glycobiology (Figure 0.7).[51]. Representative of O-linked glycans, Fuc, Gal, GalNAc, Glc, GlcNAc, Man, and Xyl were individually appended to both methanol (Ser analog) and ethanol (Thr analog). Each O-linked glycan was also constructed in both an α and β linkage to both serine and threonine analogs. N-linked GlcNAc, $GlcNAc_2$, $Man(GlcNAc)_2$, $Man_3(GlcNAc)_2$, $Man(\alpha\text{-}1,3)Man(GlcNAc)_2$, and $Man(\alpha\text{-}1,6)Man(GlcNAc)_2$ were appended to acetamide (Asn analog). Monomeric O-glycans were selected because of the universal prevalence across most if not all kingdoms of life. All O-glycan's tested rank among the most common naturally occurring and are all observed in human glycoproteins to at least some capacity. N-glycan tested show the piecewise assembly of the Man3GlcNAc2 motif that forms the basic template from which most other N-glycans are build.

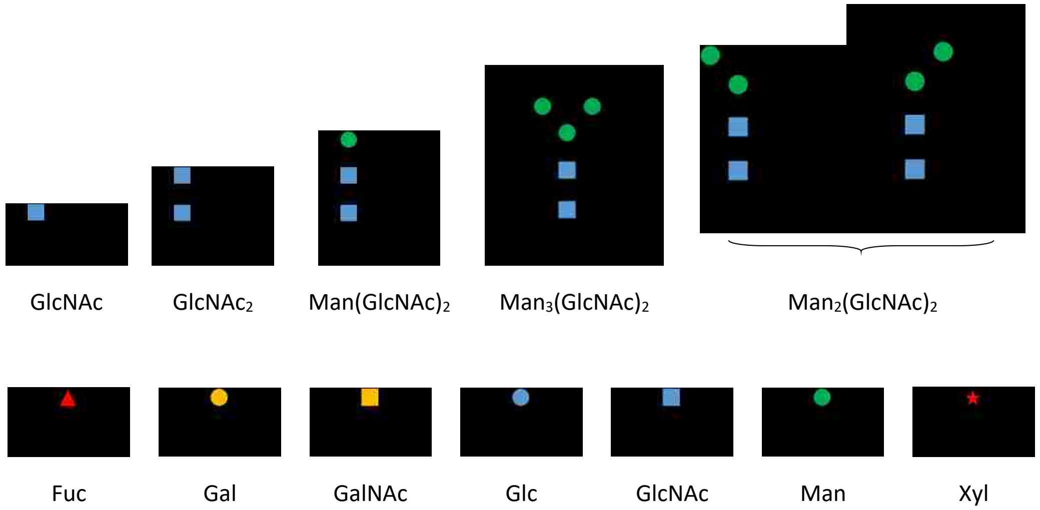

Figure 6.3. O- and N-linked glycoforms

Glycans for which free energy of solvation was determined in this study. Symbol nomenclature for simulated glycans as depicted in "Essentials of Glycobiology. 2nd edition:" [51] *NglcNAc (blue square); Man (green circle); Fuc (red triangle); Gal (yellow circle); GalNAc (yellow square); Glc (blue circle); and Xyl (red star). α and β linkage conformations for N-glycans are as noted.*

The models were built, minimized, and equilibrated in CHARMM. [76, 134 135] The water was minimized for 1,000 steps using the steepest descent algorithm. The entire system was then minimized for an additional 1,000 steps of steepest descent minimization. The systems were equilibrated in the NPT ensemble at 300K for 20 ps. Temperature control was performed using the Nosé-Hoover thermostat. [136-137] SHAKE was used to fix the distances to hydrogen atoms.[138] Non-bonded interactions were truncated with a 13 Å cutoff, and the Particle Mesh Ewald method with a 6th order b-spline, a Gaussian distribution width of 0.320 Å, and a mesh size of 24 x 24 x 24 was used to describe the electrostatics. [139] All equilibration simulations used a 2 fs time step. The CHARMM force field with CMAP correction was used to describe the protein analogs, while the glycans used the CHARMM carbohydrate force field. [76] Water was modeled using the TIP3P force field.[140]

Following equilibration, all systems were integrated with NAMD using identical parameters.[141] Langevin dynamics were used to maintain the system temperature at 300 K with a damping constant of 5 ps^{-1}. System pressure was maintained with a Nosè-Hoover barostat coupled to a Langevin piston at 1 atm with a decay of 100 fs and period of 200 fs.[137, 142-144] The glycans and the amino acid models were solvated in a 26 Å water box of TIP3P water.[88, 140] All simulations were run for 100 ns with a 0.5 ns equilibration time.

The dual topology approach in NAMD was used to achieve an alchemical change between the unmodified amino acid model compound and its corresponding glycosylated form. The approach entails simulation of a single "hybrid" system containing both the analog and the glycosylated analog. The atoms of each compound do not interact with one another but interact with the rest of the system via standard bonded and nonbonded interactions scaled by a coupling parameter, λ, from a reactant to product state in windows over λ. The overall alchemical pathway used calculate glycan impact on solvation free energy ($\Delta\Delta G_{solvation}$) is illustrated in Figure 6.4. Each simulated solvation free energy is the result of two independent simulations in which: 1) the contributions of protein solvation energy are measured via "disappearing" the relevant analog in solution and 2) by "disappearing" the glycoprotein complex between analog and glycan. The resulting effect on glycosylation on protein solvation free energy can then be isolated by taking the difference between the two simulations

[redacted]

Figure 6.4. Thermodynamic pathway used to measure glycan impact on solvation free energy.

The transition was executed piecewise over 15 discrete windows with non-uniform step sizes, such that incremental changes to λ were smallest near 0 and 1 to avoid “end-point catastrophes.” Contributions to solvation free energy from coulombic and van der Waals forces were evaluated separately in each window (sampling at every 2 fs step). Histograms for each window of the N-glycan $Man_3(GlcNAc)_2$ calculation are shown in the Figure 0.8 to demonstrate simulation convergence and sufficient window overlap. A first order trapezoidal rule was sufficient to numerically integrate average dU/dλ over λ=0,1 for each 100 ns simulation window. Error estimation was performed via the methods of Steinbrecher’s et al.[145]

6.3. Results

6.3.1. Thermodynamic Integration

Calculated solvation free energies for each glycan motif (Figure 6.3**)** indicate basic glycosylation motifs will affect protein solubility. As with polysaccharides in general, data indicates a preference for beta conformations and branching. Due to high sampling, solvation free energies are recorded in Table 6.1 exhibit extreme precision, with errors at or below 0.2 kcal/mol. Free energies listed indicated the resulting change in solvation free energy ($\Delta\Delta G_{solvation}$) achieved from glycosylation with each listed glycan.

Table 6.1. Glycan Free energy of solvation (in kcal/mol)

Free energy for each glycan and glycan conformation simulated. Standard error was estimated via the methods of Steinbrecher et al.

	$\Delta\Delta G_{solvation}$		Serine		Threonine	
N-Glycan	(kcal/mol)	O-Glycan	α	β	α	β
GlcNAc	-24.6 ± 0.1	Fuc	-13.3 ± 0.1	-13.4 ± 0.1	-12.2 ± 0.1	-12.2 ± 0.1
GlcNAc$_2$	-37.6 ± 0.1	Gal	-15.1 ± 0.1	-16.5 ± 0.1	-14.1 ± 0.1	-15.2 ± 0.1
Man(GlcNAc)$_2$	-48.2 ± 0.1	GalNAc	-16.7 ± 0.1	-18.7 ± 0.1	-15.7 ± 0.1	-17.5 ± 0.1
Man$_3$(GlcNAc)$_2$	-69.2 ± 0.2	Glc	-14.4 ± 0.1	-15.0 ± 0.1	-13.4 ± 0.1	-13.7 ± 0.1
Man$_2$-1,3-α-(GlcNAc)$_2$	-60.8 ± 0.2	GlcNAc	-16.8 ± 0.1	-18.7 ± 0.1	-15.7 ± 0.1	-16.7 ± 0.1
Man$_2$-1,6-α-(GlcNAc)$_2$	-54.6 ± 0.2	Man	-14.3 ± 0.1	-15.2 ± 0.1	-13.3 ± 0.1	-14.0 ± 0.1
		Xyl	-11.1 ± 0.1	-11.7 ± 0.1	-10.0 ± 0.1	-10.5 ± 0.1

6.3.2. O-Glycans

O-Glycan simulations capture the impact of glycan monomer diversity and linkage conformation on glycoprotein complexes with beta linkages, serine linkages, and more functionalized glycan monomers show more favorable impacts on protein stability compared to their corresponding counterparts. Tabular values for serine and threonine linked glycans from Table 6.1 are displayed in graphical format (Figure 6.5) for ease of comparison. Serine linkages yielded uniformly more favorable changes in solvation free energy compared to threonine. For both threonine and serine linkages, the β-linked conformation generally resulted in the more soluble system, although this phenomenon was largely dependent on glycan functional groups. In the case of Fuc, which exhibits less functionalization of the pyranose ring, differences between the α and β linkages were negligible, effectively identical within error. However, for glycans with multiple functional groups, such as GalNAc and GlcNAc, differences between the two conformations where quite high, differing by more than 10% in favor of the β-linkage, although the differences between the α- and β-linkages where typically less extreme in threonine-linked glycans compared to serine.

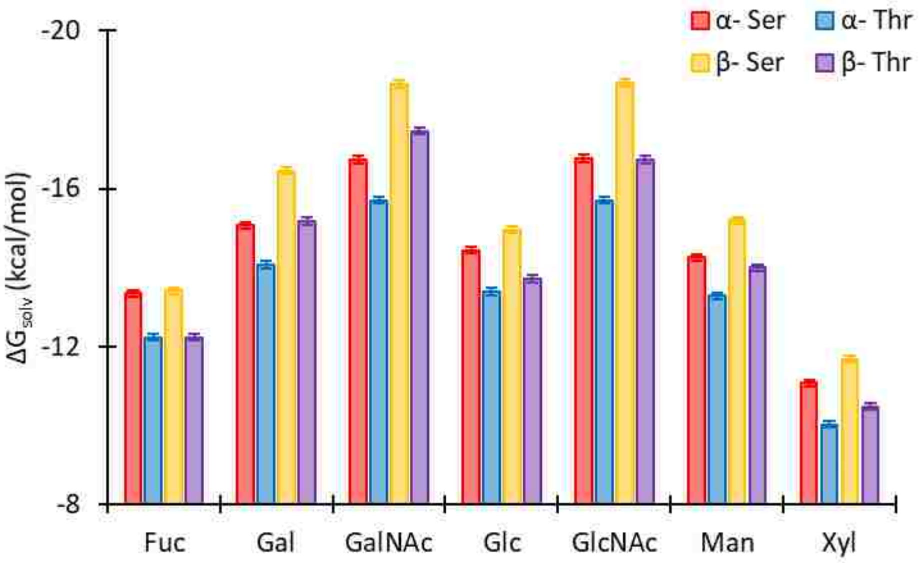

Figure 6.5. Comparison of the $\Delta G_{solvation}$ of O-linked glycans.

Data is grouped by glycan. Each glycan was modeled in both the α- (red, blue) and β-linked (yellow, purple) conformations and linked to either serine (red, yellow) or threonine (blue, purple). Error bars denote standard error estimated via the method of Steinbrecher et al.

