The Pennsylvania State University

The Graduate School

Department of Chemistry

MOLECULAR DYNAMICS SIMULATIONS AND QUANTUM MECHANICAL/MOLECULAR MECHANICAL METHODS FOR PROTON AND HYDRIDE TRANSFER IN THE ENZYME DIHYDROOROTATE DEHYDROGENASE

A Thesis in

Chemistry

by

Yolanda A. Small

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2007

The thesis of Yolanda A. Small was reviewed and approved* by the following:

Sharon Hammes-Schiffer
Eberly Professor of Biotechnology and Professor of Chemistry
Thesis Advisor
Chair of Committee

A. Welford Castleman, Jr.
Evan Pugh Professor of Chemistry and Physics
Eberly Distinguished Chair in Science

Juliette T. J. Lecomte
Associate Professor of Chemistry

Kristen Fichthorn
Merrel R. Fenske Chair Professor
Professor of Chemical Engineering
Professor of Physics

Ayusman Sen
Professor of Chemistry
Head of the Department of Chemistry

*Signatures are on file in the Graduate School

ABSTRACT

Inhibiting the enzyme Dihydroorotate Dehydrogenase (DHOD) has the effect of interrupting the pyrimidine biosynthetic pathway responsible for the generation of precursors for DNA and RNA synthesis. Given this impact, therapeutic drugs for cancer and arthritis have been designed to target DHOD, but it is not clearly understood how the drugs function because details about the enzyme's mechanism still remain elusive. Experimental researchers have contributed structural and kinetic results in an attempt to elucidate the mechanism in DHOD. Their proposals for a proton and hydride transfer reaction form the basis for our computational contributions to the ongoing dialogue about catalysis in DHOD. In this study, three methods have been applied to the DHOD system. Quantum mechanical calculations were performed to determine transition states for proton and hydride transfer in DHOD. Molecular dynamics simulations were conducted to study proton relay pathways that facilitate proton transfer in the weakly basic active site environment. Finally, combined quantum mechanical/molecular mechanical methods were used to quantify the likelihood of a sequential or concerted mechanism for proton and hydride transfer.

In this thesis, Chapter 1 provides the framework and general motivation for studying enzymes with computational methods. Chapter 2 provides background on the DHOD system, including discussions about the experimental techniques used, relevant data obtained and how useful this data will be for comparison to computational methods. Chapter 3 discusses the computational methods applied to DHOD. Chapter 4 reports the results of molecular dynamics simulations that were done to investigate proton relay

pathways at the active site of DHOD. Chapter 5 reports the results of QM/MM calculations that were done to investigate the proton and hydride transfer mechanism. Chapter 6 concludes the study with perspectives on what has been gained by this study and what future directions can be taken.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

DHOD	Dihydroorotate Dehydrogenase
DHO	Dihdyroorotate
ORO	Orotic acid
FMN	Flavin Mononucleotide
FAD	Flavin Adenine Dinucleotide
FES	Iron-Sulfur cluster
NAD	Nicotinamide Adenine Dinucleotide
QM	Quantum mechanical
MM	Molecular mechanical
QM/MM	Quantum mechanical/molecular mechanical
DHFR	Dihydrofolate Reductase
MD	Molecular Dynamics
HF	Hartree-Fock
SCF	Self consistent field
DFT	Density Functional Theory
PES	Potential Energy Surface
TS	Transition state
KIE	Kinetic Isotope Effect
Cys	Cysteine residue
Ser	Serine residue
Thr	Threonine residue
Ala	Alanine residue
NMR	Nuclear Magnetic Resonance
MS	Mass Spectroscopy
SPC	Simple point charge water model
RNA	Ribonucleic Acid
DNA	Deoxyribonucleic Acid

ONIOM	Morokuma group N-layered Integrated molecular Orbital and molecular Mechanics
EFP	Effective Fragment Potential
EVB	Empirical Valence Bond
OPLS-AA	Optimized Potential for Liquid Simulations – All-atom forcefield
AMBER	Assisted Model Building with Energy Refinement forcefield
CHARMM	Chemistry at HARvard Molecular Mechanics forcefield
GROMACS	Groningen Machine for Chemical Simulations dynamics package
RMSD	Root Mean Square Displacement
RDF	Radial Distribution Function
B3LYP	Becke 3-Parameter (exchange) Lee, Yang, Parr (correlation) Functional

ACKNOWLEDGEMENTS

I thank my advisor, Professor Sharon Hammes-Schiffer for her support of my progress, for opportunities to grow as a researcher and for the technical resources to facilitate my studies. I also extend thanks to my committee members Professor A. Welford Castleman, Jr., Professor Kristen Fichthorn, Dr. Juliette Lecomte for their support not only during the program but also in helping to guide my next professional steps.

I would like to thank the entire Hammes-Schiffer group, past and present members, for numerous scientific discussions from which I learned a great deal. Particular thanks to Dr. Kim F. Wong, Dr. Alexander V. Soudackov, Dr. Michael Pak and Dr. Arindam Chakraborty for their patience in helping me to learn different aspects of this field. I gratefully acknowledge Dr. Victor Guallar and his group at Washington University in St. Louis and the Barcelona Supercomputing Center for their advice, proactive support on the DHOD project and vibrant discussions about enzyme catalysis.

I owe special thanks to those in the university community who were a constant source of encouragement throughout my experience here; Mr. Henry W. McCoullum, Dr. Cathy Lyons, Dr., Mrs. and Miss Fraser. I am especially grateful for the new friendships and stimulating conversations of Amber, Franklin, Nellie, Blaise, Soo Young, Qian, Sanchester, Becky, Kevin, Justin, Yasu, Hiro, Rachel, Dana, Garry, Mahy, Raul, Malika, Vindhya, Sherigo, Roxana and Corina. I am thankful for financial support from National Institutes of Health Grant GM56207 and the Alfred P. Sloan Foundation.

DEDICATION

This thesis is dedicated to my wonderful family whose love and undying support served as the foundation upon which I found the strength to navigate this course. I am deeply grateful for all of your guidance and input. Thank you for keeping me firmly grounded and completely reinforced when I needed it the most.

Chapter 1

Introduction

The efficiency in enzyme catalysis is a well recognized aspect of biochemistry that still remains a mystery in many enzymes. Significant progress has been made to pinpoint the source of catalysis in enzyme systems by utilizing the advances of current computational power. In recent years, it is widely accepted that studying enzyme catalysis can no longer be efficiently handled by using truncated representations of protein systems. However, the large number of degrees of freedom in a protein makes matters complicated in terms of distinguishing the contributions that lower the activation energy barrier in catalysis.

Exact methods of solving the time-dependent Schrödinger equation have been employed to study enzyme reactions. These *ab initio* quantum mechanical (QM) methods provided highly accurate results, but were limited to a small number of atoms, usually several orders of magnitude smaller than the number of atoms in an enzyme. Thus, these models represented an isolated description of the reaction in the gas phase and often could not be directly compared to the rate of the reaction in the enzyme. The benefit of calculations on small models was to ensure that the theory properly described well-defined molecular interactions that could then be extrapolated to describe interactions in the larger system. At the same time, molecular mechanical (MM) descriptions of large systems promise a more complete explanation of catalysis; however, the timescales of enzyme reactions require unreasonably lengthy computational time and

large volumes of data collection in order to describe the reaction accurately. To overcome the limitations of both the QM and MM methods, mixed quantum classical methods were introduced by Warshel and Levitt in 1976.[1] Today, several flavors of QM/MM methods[2, 3] have surfaced to offer an answer to the ultimate question: 'What causes enzyme catalysis?'

Using QM/MM methods, many researchers have proposed a variety of explanations to describe the driving force in enzyme catalysis. Some believe that electrostatic interaction is the key to catalysis, giving way to the solvent reorganization theories of Marcus.[4] Others attribute it to steric effects or van der Waals forces that support the near attack conformation where the substrate is placed in a favorable position to facilitate reactivity.[5-9] Other suggestions for catalysis include the influence of entropic effects and dynamical effects. Some suggest that the efficiency in enzyme catalysis is a manifestation of the protein destabilizing the reactant state thereby decreasing the reaction activation energy, while others propose that it is the stabilization of the transition state.[10] Yet others state that enzyme motion offers key contributions to catalysis.[11-13] While all of these effects could play a role in catalysis, they lack a measurable way to quantify how much they influence the reaction with respect to other enzyme effects. The common thread in all of these arguments is that the behavior of the protein influences the reaction at the active site but the exact method by which this happens is still elusive.

Many QM/MM methods are available, including ONIOM,[14] quantum mechanical/effective fragment potential (QM/EFP),[15, 16] empirical valence bond/molecular mechanics (EVB/MM),[12, 17-25] *ab initio* QM/MM,[26-37] semi-

emipirical/molecular mechanics (PM3/MM) and other semi-empirical/MM methods,[38-41] and path integral/MM (PI/MM).[42, 43] Each of these methods presupposes a characteristic trait of enzyme catalysis that is important and chooses to treat this characteristic most accurately.[44, 45] For example, the EVB/MM and PM3/MM approaches place a higher emphasis on sampling phase space over obtaining accurate energetics. As such the results of these methods can provide quantitative insight into the contribution of entropic and dynamical effects to catalysis very efficiently. The DFT/MM[46, 47] or GAMESS-UK/MM[48] methods offer a highly accurate treatment of the reactive region at the expense of entropic contributions to the reaction.

The redox reaction catalyzed by DHOD involves a reduction half-reaction, where dihydroorotate (DHO) is converted to orotate (ORO) via a proton and hydride transfer, and an oxidative half-reaction, where an electron is removed from the flavin mononucleotide (FMN) cofactor and returns it to the reactant state. The electron acceptor can be either ubiquinone, NAD or fumarate, depending on the family of DHOD. This study focuses on the reduction half-reaction of DHOD's redox process. The mechanism for the proton and hydride transfer is not clearly understood, although many experimental contributions have proposed a multitude of ideas. It is also not clear how the enzyme residues and solvent contribute to the reaction at the active site.

This study is an attempt to offer a computational perspective on the nature of DHOD's mechanism and how the protein and solvent environment assist in the mechanism. This is the first known study of DHOD using molecular dynamics and quantum mechanical calculations to address DHOD's mechanistic questions. A range of methods have been applied to DHOD, from high level quantum mechanics on a subset of

atoms at DHOD's active site, to mixed QM/MM methods on the enzyme and a subset of solvent molecules, and finally to molecular dynamics methods on the fully solvated 33000 atom model of DHOD. The range of methods enables us to probe proton and hydride transfer at the molecular level, while monitoring the influence of residues and solvent on these light atom transfers. The choice of methods will also highlight the importance of using multiple approaches to handle the complexities in enzyme catalysis.