6.3.3. N-Glycans

The calculated free energy of solvation for N-linked glycans is more favorably impacted by 1,3 linkages compared to the 1,6, which synergistic effects upon branching being observed. The length of the glycoform also favorably increases solvation free energy. Tabular values for asparagine linked glycans from Table 6.1 are displayed in graphical format (Figure 6.6) for ease of comparison. Direct comparison of the N-linked monomeric glycan, GlcNAc, and its most stable O-linked form, β-serine GlcNAc, indicate N-linkages have a significantly more favorable impact on protein solubility than their O-linked counterparts.

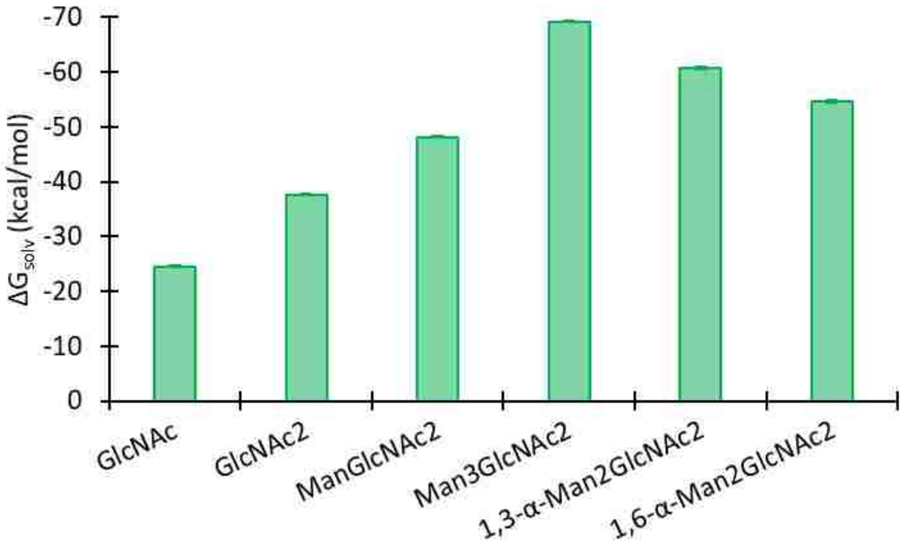

Figure 6.6. Comparison of the $\Delta G_{solvation}$ of select N-linked glycans.

Error bars denote standard error estimated via the method of Steinbrecher et al.

The impact of monomeric additions to the glycan polymer can be observed by comparison of GlcNAc with $(GlcNAc)_2$, as well as $Man(GlcNAc)_2$ and $Man_3(GlcNAc)_2$. As chain length increases, free enegy of solvation becomes more favorable, improving protein solubility; however, multiple monomeric additions to the glycan appears to have a diminishing return on the favorability of the free energy change, with subsequent additions to the chain yielding less of a positive effect than the previous. The effects of linkage position on the hexose ring can be observed in the Man addition to $(ManGlcNAc)_2$. The α-1,3 linkage to form Man-1,3-α-$(ManGlcNAc)_2$ elicits a considerably more favorable impact on solvation free energy than its α-1,6-linked counterpart. A piecewise comparison on the individual contributions of each N-glycan addition is illustrated in Figure 0.9.

6.4. Discussion

6.4.1. Effect of Branching and Bond Location

In general, N-Glycans demonstrated the effects of chain elongation and branching on solvation energy via piece wise assembly of the fundamental motif ($Man_3(GlcNAc)_2$) that initiates most N-glycans in nature. Increasing chain length between the first monomeric GlcNAc and subsequent additions of GlcNAc and Man, up to $Man_3(GlcNAc)_2$, show that, while addition of monosaccharides does impact solvation free energy favorably, there is a diminishing return. This effect is most easily observable between GlcNAc at 24.6 kcal/mol and the less than doubled solvation free energy of $GlcNAc_2$ at 37.6kcal/mol.

Branching and bond position also affects the free energy of solvation, as observed between the $Man_3(GlcNAc)_2$ and $Man_2(GlcNAc)_2$ forms. Between the α-1,3 and α-1,6 linkage, the 1,3 linkage was more favorable. The α-1,3 linkage is prevalent in both unbranched glycans, as well as many polysaccharides used for energy storage across the kingdoms of life, likely due in part to this reason.[51] Isolating the individual contribution of the 1,3- and 1,6-linked Man by deducting the solvation free energy of the $Man(GlcNAc)_2$ simulation yields -12.6 kcal/mol and -6.4 kcal/mol, respectively, totaling 19.0 kcal/mol. However, when similar rational is used to isolate the contributions of both the α-1,3 and α-1,6 Man by subtracting the $Man(GlcNAc)_2$ simulation from the $Man_3(GlcNAc)_2$, the resulting difference is a higher, 21.0 kcal/mol, implying a synergistic effect of branching on solvation free energy. Graphical comparisons of each isolated contribution discussed are illustrated in Figure 0.9.

6.4.2. Influence of Linkage Conformation and Residue Type

The β-1,4 linkage yielded a considerably more stable glycoprotein across all glycans tested. In general, β-linkages are more stable in oligosaccharides due to the more linear geometry resulting from the β-1,4 bond positioning; [51] linked molecules are situated in a more favorable location for hydrogen bonding and experience reduced opportunities for steric clashes between accompanying functional groups on the pyranose ring. When considering glycans specifically, the glycosidic linkage directly between the glycan and protein does not necessarily gain from the prospect of hydrogen bonding, depending on the neighboring residues from the protein side of the linkage. Thus, the 180° bond angle provided by the β-linkage in glycans is more favorable than the α-linkage, due primarily to its linear conformation reducing steric hindrance, by angling the glycan out away from the protein. While this phenomenon was observed for all O-glycans, it is most apparent when considering the diminished significance of the glycosidic linkage conformation on the less functionalized saccharide, Fuc.

In the case of O-glycosylation, most glycans across the kingdoms of life can be appended to any appropriately positioned serine or threonine, yet our results indicate a clear advantage for appendage to serine from a solubility standpoint. [51] It is, however, important to note that both Xyl, the glycan with the least favorable solvation free energy amongst those tested, and Glc appear appended almost exclusively to serine over threonine. [146] While this limitation, like many concerning glycosylation, is a function of the enzymes used produce the initial protein-saccharide linkage, the fact that such a phenomenon exists for a commonly used monosaccharide points to solvation free energy as a possible evolutionary driving force behind the function of select glycosyltransferases. Additionally, the use of Xyl as an initiating glycan in nature occurs largely in the Animalia kingdom proteoglycans, where they are typically attached to multiple sites on the same protein and then further elongated linearly with non-Xyl monomers, usually beginning with Gal.[68, 147] The use of Xyl on multiple sites, with longer chains and with the more favorable serine linkages, collectively suggests the naturally occurring Xly glycans serve to impart solubility.

6.4.3. Impact of Monosaccharide Diversity

Man, used in both initiation and elongation of glycans, is the sole component of nearly all fungal O-glycans, particularly yeast.[65-67] The impact of Man on solvation free energy was moderate, likely due to the absence of the functional groups present in the more favorable GlcNAc and GalNAc; however, this lower degree of functionalization in conjunction with the more epimoric C2 carbon position allows for Man to serve as a branching point, sustaining multiple glycosidic linkages, both α and β in conformation. Even in N-Glycans, the branching points often begin with Man. Among the moderately favorable glycans tested, Glc, Gal, and Man were not N-acetylated, lacking any large functional group likely to introduce steric hinderance in an elaborately branched chain. From the stand point of solvation free energy, Man serves as the more versatile of the three, providing a more positive average impact on solvation free energy than Glc, while being less dependent on location and conformation than Gal. This optimization of versatility in conjunction with solubility may, in part, explain why it is utilized so frequently in branching as well as why organisms with little diversity among monosaccharide "building blocks," like fungi, favor Man appendages.[51]

Among plants, there are no conserved O-glycan motifs with different organism using linear, branched, and monomeric glycans.[51] However, rather than the more exclusive Man favored by fungi, Gal is occasionally used for branching along with Fuc, GlcNAc, and other less common glycans.[65, 148] Offering they highest solvation free energy shift of the unacetylated glycans Gal makes a good candidate for branch points. Commonly employed monomerically or in very small chains in vertebrates, often for structural extracellular matrix secretions like collagen. [17]

While organisms exhibit preferences for specific O-glycans, most appear in all kingdoms of life to at least some degree, although in simpler life forms such as prokaryotes and archaea, glycosylation patterns while diverse in monomeric components, remain simple with very little elongation or branching.[51] In humans, Fuc, Gal, Glc, GlcNAc, and GalNAc are among the most common monosaccharides used to form glycans, but they are typically built into highly complex structures.[149] Glc, as previously discussed, attaches almost exclusively to serine and is most commonly elongated with linear Xyl additions in humans, where it is used to help fold epidermal growth factor (EGF)-like domains.[51, 68] From the standpoint of solubility, there are certainly more favorable combinations than Glc-Xyl-based glycans, but the known role in protein folding of a commonly occurring domain indicates its functions stretch beyond simple enhancing solubility. GlcNAc and Fuc, while not exclusive to serine attachment, are so able to form linkages with threonine residues as well serve a similar role to Glc. Both are attached to EGF-like domains; however, unlike Glc, both appear somewhat commonly as initial glycans in some plants.[51, 148] Common motifs initiated by both GalNAc and Fuc both are numerous and diverse and, thus, beyond generalizations.[150-152] While GalNAc is among the most advantageous additions examined, Fuc was among the least favorable. However, the calculations suggest the impact of Fuc on solvation free energy is relatively independent of linkage conformation, affording a certain degree of versatility similar to Man reduced preference between serine and threonine.

GlcNAc ranked among the most advantageous of glycan additions in terms of solvation free energy. Much like in N-linked glycans, O-linked GlcNAc can serve as an initial base monomer for construction longer glycan chains. Additionally, GlcNAc can be added monomerically to achieve an effect similar to phosphorylation, and at times, can even target phosphorylation sites on a protein.[51, 68] Like phosphorylation this has the potential to trigger conformational changes. It is possible that, with these small chain additions being so much more impactful to solvation free energy than other glycan options, GlcNAc potentially serves the dual purpose of both initiating conformational change as well as stabilizing the new conformation with its increased effect on protein solubility.

6.5. Conclusion

Calculation of the free energy of solvation for these fundamental glycan building blocks provided insight into the effects of glycosylation on protein stability, not only elucidating aspects of naturally occurring motifs, but also introducing tools for engineering new glycans for applications in pharmaceuticals and other pertinent fields. Calculations clearly illustrate the benefits of branching and chain elongation to solubility when assembling glycans, as well as confirm and quantify the value of β-linkages over the alternative α conformation. Exploration into the effect of residue on glycosidic linkages revealed N-glycan asparagine initiated chains to have a more favorable effect than O-glycans, followed by serine being more impactful than threonine in every glycan tested (N-

Asn > O-Ser > O-Thr). When considering the effect of individual saccharides used in glycan construction, the most heavily functionalized, N-acetylated GalNAc and GlcNAc, produced the greatest effect on solvation free energy. In the case of fungi-favored O-Man glycans, the less functionalized pyranose ring and more optimally positioned C2 carbon provides more opportunities for branching without steric clashes while still providing an appreciable increase in solvation free energy.

While our study provides a valuable starting point for understanding how protein glycosylation impact protein solubility and, likely, stability, the extreme diversity among glycans in terms of both monomeric composition and conformations merits further investigations into elongated O-glycans and further branched N-glycans in order to expand our understanding of naturally occurring glycan motifs. Our results provide a tool, moving forward, for assessing existing glycans and determining their impact on protein stability.