Chapter 2

Background on Dihydroorotate Dehydrogenase

2.1 Introduction

Dihydroorotate dehydrogenase (DHOD) is involved in the rate limiting step of the pyrimidine biosynthetic pathway (figure 2-1). The key function is to catalyze the conversion of dihydroorotate (DHO) to the pyrimidine product orotate (ORO). The enzyme accomplishes this feat by abstracting the pro-*S* hydrogen from the C5 position of DHO, and a hydride is transferred to the flavin mononucleotide (FMN) cofactor that is non-covalently bound to the enzyme. How DHOD executes this conversion remains a puzzle despite many experimental attempts to elucidate the nature of DHOD catalysis.

A number of experiments have been performed on DHOD systems that offer varying perspectives on DHOD catalysis. Theoretical and computational studies are often inspired by the availability of existing experimental data. However, in many instances the experimental findings leave unresolved, a number of interesting questions that computational chemists take on as a challenge. Such is the case with DHOD. In this chapter we pay tribute to the existing experimental results obtained and highlight the areas where computational methods can contribute to the determination of the key features involved in catalysis. Here we review the experimental data available for DHOD: crystal structures, kinetic isotope effects (KIE), mutagenesis, steady-state kinetics, stopped-flow and single molecule experiments.

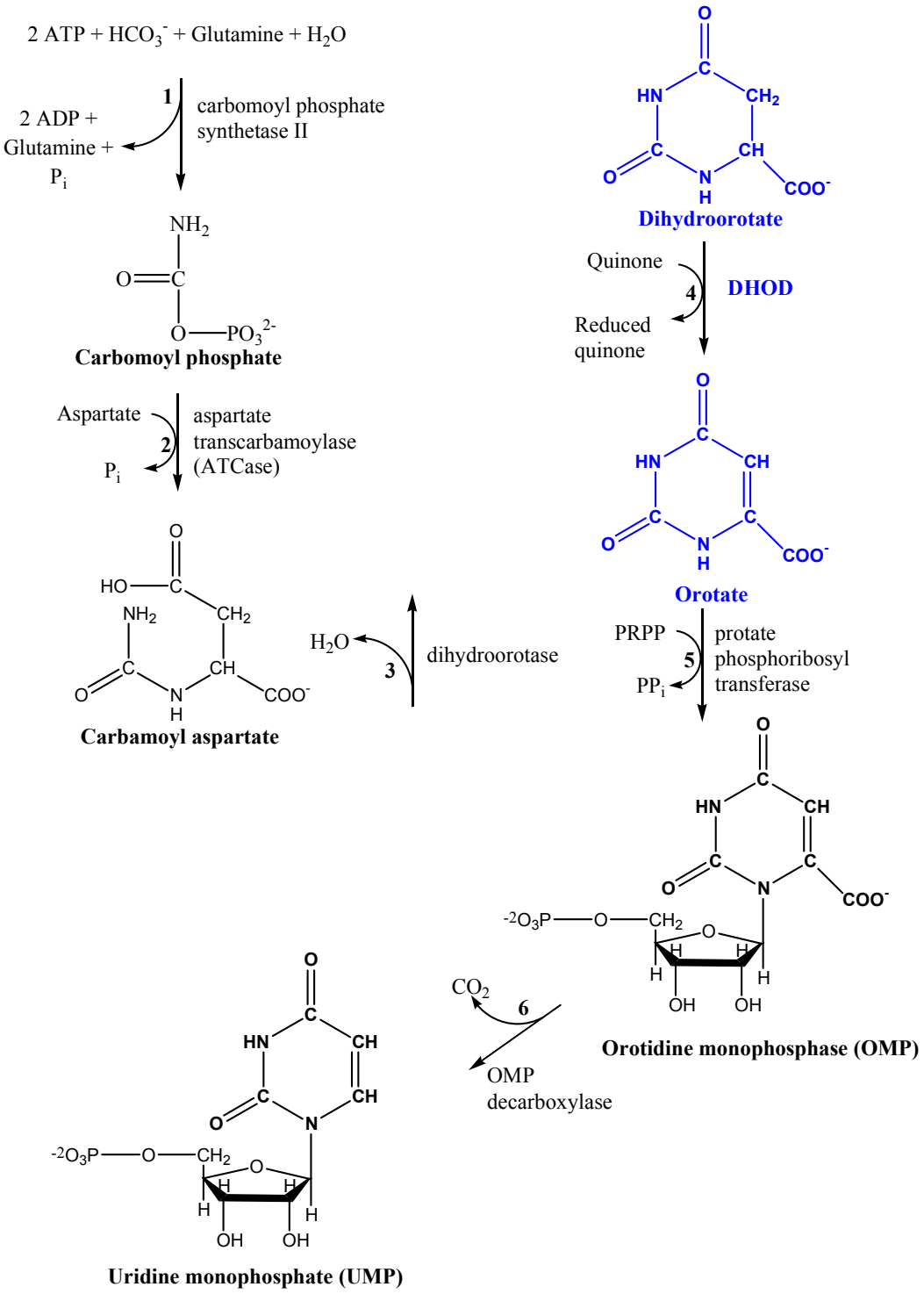

Figure 2-1: Six-step pyrimidine biosynthetic pathway. Step four is the rate-limiting redox reaction catalyzed by Dihydroorotate Dehydrogenase.

2.2 Crystal Structures

Several crystal structures have been solved for a number of DHOD species and are summarized in Table 2-1 .[49-55] The structural information and sequence alignment clearly established that three forms of the enzyme exist. These forms are divided into two main classes. Class I is found in gram-positive bacteria and parasitic life forms and resides in the cytosol of the cell. Class I is further sub-divided into forms IA and IB, where form IA is a homodimer and form IB is a heterotetramer containing an iron-sulfur cluster (FES) and flavin adenine dinucleotide (FAD) cofactor in addition to the FMN shared by all classes of DHOD enzymes.[56-59] Class II is a monomeric structure that is found in gram-negative bacteria and higher life forms such as humans. Figure 2-2 illustrates a general picture of the structures of classes IA, IB and II respectively. Class II, being the smallest of the DHOD systems, has been used in this study. In all of the class II structures, the substrate is bound in the enzyme in the product state (ORO), while the enzyme has been captured in the reactant state in the crystal. This mixture of states appears to be a common artifact of the experimental technique in which crystals are soaked in a solution of the substrate, a color change of the crystal indicates when the reaction has occurred, and the resulting crystal is used for the X-ray data collection. Crystal structures are obtained by sufficiently slowing down the reaction such that the crystals will form. As such, a majority of the class II structures have an inhibitor bound to the enzyme. The inhibitor can be molecules ranging in size from 19 to 46 heavy atoms, making the inhibitor molecule highly invasive to the reactive state of the enzyme. Subsequently, a key consideration from the computational perspective is to select a

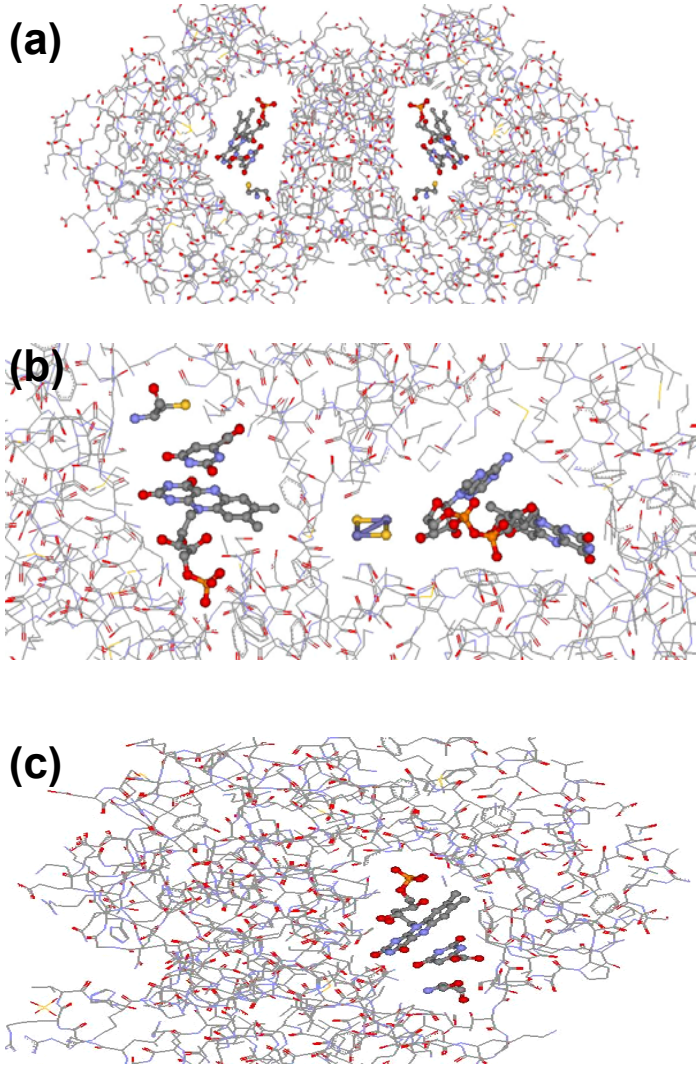

Figure 2-2: General picture of the structural features of three classes of DHOD. (a) The homodimeric structure of class IA (PDB code 1F76). (b) The heterotetrameric structure of class IB[60] with a bound Iron-Sulfur cluster and Flavin Adenine Dinucleotide (FAD) (PDB code 1EP2). (c) The monomeric structure of class II (PDB code 1D3G).

structure obtained under the least invasive conditions for the initial coordinates of the study. Crystallization conditions of a less invasive nature can either be mutations of a residue or binding of a substrate analogue.

Table 2-1: Summary of available crystal structures[49-53] for DHOD. Class 2 structures are in the first box, class IA[61] are in the middle box and class IB are in the third box. All structures are obtained from the Protein Data Bank.