Chapter 7 – Conclusions and Future Work

This dissertation investigated the enzyme catalyst of the rate-limiting step of the 4S pathway, *DszB*. Characterization of the inhibitive effect of pathway product HBP, along with several other crude oil contaminants, was performed to increase understanding of poor overall enzyme performance and to identify a means for structure-guided activity improvements. Poor kinetic stability and solvent sensitivity of *DszB* led to the investigation of optimal solvation environments, which revealed buffer additives more effective at stabilizing *DszB*, but ultimately fell short of providing an industrially viable solution. The fundamentals of glycosylation, a common post-translation modification, were also investigated as a potential alternative method for improving protein stability, and a library of basic glycan motifs was examined to enhance understanding of more complex and effective glycans in nature.

7.1. Overview

Inhibition studies elucidated basic mechanisms that limit *DszB* activity. Circular dichroism demonstrated a large-scale conformational response to HBP consistent with that observed crystallographically in response to substrate binding, implying similarities between the two conformational responses. Florescent kinetic assays revealed inhibition that follows a competitive model, further supporting the product induced conformation change that mirrors substrate binding. From MD simulations, center of mass analyses and root mean square fluctuation showed mobility in both the protein fold and HBP position, while nonbonded interactions demonstrated only a slight affinity for several residues buried in the protein core. In conjunction with free energy perturbation calculations, this revealed HBP's affinity for the active site, and, thereby, its capacity to inhibit *DszB* was derived not only from specific interactions with active site residues, but also an entropic effect driving the more hydrophobic HBP to remain buried in the less solvent-exposed pocket away from the aqueous solvation environment.

Time-lapse activity studies demonstrated *DszB*'s lack of long-term stability. Exploration of a wide range of solvation environments was conducted via differential scanning fluorimetry to both improve solvation free energy as well as to measure the stabilizing impact of amphiphilic buffer elements that may reduce production inhibition. Results indicated that select buffer additives could yield a positive impact on *DszB* stability though such effects were largely limited to short-term stability on the time scale of hours. Long-term stability over several days still showed significant decline in activity and all amphiphilic detergent additives proved deleterious to *DszB* stability.

In an effort to explore alternative means of enhancing protein stability, a fundamental investigation into the impact of glycans on the solvation free energy of glycoproteins was conducted. Thermodynamic integration calculations of the solvation free energies of basic O-glycan building blocks and common N-glycan motifs quantified glycan potential for enhancing the solution stability of a decorated protein. The conformation of glycosidic bonds formed, the individual glycan units used, and the position along the glycan ring of each subsequent addition made were all found to impact the degree to which protein solvation free energy was changed.

7.2. Future work

In addition to *E. coli*, Chinese Hamster Ovarian (CHO) cells and yeast serve as popular cell lines for commercial production of proteins. [153] Yeast offers low diversity, highly predictable glycosylation patterns as an expression host, while CHO cells favor more human-like glycosylation motifs. While expression of bacterial proteins in mammalian cells is a less common practice and proposed as a secondary option to the preferred yeast expression host, select stains of CHO cells have some glycosylation mapping data available allowing for strain selection based partly on a desired glycoform.[51, 153-154] Future work will focus on these organisms in their capacity as expression hosts.

With the completion of the glycan library, the next logical steps for this research fall along two distinct paths: (1) further development of the glycan library and (2) application of the glycan library data to stabilize *DszB*. Expansion of the glycan library would take a "proof-of-concept" tool of limited use and expand it into a more comprehensive library of solvation free energies, potentially including additional metrics for evaluating increasingly complex glycans. Glycan motifs of more immediate interest to be added to the library include: (1) O-linked mannose elongated chains with branching glycans like observed in yeast and (2) specific extensions of the common $Man_3(GlcNAc)_2$ N-linked motif observed in CHO cells. Glycosylation motifs of these two organisms are of particular interest given their commercial significance. Many therapeutic molecules and industrially produced enzymes are manufactured through the use of CHO and yeast strains, both of which exhibit different glycosylation patterns (even from strain to strain), and, thus, naturally occurring glycans within these two species are among the most readily applied to commercially produced proteins. Beyond just the addition of new glycans to a library of solvation free energies, deeper probes into the conformation of each glycan would further the usefulness of the database. While free rotation about glycosidic bonds and the potential for branched structures create a considerably more dynamic biomolecule relative to a classical protein, producing many rotamers that a given glycan rotates through, it is most probable that some rotamers are considerably more stable than others. In this regard, glycans might be visualized more like an intrinsically disordered protein, sampling/interchanging between multiple different nearly equally preferred conformations. Applying replica-exchange molecular dynamics (REMD) to the $Man_3(GlcNAc)_2$ and other longer chain motifs would allow the isolation of the more frequent rotamers each glycan favors, thus providing a conformational ensemble of each glycan tested. Expanding the library to include such structural data would allow its usefulness to extend beyond stability measurements and perhaps touch on the role particular glycans may play in signaling and recognition.

Given *DszB* limitations regarding stability, further research is necessary in order for the 4S pathway to ever be considered for implementation on an industrial scale. Prior testing demonstrated *DszB* sensitivity to solvation environment regarding both prolonged kinetic stability and degree of product inhibition, and while further investigations into varied solvation environments may yield moderate gains in *DszB* properties, ultimately, direct modification of the enzyme is likely necessary to produce performance levels desirable for commercial use. Many of the residues responsible for complicating product inhibition are either buried in the active site (a notoriously difficult region to stably mutate) or are highly conserved and, thus, likely critical to catalysis. While these challenges do not preclude directed evolution as a means to garner desired effects, glycosylation offers a potentially more versatile solution, less likely to generate deleterious variants. Expression in an *E. coli* host prevents glycosylation given that the organism lacks relevant cellular machinery. Glycan simulation data of naturally occurring glycan motifs indicated O-linked mannose offers versatility that ultimately can lead to large increases in solvation free energy due to its prevalence for branching. While alternative glycans may offer larger swings in stability, locating or designing an expression host to obtain such a specific pattern would likely prove difficult and unnecessary. Multiple strains of yeast designed for commercial production of protein are already employed in lab scale purification. Given yeast’s tendency to heavily glycosylate with mannose, yeast-produced *DszB* would likely show a high increase in aqueous stability. For alternative more diverse glycosylation patterns, expression of *DszB* in CHO cells, which also have strains available for commercial protein production, could be explored. Furthermore, changes in glycans attached to *DszB* could be achieved through direct surface residue mutations. Residues surrounding serine, threonine, and asparagine targeted for glycosylation influence the specific glycotransferase involved in attachment, thus altering the glycan attached and allowing for relatively simple, usually benign surface/solvent exposed residue mutations to impact glycosylation patterns.

7.3. Summary

In summary, the 4S pathway presents a potentially viable biodesulfuration option for capable of cutting current sulfur emissions in the liquid transportation sector in half given the prevalence of DBTs in current fuels. However, the native performance of *DszB* precludes industrial application as of now. Investigations into the inhibitive nature of principle pathway product, HBP, demonstrated molecule hydrophobicity in an aqueous solvation environment is a large factor in *DszB*’s relatively slow turnover. Additionally, further investigations into the long-term kinetic stability revealed a rapid decrease in enzyme activity over a relatively short time range. Alterations in the solvation environment of *DszB* can have a profound effect on enzyme stability but, ultimately, fall short of developing significant long-term kinetic activity, and the introduction of amphiphilic elements to aqueous buffers in order to reduce product inhibition destabilized *DszB* immediately. Analysis of the basic principles of glycosylation as they pertain to protein stability indicate expression of *DszB* in an alternative glycosylation-prone host organism such as yeast or CHO cells may yield more stable *DszB* variants capable both of maintaining activity for longer periods and with increased tolerance of inhibition reducing amphiphile buffer additives.

Appendix

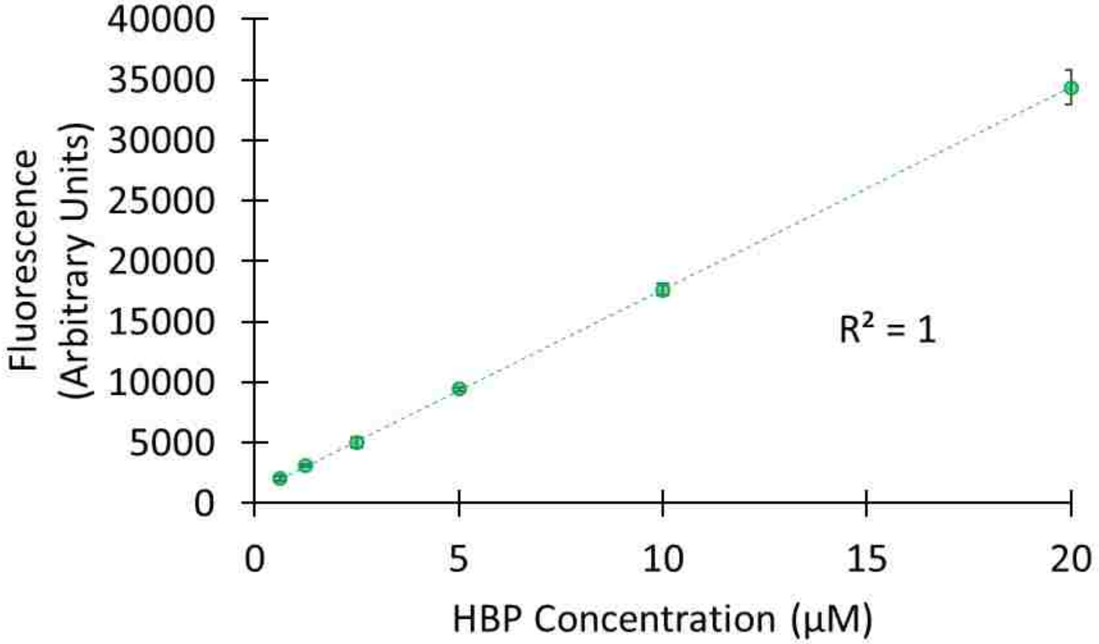

Figure 0.1. HBP concentration calibtration curve

Dotted line indicates calibration curve with resulting R^2 *value for linear regression displayed in the plot. Error bars represent standard deviation with n=3.*

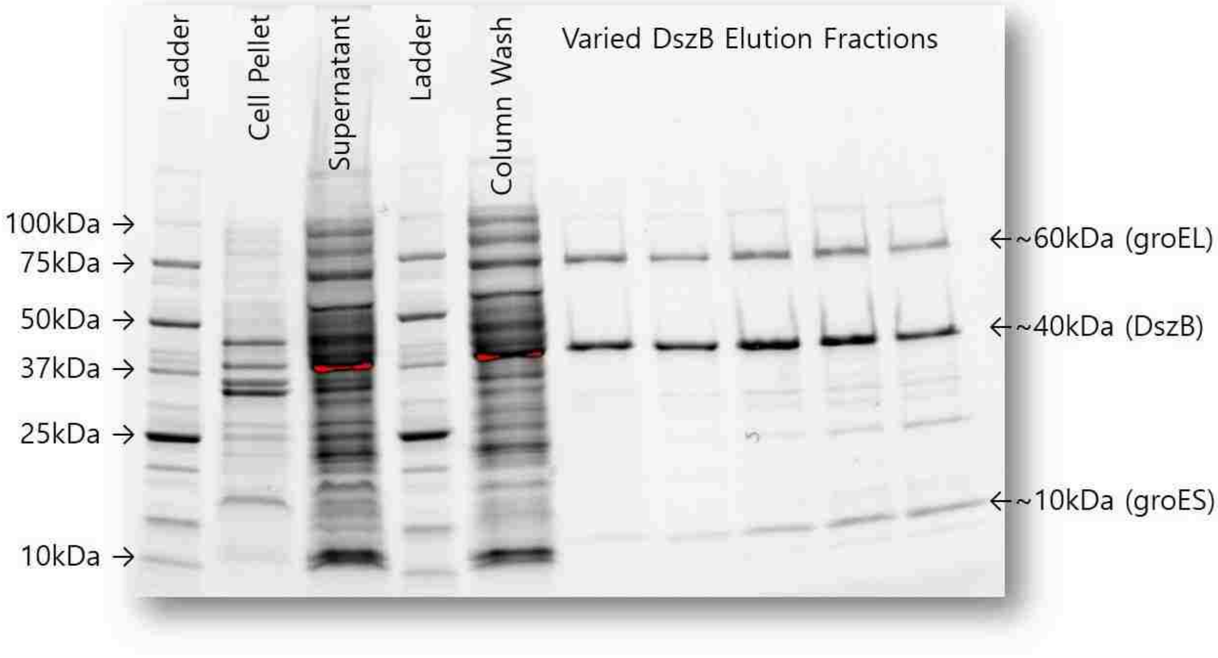

Figure 0.2. SDS PAGE for IMAC Purified DszB.