PDB code	Species	ligand	Temp(K)	phValue	resolution	descriptor	chain ID	Year
1D3G	*Human*	DDQ, FMN, BRE, ACT, SO4	293	n/a	1.6	DHOD 2 COMPLEXED WITH BREQUINAR ANALOG	A	2000
1D3H	*Human*	ORO, FMN, A26, ACT, SO4	293	n/a	1.8	DHOD 2 COMPLEXED WITH ANTIPROLIFERATIVE AGENT *A771726*	A	2000
1UUM	*Rat*	AFI, ORO, FMN, BOG	n/a	8.5	2.3	DHOD 2 IN COMPLEX WITH ATOVAQUONE	A, B	2004
1UUO	*Rat*	ORO, FMN, BRF, Na, Ni	n/a	8.5	2.44	DHOD 2 IN COMPLEX WITH BREQUINAR	A	2004
2B0M	*Human*	ORO, FMN, 201	277	4.6	2	DHOD 2 bound to a novel inhibitor	A	n/a
1F76	*E Coli*	FMT, ORO, FMN	298	3.5	2.5	DHOD 2	A, B, D, E	2002
1DOR	*L Lactis*	FMN	n/a	8.5	2	DHOD A	A, B	1997
1JQV	*L Lactis*	ACY, ORO, FMN, MG	293	n/a	2.1	The K213E mutant of DHOD A	A, B	2003
1JQX	*L Lactis*	GOL, ORO, FMN, MG	293	n/a	1.7	The R57A mutant of DHOD A	A, B	2003
1JRB	*L Lactis*	ORO, FMN	293	n/a	1.9	The P56A mutant of DHOD A	A, B	2003
1JRC	*L Lactis*	ORO, FMN	293	n/a	1.8	The N67A mutant of DHOD A	A, B	2003
1JUB	*L Lactis*	GOL, FMN, MG	293	8.5	1.4	The K136E mutant of DHOD A	A, B	2003
1JUE	*L Lactis*	GOL, ACY, FMN, MG	293	8.5	1.8	native DHOD A	A, B	2003
1OVD	*L Lactis*	GOL, ORO, FMN, MG	293	8.5	2.25	THE K136E MUTANT OF DHOD A IN COMPLEX WITH OROTATE	A, B	2003
1TV5	*P. Falciparum*	N8E, ORO, FMN, A26, SO4	277	4.3	2.4	DHOD with a bound inhibitor	A	n/a
2B4G	*C3WBO3*	GOL, ORO, FMN, Br	298	4	1.95	DHOD	A, B, C, D	n/a
2BSL	*L Lactis*	GOL, DHB, FMN, Mq, ACT	n/a	8.5	2.3	DHOD A IN COMPLEX WITH 3,4-DIHYDROXYBENZOATE	A, B	n/a
2BX7	*L Lactis*	GOL, 34D, FMN, Mq, ACT	n/a	8.8	2.04	DHOD A IN COMPLEX WITH 3,5-DIHYDROXYBENZOATE DUAL BINDING MODE OF A NOVEL SERIES OF DHODH INHIBITORS	A, B	n/a
2BXV		3FT, ORO, FMN, ACT, SO4	n/a	4.6	2.15		A	2006
2DOR	*L Lactis*	ORO, FMN	n/a	8.5	2	DHOD A COMPLEXED WITH OROTATE	A, B	1998
1EP2	*L Lactis*	FES, FAD, ORO, FMN	298	4.6	2.4	DHOD B COMPLEXED WITH OROTATE	A, B	2000
1EP3	*L Lactis*	FES, FAD, FMN	298	4.5	2.1	DHOD B. COLLECTED UNDER CRYOGENIC CONDITIONS.	A, B	2000
1EP1	*L Lactis*	FES, FAD, FMN	298	4.6	2.2	DHOD B	A, B	2000

Based on the proton and hydride donor-acceptor distances seen in the active site of the crystal structures, it has been proposed that a cysteine (CYS) residue functions as the base in the reaction of class I DHOD enzymes while a serine (SER) functions as the base in class II DHOD. SER has a native pKa of ~13, and CYS has a pKa of ~8.3, implying that the CYS can give up its native proton more easily than the SER so that it is more capable of abstracting a proton from another atom (figure 2-3).

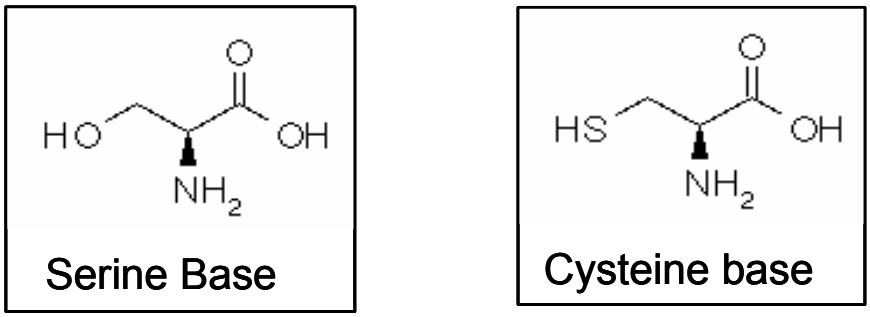

Figure 2-3: Amino acid residues thought to be involved in proton abstraction during Dihydroorotate Dehydrogenase catalysis.

This finding raises an interesting question as to how serine is able to assist in the reaction at the active site, if it is less likely to give up its native proton and exist as a negatively charged species, to abstract a proton. The crystal structures also point to a stacking configuration of the active site components. The six-membered ring of the substrate and the isoalloxazine ring of the cofactor are arranged in a pi overlapping configuration which may influence the level of quantum mechanical theory required to describe this system accurately because stronger dispersion interactions and electron correlation need to be considered.