"Cell Pellet" and "Supernatant" samples were taken after cell lysis centrifugation. The "Column Wash" sample was taken from the cumulative column flow through prior to addition of elution buffer. "DszB Elution Fractions" were taken from 1 ml fractions throughout the elution processes. Elution fractions contain some impurities due to a close association with co-expressed molecular chaperones groES and groEL.

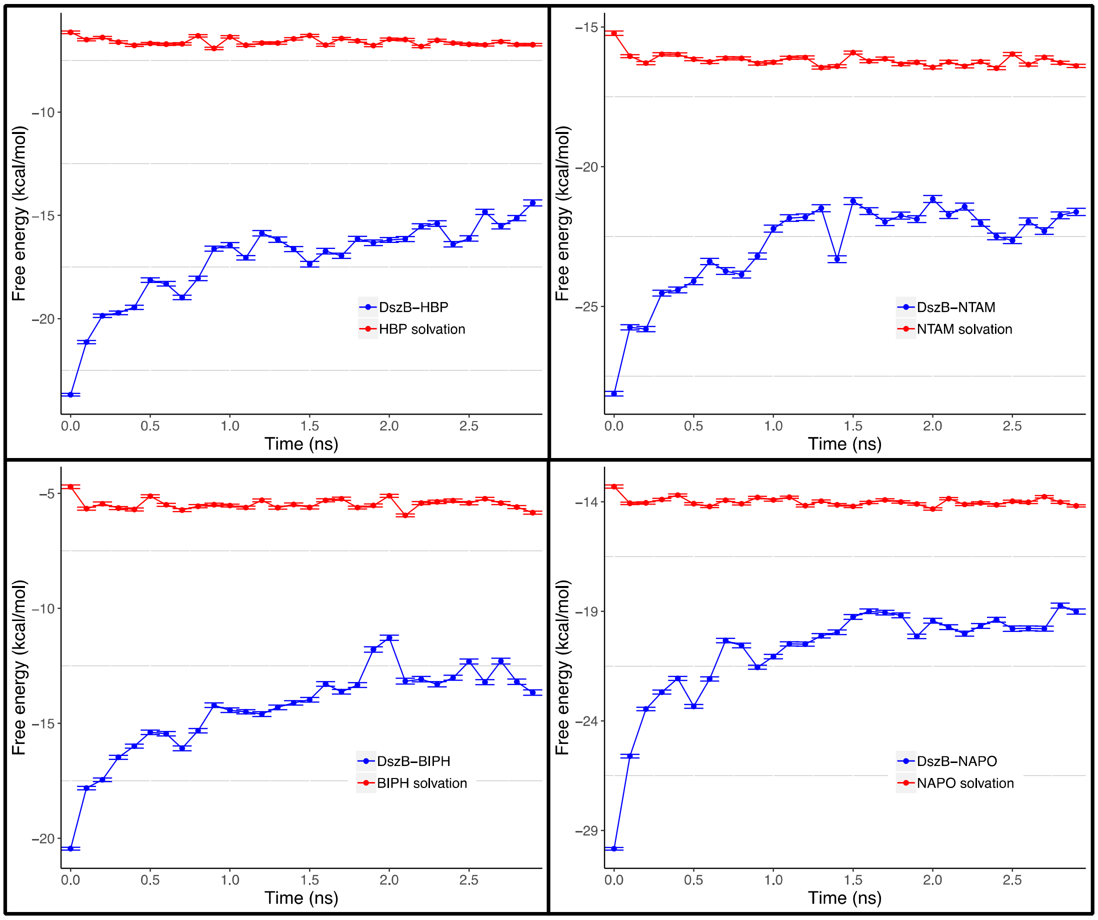

Figure 0.3. Gibbs free energy over 30 consecutive 0.1-ns simulations, as determined using FEP/λ-REMD for selected systems.

Blue lines represent the decoupling of the bound ligand from solvated DszB to a vacuum. Red lines represent ligand solvation free energy. Convergence was assessed based on time progression of the discrete free energy values. The last 1 ns data was used for all determinations of change in free energy.

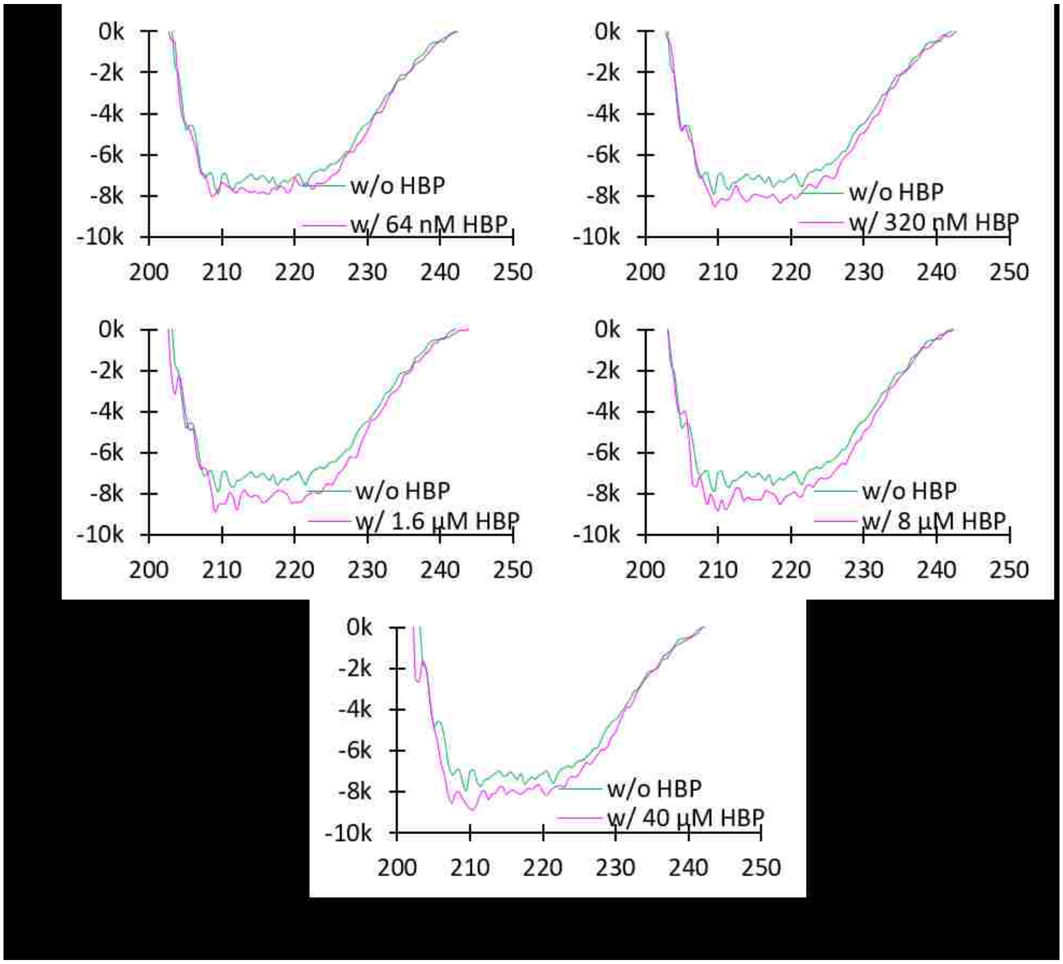

Figure 0.4. Circular dichroism spectra for DszB interacting with HBP

Circular dichroism spectra for DszB interacting with HBP (fuchsia) and unbound (green) for multiple concentrations of HBP (64 nM, 320 nM, 1.6 µM, 8 µM, and 40 µM) over a wavenumber range indicative of helical structure (200 to 250 nm).

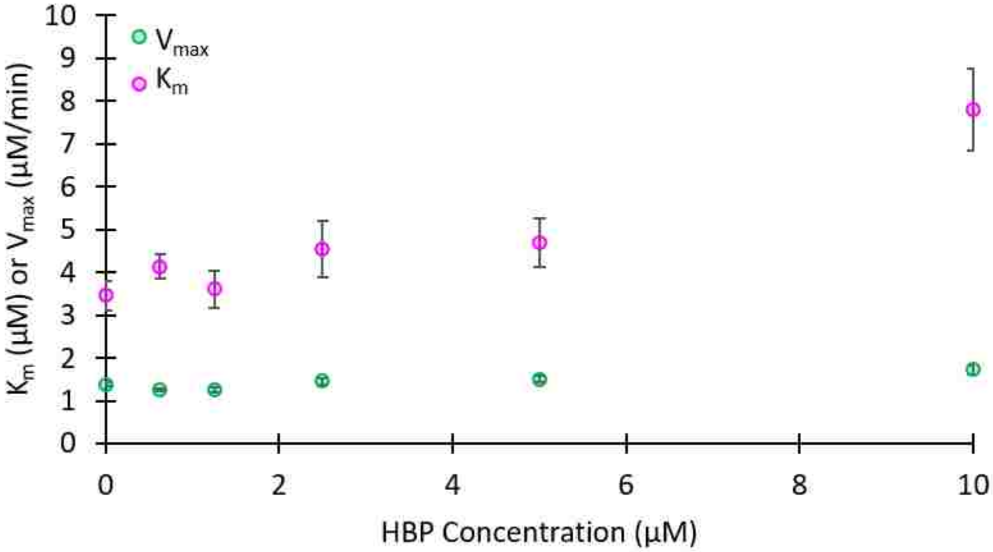

Figure 0.5. Product Inhibition of DszB

Demonstrated by apparent Michaelis-Menten parameters for each concentration of HBP. Each parameter was determined numerically using a Generalized Reduced Gradient (GRG) algorithm. Each parameter was then considered as a function of HBP concentration. The theoretical maximum rate of reaction (V_{max}) (green) appeared to be independent of HBP concentration, while the apparent Michaelis constant (K_m) (pink) increased with increasing concentrations of HBP. Error bars indicate standard error (n=3).

Table 0.1. Occupancy of each hydrogen bond formed between a ligand and DszB.

Inhibitors, including HBP, NTAM, and BIPH, are highlighted in blue, and non-inhibitory molecules, NAPO and BCA, are highlighted in gray. Values are obtained from single simulation CHARM outputs.

	HBPS			HBP		
Inhibitor	Donor	Acceptor	Occupancy (%)	Donor	Acceptor	Occupancy (%)
	R70-side	HBPS-O3	77.31	R70-side	HBP-O	1.05
	R70-side	HBPS-O2	55.31	HBP-O	H60-side	2.51
	G73-main	HBPS-O2	49.05	H60-side	HBP	1.01
	NTAM			BIPH		
	Donor	Acceptor	Occupancy (%)	Donor	Acceptor	Occupancy (%)
	G73-main	NTAM-O2	7.71	G73-main	BIPH-O1	4.55
	R70-side	NTAM-O1	1.05	G73-main	BIPH-O2	5.46
	G73-main	NTAM-O1	1.01	-	-	-
Non-inhibitory	NAPO			BCA		
	Donor	Acceptor	Occupancy (%)	Donor	Acceptor	Occupancy (%)
	-	-	-	R70-side	BCA-O2	1.57
	-	-	-	BCA-O1	H60-side	5.45

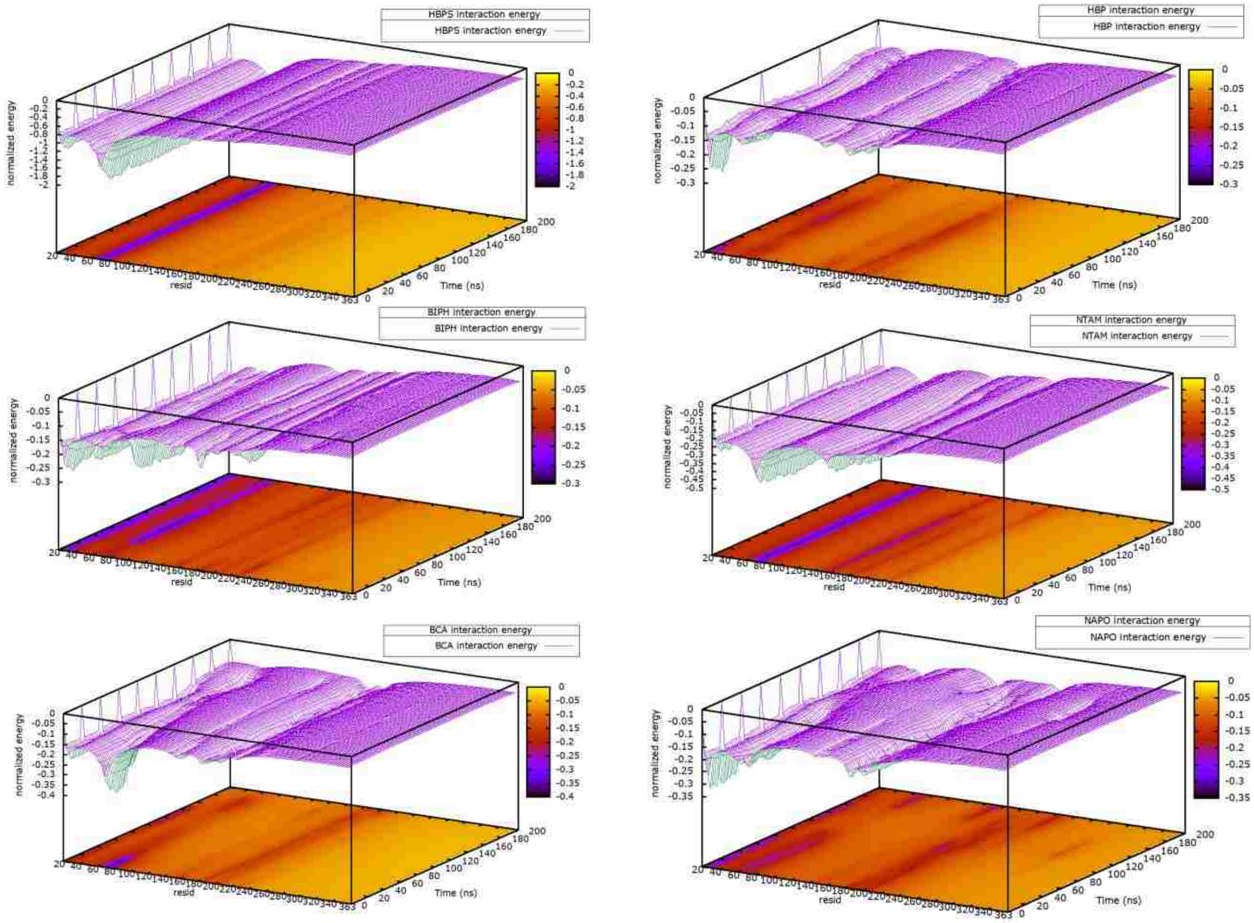

Figure 0.6. Total nonbonded interaction energy between a ligand and DszB with respect to time.

The energy value was calculated as a weighted average using a grid size of 100. The x-axis represents residue number, ranging from 20 to 363. The y-axis is the time series from 0 to 200 ns.

Table 0.2. Tabular values for resolvable protein melting points measured by DSF.

Solution additive column indicates the Hampton Solubility and Stability name and concentration for each additive that was combined in equal volumes with 0.2 mg/ml DszB in elution buffer. Group column indicates the general mechanism group the additive fall under. Groupings defined by Hampton's existing buffer additive classifications. In the case of multiple inflection points in fluorescence intensity vs temperature data, the recorded melting temperature listed indicates the lowest temperature inflection point.

Solution Additive	**Group**	**Melt °C**
Water Control	Water Control	45
500 mM Glycine	Amino Acid	44
500 mM β-Alanine	Amino Acid	45
2,500 mM Betaine monohydrate	Osmolyte	49
2,000 mM Xylitol	Osmolyte	52
2,000 mM Sucrose	Osmolyte	52
500 mM Spermine tetrahydrochloride	Polyamine	33
80 mM Adipic acid	Linker	32
500 mM Ethylenediamine dihydrochloride	Linker	21
500 mM Urea	Chaotrope	36
5% w/v Benzamidine hydrochloride	Inhibitor	42
1,000 mM Non-Detergent Sulfobetaine 201 (NDSB-201)	Non Detergent	51
500 mM Acetamide	Organic Acid	43
5% v/v Tacsimate pH 7.0	Organic Acid	40
250 mM 4-Aminobutyric acid (GABA)	Salt	31
500 mM Lithium nitrate	Salt	37
250 mM Sodium benzenesulfonate	Salt	43
50% v/v Glycerol	Polyol	47
10% v/v Ethylene glycol	Polyol	41
10% v/v Polyethylene glycol 200	Polyol	43
5% v/v Polyethylene glycol monomethyl ether 550	Polyol	42
10% v/v Polyethylene glycol 400	Polyol	43
10% w/v 1,2-Propanediol	Polyol	43
3% w/v Polyethylene glycol monomethyl ether 1,900	Polymer	37
3% w/v Polyethylene glycol 3,350	Polymer	34
50 mM Methyl-β-cyclodextrin	Cyclodextrin	55

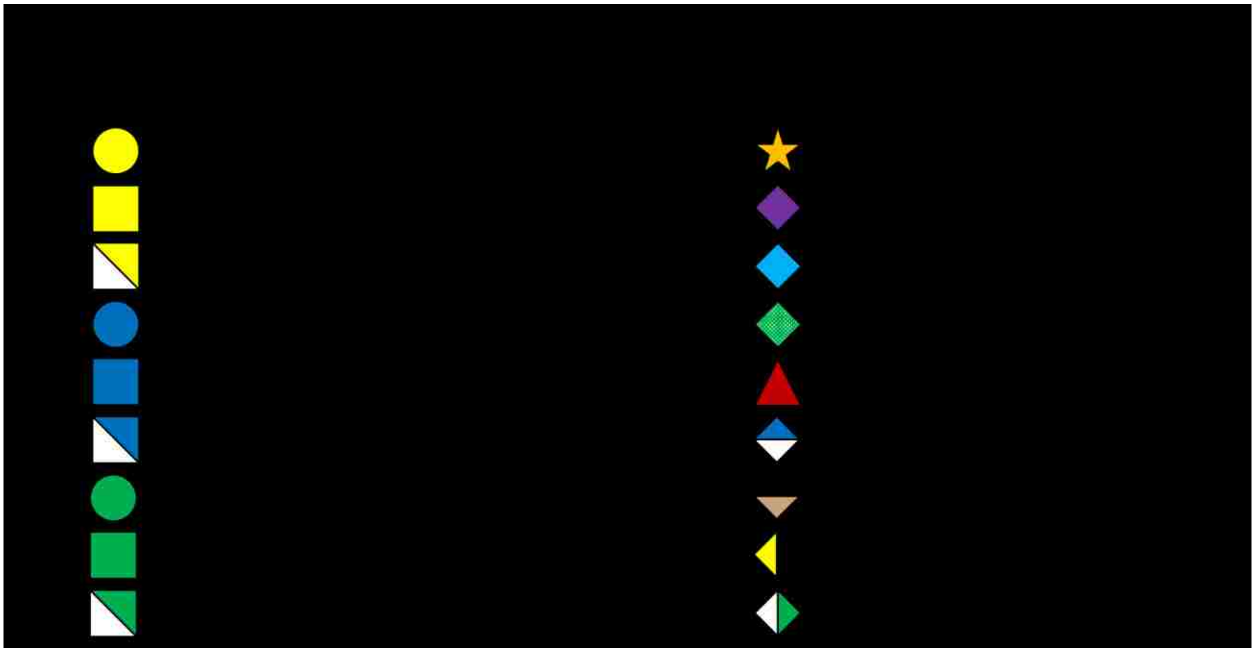

Figure 0.7. Key for symbolic representation of complex glycan structures.

Reproduce from Essentials of Glycobiology 3rd Edition. [51]

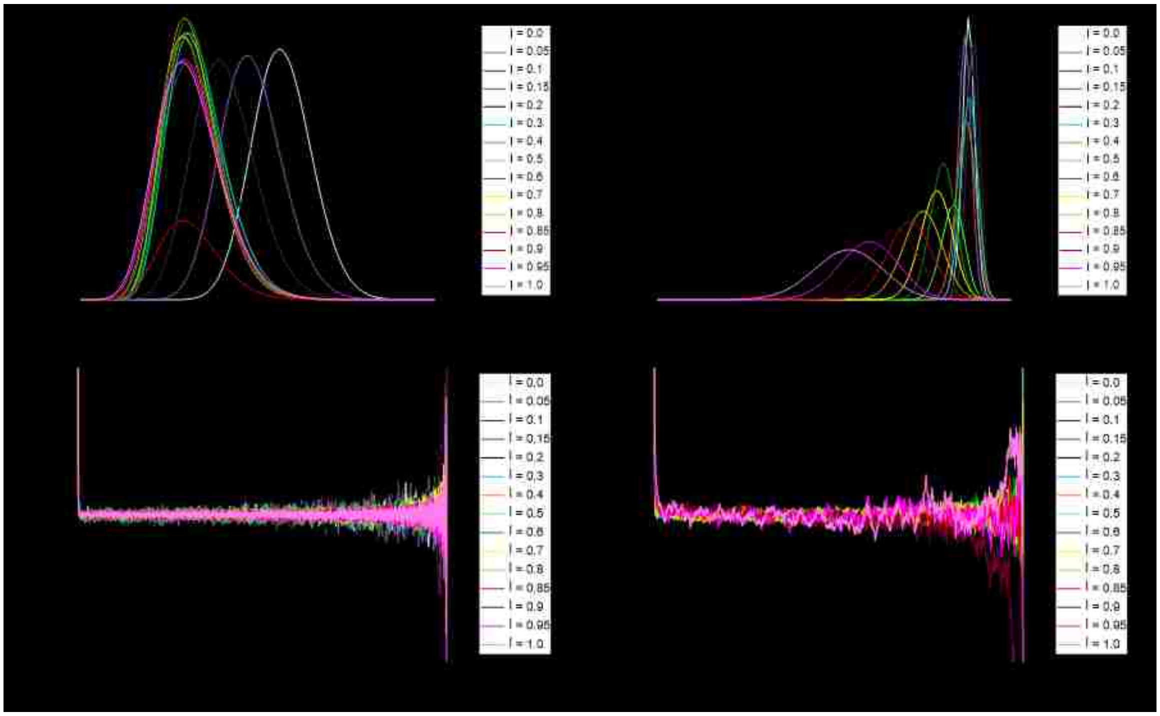

Figure 0.8. Histogram and autocorrelation function for Man3GlcNAc2

Left hand figures demonstrate convergence of van der Waals contributions to solvation free energy while right hand figures show electrostatic convergence. Bottom row figures depicted autocorrelation functions while top row figures depict histograms for each step along the alchemical path for our dual topology model. All graphs are color coded according to reaction coordinate.