2.3 Kinetic Isotope Effects

Kinetic isotope effect experiments are widely performed in enzyme studies. One key feature of KIE's is that the effect of hydrogen tunneling for light-atom transfer reactions can be quantified.[62-64] Deuterium isotopic substitution of the DHO substrate was performed and was characterized by NMR or mass spectroscopy. DHO was isotopically labeled at the C5 and C6 positions, independently to determine if each site was involved in the mechanism. Catalytic rates were then measured as a function of pH to determine that pH between 6.0 and 7.2 is the range with maximal activity. For both of the labeled substrates, similar trends were observed where KIEs are high (2 to 4) at low pH (6.5) and are near unity at high pH (8.8).[65] After establishing that both the C5 and C6 positions are active in the same pH range, double isotopic labeling, at both the C5 and C6 position, was done in an attempt to identify either a concerted or sequential transfer of the proton and hydride from the substrate. In comparison to the individually labeled isotopes, if substitution at the first position lowers the isotope effect at the second position in the doubly substituted isotope, then this implies that the two steps may have a separate catalytically relevant transition state and the reaction is thereby sequential. Otherwise, if there is no change in the isotope effect of the doubly substituted compound with respect to the individually substituted compounds in the catalytically relevant pH range, then the two steps may be concerted. A more rapid interpretation of the KIE data is to take the product of the isotope effects of the individually labeled sites and if this product is equal to the KIE of the doubly substituted species then the reaction is

concerted but if this product is larger than the doubly substituted KIE, then the mechanism may be stepwise.

Johnston and coworkers performed kinetic isotope effect studies on DHOD.[65] They were able to decipher that when the substrate dihydroorotate (DHO) binds in the active site, the chemistry evolves through the loss of the pro-*S* proton to a base in the protein environment and the loss of a hydride to a flavin mononucleotide (FMN) cofactor. The proposed mechanism is believed to involve a proton transfer from the C(5) carbon of the substrate to the hydroxyl end of the base residue.[66] Either in sequentially or concertedly, a hydride is transferred from the C(6) carbon of the substrate to the N(5) nitrogen of the cofactor. The products are orotate, formed from the substrate, and a reduced form of the cofactor. Orotate is then used in the next step of pyrimidine biosynthesis, and the cofactor gets re-oxidized by ubiquinone or nicotinamide adenine dinucleotide (NAD^+) depending on the form of the enzyme. Their kinetic isotope effect results are consistent with a concerted mechanism for the class II DHOD. However, the authors acknowledge the large error bars for the results obtained and do not completely eliminate the possibility of a stepwise mechanism.

2.4 Mutation Studies

The proton and hydride transfers in DHOD comprise the reduction half-reaction of a redox reaction. An oxidation half-reaction follows to remove the electron from the FMN. Mutation studies were carried out to determine the electron acceptor used by each class of DHOD.[54, 66] The results showed that in DHOD 1A fumarate is the electron

acceptor, in DHOD 1B Nicotinamide Adenine Dinucleotide (NAD) is the electron acceptor, and in DHOD 2 ubiquinone is the electron acceptor. Site directed mutagenesis was performed to confirm the identity of the reacting base residue. In the class I enzyme, the nearest base, cysteine, was mutated to the predicted class II base, serine, resulting in a nearly undetectable activity level at physiological pH.[67] Likewise in the class II enzyme, a similar Ser to Ala mutation proved to reduce enzyme activity significantly (100000-fold in *E. Coli*). Moreover, a Ser to Cys mutation also reduced enzyme activity but not as significantly as the Ala mutation (500-fold in *E. coli*).[68] This may indicate that the Ser is not absolutely critical for catalysis but may play a role in maintaining certain active site hydrogen bonded configurations necessary for proton transfer. Additionally, Walsh and coworkers performed site-directed mutagenesis studies on DHOD.[66] They proposed that the base residue differed between Class I and Class II of DHOD. Class I uses cysteine as the active site base, while Class II uses serine as the base. Additionally, mutation studies were performed to quantify the importance of lysine which hydrogen bonds to FMN. They concluded that the role of lysine is for FMN binding and stabilization.[69]

These findings from KIE and mutation experimental studies pose three interesting questions that we want to target with theoretical techniques: (1) When the two charged species are being transferred from the substrate, are they transferred concertedly or sequentially? (2) If the process is sequential, which transfer is rate limiting and what is the order of the transfer? (3) What is the energetic or environmental significance of the different base residues used in the class I and class II enzymes?

2.5 Kinetics Experiments: Steady State, Stopped-flow, Single Molecule

In steady state kinetics experiments, turnover rate can be captured by measuring the rate at which the substrate is consumed or the product is produced. However, the formulation of the final catalytic rate (k_{cat}) is based on an algebraic function of the rate constants from individual steps in the catalytic cycle. In the case of DHOD, it is likely that the k_{cat} generated by steady-state experiments will describe both the reduction and oxidation half-reactions, making it difficult to compare to our computational studies of just the reduction half-reaction. However, if comparable, the steady state experiments report a turnover rate that can be converted to an activation free energy ($\Delta G^{\neq}$) using the following transition state theory rate expression Eq. 2.1

$$k = \left(\frac{k_B T}{h}\right)\exp\left(-\frac{\Delta G^{\neq}}{RT}\right) \qquad \text{[Eq. 2.1]}$$

where k_B is the Boltzmann constant, T is the temperature, h is Planck's constant, R is the gas constant and *k* is the experimentally determined catalytic rate.