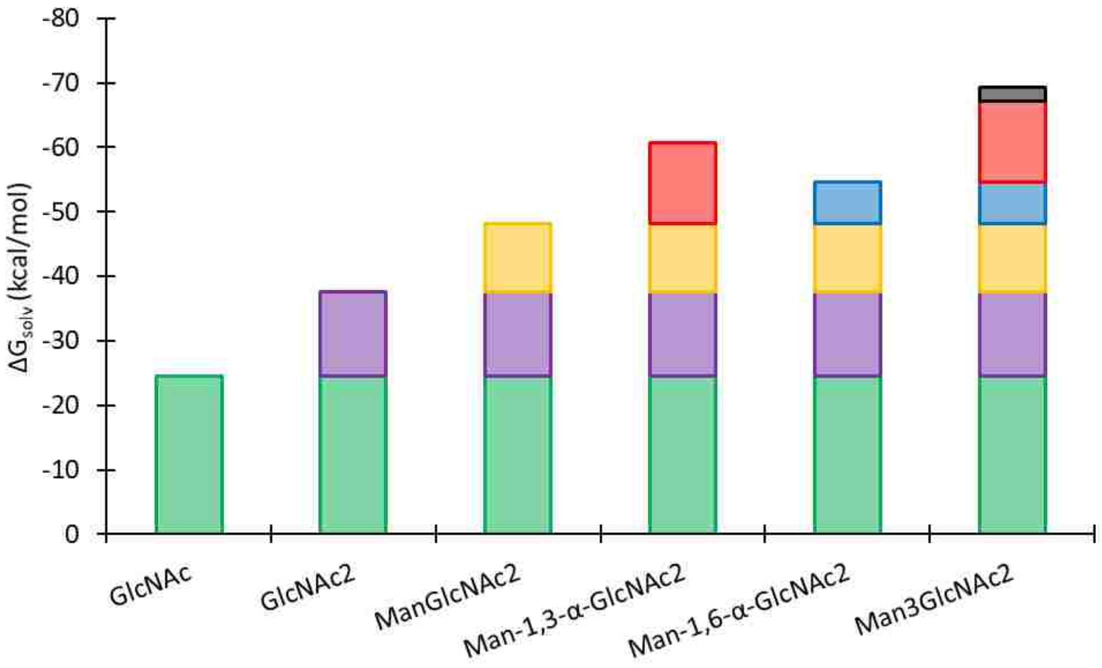

Figure 0.9. N-Glycan solvation free energy break down

Green: initiating GlcNAc unit

Purple: GlcNAc addition

Yellow: β-1,4 Man addition

Red: α-1,3 Man addition

Blue: α-1,6 Man addition

Black: Synergistic effect of branching from concurrent 1,3 & 1,6 Man additions

References

1. Caro, A.; Boltes, K.; Leton, P.; Garcia-Calvo, E., Description of by-product inhibiton effects on biodesulfurization of dibenzothiophene in biphasic media. *Biodegradation* **2008,** *19* (4), 599-611.
2. Perry, H., Coal in the United States: a status report. *Science* **1983,** *222*, 377-385.
3. Secretariat, U. N. E. C. f. E., *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)*. United Nations Publications: 2009.
4. Shah, P. S.; Balkhair, T., Air pollution and birth outcomes: A systematic review. *Environment International* **2011,** *37* (2), 498-516.
5. Kabe, T.; Ishihara, A.; Tajima, H., HYDRODESULFURIZATION OF SULFUR-CONTAINING POLYAROMATIC COMPOUNDS IN LIGHT OIL. *Ind. Eng. Chem. Res.* **1992,** *31* (6), 1577-1580.
6. Gray, M. R.; Ayasse, A. R., KINETICS OF HYDRODESULFURIZATION OF THIOPHENIC AND SULFIDE SULFUR IN ATHABASCA BITUMEN. *Energy & Fuels* **1995,** *9* (3), 500-506.
7. Javadli, R.; de Klerk, A., *Desulfurization of heavy oil*. 2012; Vol. 1.
8. Leprince, P., Petroleum Refining : Vol. 3: Conversion Processes. **2000**.
9. Javadli, R.; de Klerk, A., Desulfurization of Heavy Oil-Oxidative Desulfurization (ODS) As Potential Upgrading Pathway for Oil Sands Derived Bitumen. *Energy & Fuels* **2012,** *26* (1), 594-602.
10. Monticello, D. J.; Finnerty, W. R., MICROBIAL DESULFURIZATION OF FOSSIL-FUELS. *Annu. Rev. Microbiol.* **1985,** *39*, 371-389.
11. Carrales, M. J.; Martin, R. W., Sulfur content of crude oils. *US Department of the Interior, Bureau of Mine* **1975**.
12. Kilbane, J. J.; Le Borgne, S., Petroleum biorefining: the selective removal of sulfur, nitrogen, and metals. *Petroleum Biotechnology: Developments and Perspectives* **2004,** *151*, 29-65.
13. McFarland, B. L.; Boron, D. J.; Deever, W.; Meyer, J. A.; Johnson, A. R.; Atlas, R. M., Biocatalytic sulfur removal from fuels: Applicability for producing low sulfur gasoline. *Crit. Rev. Microbiol.* **1998,** *24* (2), 99-147.
14. Swaty, T. E., Global refining industry trends: the present and future. *Hydrocarb. Process.* **2005,** *84* (9), 35-+.
15. Linguist, L.; Pacheco, M., Enzyme-based diesel desulfurization process offers energy, CO2 advantages. *Oil Gas J.* **1999,** *97* (8), 45-+.
16. Li, G. Q.; Li, S. S.; Qu, S. W.; Liu, Q. K.; Ma, T.; Zhu, L.; Liang, F. L.; Liu, R. L., Improved biodesulfurization of hydrodesulfurized diesel oil using Rhodococcus erythropolis and Gordonia sp. *Biotechnology letters* **2008,** *30* (10), 1759-64.
17. Oldfield, C.; Pogrebinsky, O.; Simmonds, J.; Olson, E. S.; Kulpa, C. F., Elucidation of the metabolic pathway for dibenzothiophene desulphurization by Rhodococcus sp. strain IGTS8 (ATCC 53968). *Microbiology-(UK)* **1997,** *143*, 2961-2973.
18. Gallagher, J. R.; Olson, E. S.; Stanley, D. C., MICROBIAL DESULFURIZATION OF DIBENZOTHIOPHENE - A SULFUR-SPECIFIC PATHWAY. *FEMS Microbiol. Lett.* **1993,** *107* (1), 31-36.
19. McFarland, B. L., Biodesulfurization. *Current Opinion in Microbiology* **1999,** *2* (3), 257-264.

20. Agarwal, P.; Sharma, D. K., Comparative Studies on the Bio-desulfurization of Crude Oil with Other Desulfurization Techniques and Deep Desulfurization through Integrated Processes. *Energy & Fuels* **2010,** *24*, 518-524.
21. Kilbane, J. J., 2nd, Microbial biocatalyst developments to upgrade fossil fuels. *Current opinion in biotechnology* **2006,** *17* (3), 305-14.
22. Le Borgne, S.; Quintero, R., Biotechnological processes for the refining of petroleum. *Fuel Process. Technol.* **2003,** *81* (2), 155-169.
23. Nakayama, N.; Matsubara, T.; Ohshiro, T.; Moroto, Y.; Kawata, Y.; Koizumi, K.; Hirakawa, Y.; Suzuki, M.; Maruhashi, K.; Izumi, Y.; Kurane, R., A novel enzyme, 2 '-hydroxybiphenyl-2-sulfinate desulfinase (DszB), from a dibenzothiophene-desulfurizing bacterium Rhodococcus erythropolis KA2-5-1: gene overexpression and enzyme characterization. *BBA-Proteins Proteomics* **2002,** *1598* (1-2), 122-130.
24. Watkins, L. M.; Rodriguez, R.; Schneider, D.; Broderick, R.; Cruz, M.; Chambers, R.; Ruckman, E.; Cody, M.; Mrachko, G. T., Purification and characterization of the aromatic desulfinase, 2-(2′ -hydroxyphenyl)benzenesulfinate desulfinase. *Archives of Biochemistry and Biophysics* **2003,** *415* (1), 14-23.
25. Gray, K. A.; Pogrebinsky, O. S.; Mrachko, G. T.; Xi, L.; Monticello, D. J.; Squires, C. H., Molecular mechanisms of biocatalytic desulfurization of fossil fuels. *Nat. Biotechnol.* **1996,** *14* (13), 1705-1709.
26. Davoodi-Dehaghani, F.; Vosoughi, M.; Ziaee, A. A., Biodesulfurization of dibenzothiophene by a newly isolated Rhodococcus erythropolis strain. *Bioresource technology* **2010,** *101* (3), 1102-5.
27. Maghsoudi, S.; Vossoughi, M.; Kheirolomoom, A.; Tanaka, E.; Katoh, S., Biodesulfurization of hydrocarbons and diesel fuels by Rhodococcus sp strain P32C1. *Biochem. Eng. J.* **2001,** *8* (2), 151-156.
28. Reichmuth, D. S.; Hittle, J. L.; Blanch, H. W.; Keasling, J. D., Biodesulfurization of dibenzothiophene in Escherichia coli is enhanced by expression of a Vibrio harveyi oxidoreductase gene. *Biotechnol. Bioeng.* **2000,** *67* (1), 72-79.
29. Monticello, D. J., Biodesulfurization and the upgrading of petroleum distillates. *Current opinion in biotechnology* **2000,** *11* (6), 540-546.
30. Kodama, K.; Umehara, K.; Shimizu, K.; Nakatani, S.; Minoda, Y.; Yamada, K., IDENTIFICATION OF MICROBIAL PRODUCTS FROM DIBENZOTHIOPHENE AND ITS PROPOSED OXIDATION PATHWAY. *Agricultural and Biological Chemistry* **1973,** *37* (1), 45-50.
31. Kilbane, J. J., II, Sulfur-specific microbial metabolism of organic compounds. *Resources, Conservation and Recycling* **1990,** *3* (2-3), 69-79.
32. Kodama, K.; Nakatani, S.; Umehara, K.; Shimizu, K.; Minoda, Y.; Yamada, K., ISOLATION AND IDENTIFICATION OF PRODUCTS FROM DIBENZOTHIOPHENE. *Agricultural and Biological Chemistry* **1970,** *34* (9), 1320-&.
33. Raheb, J., *The Role of Microorganisms and Productions in Biodesulfurization of Fossiel Fuels*. 2016; Vol. 08.
34. *Bioprocessing and biotreatment of coal*. Marcel Dekker, Inc.: New York,, 1990; p xiii + 744.