Stopped-flow experiments offer a more controlled solution to studying reactions with multiple steps and are highly useful in measuring the reduction and the oxidation half-reactions separately in DHOD. Flavin containing enzymes are, in general, good candidates for this method because of the intrinsic fluorescence of the isoalloxazine FMN group native to this enzyme.[70] FMN absorbs or fluoresces at wavelengths that are signatures of the oxidized or reduced states. In DHOD, the flavin absorbs at 450 nm and the enzyme substrate complex absorbs at 550 nm, making it possible to identify the stage

of the reaction. Additionally, fluorescence is used to verify the absorbance data. When the DHOD flavin is excited at 450 nm, the species will emit, exhibiting a signature peak at 520 nm. However, one limitation of this method is that side reactions in the enzyme are indistinguishable from the main reaction and may affect the measured overall rate. Stopped flow experiments were performed at pH 8.5 and 277 K. The results are represented in Figure 2-4 (written communication with B. Palfey). The turnover number is 16 s^{-1} at the same temperature and pH. Based on this value of k_{cat} = 16 s^{-1} at T = 277 K, the activation free energy is estimated to be 14.6 kcal/mol by the transition state theory rate expression above. However, it is not clear that this number isolates the proton or hydride transfer step, so it is not a viable rate for comparison to computationally calculated reaction energies.

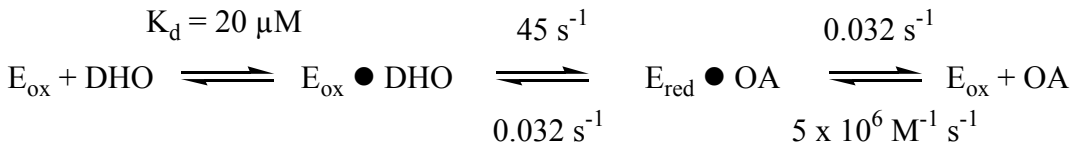

Figure 2-4: Experimental rate constants from stopped flow experiments on *E. coli* DHOD.[71]

Single molecule experiments are similar to stopped-flow in that half reactions can be isolated from the overall catalytic rate.[72-76] Again, the fluorescent nature of FMN makes DHOD a good candidate for these experiments.[77-80] For DHOD, FMN will fluoresce at 520 nm in the oxidized state and does not emit a signal in the reduced state. By measuring a sufficient number of events for proper statistics, lifetimes of fluorescence versus non-fluorescence can be calculated to provide a rate for the reaction. However, a

number of factors contribute to the uncertainty in rates measured by single molecule techniques. The method may suffer from poor statistics of the number of turnovers observed, insufficient signal-to-noise ratios, and improper binning of the data which may bias the interpretation of the results and identify arbitrary kinetic steps. Rates reported for DHOD from single molecule experiments are somewhat unclear.

2.6 Conclusions

When studying enzyme catalysis with computational methods, individual steps of the catalytic cycle can be modeled. A measure of how accurately the model reflects the behavior of the system is to compare to experimental data such as reaction rates. However, there are some cases where the inherent limitations of the experiment prevent a direct comparison to calculated results. In steady state kinetic experiments, individual steps along the catalytic cycle are not always distinguishable. In stopped flow experiments small, side reactions occurring during individual steps of the catalytic cycle are also indistinguishable. Finally, in single molecule experiments the results obtained may be biased if an insufficient number of observable turnover events is used for the analysis. With the wealth of experimental data available for DHOD, there is still no well-defined mechanism established. Furthermore, from the computational chemists' perspective, deciphering the experimental findings is a circumstance that fosters the much-needed collaborations with experimental groups to tackle these issues of enzyme catalysis.

This overview of the existing experimental data highlighted the areas in which computational and experimental results can be directly compared and aspects of catalysis in which general trends can be inferred. This encourages the on-going discussion on DHOD catalysis and engages computational chemists to devise approaches that may shed light on the matters of proton and hydride transfer. Issues of interest include sequential versus concerted, and the effect of mutations in the active site.

Chapter 3

Quantum Mechanical/Molecular Mechanical Method and Protocol for DHOD Calculations

3.1 QM and MM in *ab initio* QM/MM Methods

QM/MM methods allow for the treatment of large protein systems, where the active site can be treated with a high level of quantum mechanics (QM) while the remaining protein environment is treated with molecular mechanics (MM). Including both QM and MM components into the same calculation offers a more accurate representation of the reactive system. This approach ensures that the quantum mechanical active site is influenced by the interactions with the surrounding protein residues. The approach combines the energetic contributions from the QM region (E_{QM}), MM region (E_{MM}) and interactions between both regions ($E_{QM/MM}$).

$$E_{TOTAL} = E_{QM} + E_{MM} + E_{QM/MM} \qquad \text{[Eq. 3.1]}$$

The QM energy is calculated at the Hartree-Fock (HF) level of theory, where self-consistent field (SCF) calculations are performed to find the ground state wavefunction and energy. In the HF approximation, equation 3.2 describes the energy of the QM region where * indicates that that electrostatic interactions with the MM point charges are included in the one-electron (core) Hamiltonian.

$$E^{*}_{QM} = \sum_{\mu\nu} P_{\mu\nu} H^{core*}_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}[2J_{\mu\nu} - K_{\mu\nu}] + \sum_{AC} \frac{Z_A Z_C}{R_{AC}} + \sum_{AC} \frac{Z_A q_M}{r_{AM}} \qquad \text{[Eq. 3.2]}$$

Here, P is the electron density matrix of the QM region, H^{core} is the core Hamiltonian, J and K are the coulomb and exchange integrals respectively, the third term describes the electrostatic interaction between the QM nuclei and the fourth terms describes the electrostatic interaction between the QM nuclei and the MM point charges. The HF formalism is a general starting point for QM calculations. However, the above expression can be extended to improve the energy by incorporating the exchange-correlation features of density functional theory (DFT). For calculations in this thesis, the B3LYP functional is used.