35. Abin-Fuentes, A.; Leung, J. C.; Mohamed, M. E.; Wang, D. I. C.; Prather, K. L. J., Rate- Limiting Step Analysis of The Microbial Desulfurization of Dibenzothiophene in a Model Oil System. *Biotechnol. Bioeng.* **2014,** *111* (5), 876-884.
36. Ansari, F.; Grigoriev, P.; Libor, S.; Tothill, I. E.; Ramsden, J. J., DBT degradation enhancement by decorating Rhodococcus erythropolis IGST8 with magnetic Fe3O4 nanoparticles. *Biotechnol Bioeng* **2009,** *102* (5), 1505-12.
37. Ma, T., The Desulfurization Pathway in Rhodococcus. In *Biology of Rhodococcus*, Alvarez, H. M., Ed. Springer: New York, 2010; Vol. 16, pp 207-230.
38. Zhang, Q.; Tong, M. Y.; Li, Y. S.; Gao, H. J.; Fang, X. C., Extensive desulfurization of diesel by Rhodococcus erythropolis. *Biotechnology letters* **2007,** *29* (1), 123-7.
39. Gray, K. A.; Mrachko, G. T.; Squires, C. H., Biodesulfurization of fossil fuels. *Current opinion in microbiology* **2003,** *6* (3), 229-235.
40. Abin-Fuentes, A.; Mohamed, M. E.; Wang, D. I. C.; Prathera, K. L. J., Exploring the Mechanism of Biocatalyst Inhibition in Microbial Desulfurization. *Appl. Environ. Microbiol.* **2013,** *79* (24), 7807-7817.
41. Chen, H.; Zhang, W. J.; Cai, Y. B.; Zhang, Y.; Li, W., Elucidation of 2-hydroxybiphenyl effect on dibenzothiophene desulfurization by Microbacterium sp. strain ZD-M2. *Bioresource technology* **2008,** *99* (15), 6928-33.
42. Folsom, B. R.; Schieche, D. R.; DiGrazia, P. M.; Werner, J.; Palmer, S., Microbial desulfurization of alkylated dibenzothiophenes from a hydrodesulfurized middle distillate by Rhodococcus erythropolis I-19. *Appl. Environ. Microbiol.* **1999,** *65* (11), 4967-4972.
43. Lee, W. C.; Ohshiro, T.; Matsubara, T.; Izumi, Y.; Tanokura, M., Crystallization and preliminary X-ray analyses of desulfurization enzyme DszB and its C27S mutant complexed with biphenyl-2-sulfinic acid. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2004,** *60*, 1636-1638.
44. Lee, M. K.; Senius, J. D.; Grossman, M. J., SULFUR-SPECIFIC MICROBIAL DESULFURIZATION OF STERICALLY HINDERED ANALOGS OF DIBENZOTHIOPHENE. *Appl. Environ. Microbiol.* **1995,** *61* (12), 4362-4366.
45. Lee, W. C.; Ohshiro, T.; Matsubara, T.; Izumi, Y.; Tanokura, M., Crystal structure and desulfurization mechanism of 2'-hydroxybiphenyl-2-sulfinic acid desulfinase. *The Journal of biological chemistry* **2006,** *281* (43), 32534-9.
46. Geronimo, I.; Nigam, S. R.; Payne, C. M., Desulfination by 2'-hydroxybiphenyl-2-sulfinate desulfinase proceeds via electrophilic aromatic substitution by the cysteine-27 proton. *Chemical Science* **2017,** *8* (7), 5078-5086.
47. Chen, H. Y.; Demidkina, T. V.; Phillips, R. S., 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), 12276-12283.
48. Xi, L.; Squires, C. H.; Monticello, D. J.; Childs, J. D., A flavin reductase stimulates DszA and DszC proteins of Rhodococcus erythropolis IGTS8 in vitro. *Biochemical and biophysical research communications* **1997,** *230* (1), 73-5.
49. Li, M. Z.; Squires, C. H.; Monticello, D. J.; Childs, J. D., Genetic analysis of the dsz promoter and associated regulatory regions of Rhodococcus erythropolis IGTS8. *Journal of bacteriology* **1996,** *178* (22), 6409-18.

50. Watkins, L.; Rodriguez, R.; Schneider, D.; Broderick, R.; Cruz, M.; Chambers, R.; Ruckman, E.; Cody, M.; Mrachko, G., Purification and characterization of the aromatic desulfinase, 2-(2′ -hydroxyphenyl) benzenesulfinate desulfinase. *Archives of biochemistry and biophysics* **2003,** *415* (1), 14-23.
51. Ajit Varki, E. E., Richard D Cummings, Jeffrey D Esko, Pamela Stanley, Gerald W Hart, Markus Aebi, Alan G Darvill, Taroh Kinoshita, Nicolle H Packer, James H Prestegard, Ronald L Schnaar, and Peter H Seeberger, *Essentials of Glycobiology, 3rd edition*. Cold Spring Harbor: 2015-2017.
52. Dalziel, M.; Crispin, M.; Scanlan, C. N.; Zitzmann, N.; Dwek, R. A., Emerging principles for the therapeutic exploitation of glycosylation. *Science* **2014,** *343* (6166), 1235681.
53. Fiedler K, S. K., The role of N-glycans in the secretory pathway. *Cell* **1995,** *81*, 309-312.
54. Karpusas, M.; Nolte, M.; Benton, C. B.; Meier, W.; Lipscomb, W. N.; Goelz, S., The crystal structure of human interferon beta at 2.2-A resolution. *Proc Natl Acad Sci U S A* **1997,** *94* (22), 11813-8.
55. van Baarsen, L. G. M.; Vosslamber, S.; Tijssen, M.; Baggen, J. M. C.; van der Voort, L. F.; Killestein, J.; van der Pouw Kraan, T. C. T. M.; Polman, C. H.; Verweij, C. L., Pharmacogenomics of Interferon-ß Therapy in Multiple Sclerosis: Baseline IFN Signature Determines Pharmacological Differences between Patients. *PLOS ONE* **2008,** *3* (4), e1927.
56. Clerico, M.; Contessa, G.; Durelli, L., Interferon-beta1a for the treatment of multiple sclerosis. *Expert opinion on biological therapy* **2007,** *7* (4), 535-42.
57. Sareneva, T.; Pirhonen, J.; Cantell, K.; Julkunen, I., N-glycosylation of human interferon-gamma: glycans at Asn-25 are critical for protease resistance. *Biochem J* **1995,** *308 (Pt 1)* (Pt 1), 9-14.
58. Mastrangeli, R.; Rossi, M.; Mascia, M.; Palinsky, W.; Datola, A.; Terlizzese, M.; Bierau, H., In vitro biological characterization of IFN-beta-1a major glycoforms. *Glycobiology* **2015,** *25* (1), 21-9.
59. Ouellette, R. J.; Rawn, J. D., 13 - Carbohydrates. In *Principles of Organic Chemistry*, Ouellette, R. J.; Rawn, J. D., Eds. Elsevier: Boston, 2015; pp 343-370.
60. Varki, A., Biological roles of glycans. *Glycobiology* **2017,** *27* (1), 3-49.
61. Philippe Van den Steen, P. M. R., Raymond A. Dwek,Ghislain Opdenakker, Concepts and Principles of O-Linked Glycosylation. *Critical Reviews in Biochemistry and Molecular Biology* **1998**, 151–208.
62. Sola, R. J.; Griebenow, K., Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci* **2009,** *98* (4), 1223-45.
63. Lu, D.; Yang, C.; Liu, Z., How hydrophobicity and the glycosylation site of glycans affect protein folding and stability: a molecular dynamics simulation. *J Phys Chem B* **2012,** *116* (1), 390-400.
64. Schaffer, C.; Graninger, M.; Messner, P., Prokaryotic glycosylation. *Proteomics* **2001,** *1* (2), 248-61.
65. Joshi, H. J.; Narimatsu, Y.; Schjoldager, K. T.; Tytgat, H. L. P.; Aebi, M.; Clausen, H.; Halim, A., SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell* **2018,** *172* (3), 632-632 e2.

66. SabineStrahl-Bolsinger, M., WidmarTanner, Protein O-mannosylation. **1999,** *1426* (2), 297-307.
67. Loibl, M.; Strahl, S., Protein O-mannosylation: what we have learned from baker's yeast. *Biochim Biophys Acta* **2013,** *1833* (11), 2438-46.
68. Spiro, R. G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002,** *12* (4), 43R-53R.
69. Medzihradszky, K. F., Characterization of Protein N - Glycosylation. **2005,** *405*, 116-138.
70. Patrick Hossler, B. C. M., Wei-Shou Hu, Systems Analysis of N-Glycan Processing in Mammalian Cells. *PLoS ONE* **2007**.
71. P. Sears, C.-H. W., Enzyme action in glycoprotein synthesis. *Cellular and Molecular Life Sciences* **1998,** *54* (3), 223-252.
72. Hossler, P.; Khattak, S. F.; Li, Z. J., Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* **2009,** *19* (9), 936-49.
73. Harper, L. T.; Watkins, L. M. Expression, purification and characterization of recombinant 2-(2'-hydroxyphenyl)benzenesulfinase desulfinase from Rhodococcus erythropolis sp. IGTS8. Texas State University, 2012.
74. Verlet, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967,** *159* (1), 98-103.
75. Verlet, L., Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Physical Review* **1968,** *165* (1), 201-214.
76. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009,** *30* (10), 1545-614.
77. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** *26* (16), 1781-802.
78. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-38.
79. Gray, M. R.; Ayasse, A. R.; Chan, E. W.; Veljkovic, M., Kinetics of hydrodesulfurization of thiophenic and sulfide sulfur in Athabasca bitumen. *Energy & fuels* **1995,** *9* (3), 500-506.
80. Oldfield, C.; Pogrebinsky, O.; Simmonds, J.; Olson, E. S.; Kulpa, C. F., Elucidation of the metabolic pathway for dibenzothiophene desulphurization by Rhodococcus sp. strain IGTS8 (ATCC 53968). *Microbiology* **1997,** *143* (9), 2961-2973.
81. Gallagher, J. R.; Olson, E. S.; Stanley, D. C., Microbial desulfurization of dibenzothiophene: a sulfur-specific pathway. *FEMS Microbiol. Lett.* **1993,** *107* (1), 31-35.
82. Wang, Z., B. P. Hollebone, M. Fingas, B. Fieldhouse, L. Sigouin, M. Landriault, P. Smith, J. Noonan, G. Thouin, AND J W. Weaver, CHARACTERISTICS OF SPILLED OILS, FUELS, AND PETROLEUM PRODUCTS: 1. COMPOSITION

AND PROPERTIES OF SELECTED OILS. *U.S. Environmental Protection Agency* **2003**.

83. Yu, Y.; Fursule, I. A.; Mills, L. C.; Englert, D. L.; Berron, B. J.; Payne, C. M., CHARMM force field parameters for 2′-hydroxybiphenyl-2-sulfinate, 2-hydroxybiphenyl, and related analogs. *Journal of Molecular Graphics and Modelling* **2017,** *72*, 32-42.
84. Kelly, S. M.; Price, N. C., The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Curr. Protein Pept. Sci.* **2000,** *1* (4), 349-384.
85. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005,** *1751* (2), 119-139.
86. Lee, W. C.; Ohshiro, T.; Matsubara, T.; Izumi, Y.; Tanokura, M., Crystal structure and desulfurization mechanism of 2′-hydroxybiphenyl-2-sulfinic acid desulfinase. *Journal of Biological Chemistry* **2006,** *281* (43), 32534-32539.
87. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry* **2009,** *30* (16), 2785-2791.
88. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79* (2), 926-935.
89. Durell, S. R.; Brooks, B. R.; Bennaim, A., Solvent-induced forces between 2 hydrophilic groups. *Journal of Physical Chemistry* **1994,** *98* (8), 2198-2202.
90. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *Journal of chemical theory and computation* **2012,** *8* (9), 3257.
91. MacKerell, A. D.; Feig, M.; Brooks, C. L., 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), 1400-1415.
92. MacKerell Jr, A.; Bashford, D.; Bellott, M.; Dunbrack Jr, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., J. Wio rkiewicz-Kuczera, D. Yin and M. Karplus. *J. Phys. Chem. B* **1998,** *102*, 3586.
93. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I., 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), 671-690.
94. 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-sulfinate, 2-hydroxybiphenyl, and related analogs. *Journal of Molecular Graphics and Modelling* **2017,** *72*, 32-42.
95. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic acids research* **2005,** *33* (suppl 2), W368-W371.

96. T. Darden , D. Y., L. Pedersen, T. Darden, D. York, and L. Pedersen, J. Chem. Phys. 98, 10089 (1993). *J. Chem. Phys.* **1993,** *98,* 10089.
97. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.
98. Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., CHARMM: the biomolecular simulation program. *Journal of computational chemistry* **2009,** *30* (10), 1545-1614.
99. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics* **1984,** *81* (1), 511-519.
100. Hoover, W. G., Canonical dynamics: equilibrium phase-space distributions. *Physical review A* **1985,** *31* (3), 1695.
101. Jiang, W., Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics. *Journal of Chemical Theory and Comutation* **2009,** *5*, 2583-2588.
102. Deng, Y.; Roux, B., Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *Journal of Chemical Theory and Computation* **2006,** *2* (5), 1255-1273.
103. Hünenberger, P. H.; McCammon, J. A., Ewald artifacts in computer simulations of ionic solvation and ion–ion interaction: a continuum electrostatics study. *The Journal of chemical physics* **1999,** *110* (4), 1856-1872.
104. Kastenholz, M. A.; Hünenberger, P. H., 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), 774-788.
105. Kastenholz, M. A.; Hünenberger, P. H., Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *The Journal of chemical physics* **2006,** *124* (22), 224501.
106. Kastenholz, M. A.; Hünenberger, P. H., 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), 124106.
107. Reif, M. M.; Kräutler, V.; Kastenholz, M. A.; Daura, X.; Hünenberger, P. H., Molecular dynamics simulations of a reversibly folding β-heptapeptide in methanol: influence of the treatment of long-range electrostatic interactions. *The Journal of Physical Chemistry B* **2009,** *113* (10), 3112-3128.
108. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics* **2008,** *129* (12), 124105.
109. Schrödinger, L., PyMOL The PyMOL Molecular Graphics System. Version: 2010.
110. Secretariat, Globally Harmonized System of Classification and Labelling of Chemicals (GHS). *United Nations Publications* **2009**.
111. Mohebali, G.; Ball, A. S., Blocatalytic desulfurization (BDS) of petrodiesel fuels. *Microbiology-Sgm* **2008,** *154*, 2169-2183.
112. Ohshiro, T.; Ishii, Y.; Matsubara, T.; Ueda, K.; Izumi, Y.; Kino, K.; Kirimura, K., Dibenzothiophene desulfurizing enzymes from moderately thermophilic bacterium Bacillus subtilis WU-S2B: purification, characterization and overexpression. *Journal of bioscience and bioengineering* **2005,** *100* (3), 266-73.

113. Ohshiro, T.; Nakura, S.; Ishii, Y.; Kino, K.; Kirimura, K.; Izumi, Y., Novel Reactivity of Dibenzothiophene Monooxygenase from Bacillus subtilis WU-S2B. *Bioscience Biotechnology and Biochemistry* **2009,** *73* (9), 2128-2130.
114. Hamada, H. A., Tsutomu; Shiraki, Kentaro, Effect of Additives on Protein Aggregation. *Current Pharmaceutical Biotechnology* **2009,** *10* (4), 400-407.
115. Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P., Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **2006,** *357* (2), 289-98.
116. Hamada, H.; Arakawa, T.; Shiraki, K., Effect of Additives on Protein Aggregation. *Current Pharmaceutical Biotechnology* **2009**.
117. Block, H.; Maertens, B.; Spriestersbach, A.; Brinker, N.; Kubicek, J.; Fabis, R.; Labahn, J.; Schäfer, F., Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). **2009,** *463*, 439-473.
118. Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.; Fedorov, O. Y.; Finerty, P. J., Jr.; Wasney, G. A.; Yeung, R.; Arrowsmith, C.; Ball, L. J.; Berglund, H.; Hui, R.; Marsden, B. D.; Nordlund, P.; Sundstrom, M.; Weigelt, J.; Edwards, A. M., Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **2006,** *103* (43), 15835-40.
119. Auton, M.; Rösgen, J.; Sinev, M.; Holthauzen, L. M. F.; Bolen, D. W., Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains. *Biophys Chem* **2011,** *159* (1), 90-99.
120. Bolen DW, B. I., The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol* **2001**.
121. Back, J. F.; Oakenfull, D.; Smith, M. B., Increased thermal stability of proteins in the presence of sugars and polyols. *Biochemistry* **2002,** *18* (23), 5191-5196.
122. Kumar, R., Role of naturally occurring osmolytes in protein folding and stability. *Arch Biochem Biophys* **2009,** *491* (1-2), 1-6.
123. Dell, A.; Galadari, A.; Sastre, F.; Hitchen, P., Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol* **2010,** *2010*, 148178.
124. Weerapana, E.; Imperiali, B., Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **2006,** *16* (6), 91R-101R.
125. Larsen, I. S. B.; Narimatsu, Y.; Joshi, H. J.; Siukstaite, L.; Harrison, O. J.; Brasch, J.; Goodman, K. M.; Hansen, L.; Shapiro, L.; Honig, B.; Vakhrushev, S. Y.; Clausen, H.; Halim, A., Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc Natl Acad Sci U S A* **2017,** *114* (42), 11163-11168.
126. Jörg Rösgen, B. M., David Wayne Bolen, Protein Folding, Stability, and Solvation Structure in Osmolyte Solutions. **2005,** *89* (5), 2988-2997.
127. Harano, Y., Application of Hydration Thermodynamics to the Evaluation of Protein Structures and Protein-Ligand Binding. *Entropy* **2012,** *14* (8), 1443-1468.
128. Konig, G.; Bruckner, S.; Boresch, S., Absolute hydration free energies of blocked amino acids: implications for protein solvation and stability. *Biophys J* **2013,** *104* (2), 453-62.
129. Bagger, H. L.; Fuglsang, C. C.; Westh, P., Hydration of a glycoprotein: relative water affinity of peptide and glycan moieties. *Eur Biophys J* **2006,** *35* (4), 367-71.

130. Paluch, A. S.; Parameswaran, S.; Liu, S.; Kolavennu, A.; Mobley, D. L., Predicting the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol mixtures via molecular simulation. *J Chem Phys* **2015,** *142* (4), 044508.
131. Liu, S.; Cao, S.; Hoang, K.; Young, K. L.; Paluch, A. S.; Mobley, D. L., Using MD Simulations To Calculate How Solvents Modulate Solubility. *J Chem Theory Comput* **2016,** *12* (4), 1930-41.
132. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S., Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *The Journal of Chemical Physics* **2003,** *119* (11), 5740-5761.
133. Shirts, M. R.; Pande, V. S., Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J Chem Phys* **2005,** *122* (13), 134508.
134. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998,** *102* (18), 3586-616.
135. Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, 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. *J Comput Chem* **2004,** *25* (11), 1400-15.
136. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **1983,** *50* (5), 1055-1076.
137. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985,** *31* (3), 1695-1697.
138. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H., *Numerical-Integration of Cartesian Equations of Motion of a System with Constraints – Molecular-Dynamics of N-Alkanes*. 1977; Vol. 23, p 327-341.
139. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.
140. Lemkul, J. A.; Huang, J.; MacKerell, A. D., Jr., Induced Dipole-Dipole Interactions Influence the Unfolding Pathways of Wild-Type and Mutant Amyloid beta-Peptides. *J Phys Chem B* **2015,** *119* (51), 15574-82.
141. James C. Phillips, R. B., Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten, Scalable Molecular Dynamics with NAMD. *J Comput Chem* **2005**, 1781–1802.
142. L.BrooksIIIMartinKarplus, A., Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* **1984,** *105* (5), 495-500.
143. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J Chem Phys* **2007,** *126* (1), 014101.
144. Kubo, R., Toda, Morikazu, Hashitsume, Natsuki *Statistical Physics II*. Springer-Verlag: Berlin Heidelberg, 1991.

145. Steinbrecher, T.; Joung, I.; Case, D. A., Soft-core potentials in thermodynamic integration: comparing one- and two-step transformations. *J Comput Chem* **2011,** *32* (15), 3253-63.
146. Götting, C.; Kuhn, J.; Zahn, R.; Brinkmann, T.; Kleesiek, K., Molecular Cloning and Expression of Human UDP-d-Xylose:Proteoglycan Core Protein β-d-Xylosyltransferase and its First Isoform XT-II. *Journal of Molecular Biology* **2000,** *304* (4), 517-528.
147. Pomin, V. H.; Mulloy, B., Glycosaminoglycans and Proteoglycans. *Pharmaceuticals (Basel)* **2018,** *11* (1).
148. Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G., Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* **1998,** *33* (3), 151-208.
149. Peter ‐ Katalinić, J., Methods in Enzymology: O ‐ Glycosylation of Proteins. **2005,** *405*, 139-171.
150. Yu, H.; Takeuchi, H., Protein O-glucosylation: another essential role of glucose in biology. *Curr Opin Struct Biol* **2019,** *56*, 64-71.
151. Rana, N. A.; Haltiwanger, R. S., Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors. *Curr Opin Struct Biol* **2011,** *21* (5), 583-9.
152. Daniel J. Moloney, A. I. L. a. R. S. H., The O-Linked Fucose Glycosylation Pathway *The Journal of biological chemistry* **1997**.
153. Muir, E.; Raza, M.; Ellis, C.; Burnside, E.; Love, F.; Heller, S.; Elliot, M.; Daniell, E.; Dasgupta, D.; Alves, N.; Day, P.; Fawcett, J.; Keynes, R., Trafficking and processing of bacterial proteins by mammalian cells: Insights from chondroitinase ABC. *PloS one* **2017,** *12* (11), e0186759-e0186759.
154. Yang, Z.; Halim, A.; Narimatsu, Y.; Jitendra Joshi, H.; Steentoft, C.; Schjoldager, K. T.-B. G.; Alder Schulz, M.; Sealover, N. R.; Kayser, K. J.; Paul Bennett, E.; Levery, S. B.; Vakhrushev, S. Y.; Clausen, H., The GalNAc-type O-Glycoproteome of CHO cells characterized by the SimpleCell strategy. *Mol Cell Proteomics* **2014,** *13* (12), 3224-3235.

Vita

Education

2014-present	Ph.D., Chemical Engineering, University of Kentucky, United states of America
2009-2014	B.S., Chemical Engineering, University of Kentucky, United states of America

Publications

*Equally contributing authors

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-sulfinate, 2-hydroxybiphenyl, and related analogs. Journal of Molecular Graphics and Modelling 2017, 72, 32-42.

Yue Yu*, **Landon C. Mills***, Derek L. Englert, Brad J. Berron, Christina M. Payne, Inhibition mechanisms of Rhodococcus erythropolis 2'-hydroxybiphenyl-2-sulfinate desulfinase (DszB) (in preparation)

Landon C. Mills, Gregg T. Beckham, and Christina M. Payne The hydrophobicity and conformations of common glycosylation motifs across the kingdoms of life (in preparation)

Landon C. Mills, Derek L. Englert, and Christina M. Payne, The Impact of Solvation Environment on the Stability of Desulfurization Enzyme 2'-hydroxybiphenyl-2-sulfinate desulfinase (DszB) (in preparation)