The energy of the MM region is calculated according to equation 3.3

$$\begin{aligned} E_{MM} = & \sum_{i=stretches} k_i (r_i - r_0)^2 + \sum_{j=bends} k_j (\theta_j - \theta_0)^2 \\ & + \sum_{k=torsions} \left[\frac{V_{k,1}}{2}(1+\cos\phi_k) + \frac{V_{k,2}}{2}(1-\cos 2\phi_k) + \frac{V_{k,3}}{2}(1+\cos 3\phi_k) \right] \\ & + \sum_{MN} \frac{q_M q_N}{r_{MN}} + \sum_{MN} 4\varepsilon_{MN} \left[\left(\frac{\sigma_{MN}}{r_{MN}} \right)^{12} - \left(\frac{\sigma_{MN}}{r_{MN}} \right)^{6} \right] \end{aligned} \qquad \text{[Eq. 3.3]}$$

where the first three terms describe the bonding interactions (bond stretching, angle bending and torsional twisting) and the last two terms describe the non-bonding interactions (Coulomb and van der Waals). This expression is specific to the forcefield which, for the calculations in this thesis, is the OPLS-AA forcefield.[81] Some terms may differ in other forcefields such as AMBER, CHARMM, or GROMACS.[81, 82]

3.2 QM/MM Treatment of the Interface

The defining feature of QM/MM methods is how the interface between the QM and MM regions is treated. If there are no covalent bonds across the surface dividing the regions, the interactions included in the total energy are the electrostatic and non-bonded interactions between the QM and MM atoms. The electrostatic energy is calculated as a Coulomb interaction between the electronic density in the QM region and the MM point charges. The van der Waals interaction terms for the interface are parameterized and are usually larger than those of the standard forcefield to account for the stronger Pauli repulsion between the atoms at the interface for certain atom types, for example, N-H---O, and N-H---S hydrogen bonds.

A more complex interface treatment is required when a covalent bond separates the two regions. For *ab initio* QM/MM methods, common treatments of the covalent interface involve either a link-atom approach[31, 47, 83] or a frozen orbital-like approach.[84] The link atom approach removes the covalency by cutting the interface bond and capping the ends with hydrogen atoms. In this way, the QM/MM calculation simplifies to the non-covalent interface case. However, in some rare cases this approach has been shown to produce incorrect final energies that are thought to be artifacts of the force representation from truncating the system. To more accurately describe a covalent interface, Gao and coworkers were the first to introduce frozen orbital-like methods.[84] The calculations in this thesis used a frozen orbital scheme implemented by Friesner and coworkers.[85] The frozen orbital method ensures that two conditions are met:

(1) Charges from the MM region polarize the QM region. (The wavefunction is allowed to feel the effects of the MM point charges, similar to the link-atom method.)

(2) Atoms at the interface are allowed to adjust during the minimization.

To describe the interface of the MM atoms, point charges at the QM/MM interface are chosen to obtain a correct total charge of the entire QM/MM system, and the charge is placed in the middle of each frozen bond. The frozen bond is defined as the Boys localized[86] σ-orbital with the electron density delocalized over only the two adjacent atoms.[87] Note that the coordinates of the adjacent atoms connecting the bond are not spatially frozen; only the orbital representation of the bond remains constant. Given the need for the parameterization, only specific types of bonds can be selected to separate the QM from the MM regions. For protein residues, currently three bond types have been parameterized: the α-carbon/amino nitrogen bond, the α-carbon/carboxyl carbon bond of the protein backbone, and the α-carbon/β-carbon bond of the sidechains. This frozen orbital at the bond is adjusted throughout the QM/MM calculation to allow the adjacent atoms to move during the minimization. The adjustments entail first zeroing out the coefficients of the basis functions centered on all atoms except the two adjacent to the frozen bond. Next, the frozen orbital is aligned to the frozen bond using a rotation matrix. Then, using another matrix transformation, the molecular orbital corresponding to the frozen orbital is re-normalized to counter the effect of changing the orbital overlap when bond distances and angles change during the minimization. Modifying the normal Roothan equations to help keep the frozen orbitals fixed during the SCF energy

calculation rounds out the frozen orbital procedure. Conditions of the frozen orbital are introduced in the expression

$$F''_{\mu\upsilon} = F_{\mu\upsilon} - \sum_{\substack{i=nonfr \\ j=frozen}} \sum_{\alpha\beta} S_{\mu\alpha} C_{\alpha i} \in_{ij} C_{\beta j} S_{\beta\upsilon} \qquad \text{[Eq. 3.4]}$$

where F represents the standard Fock matrix, F'' represents the modified Fock matrix, S is the overlap matrix. C are the molecular orbital coefficients including fixed coefficients for frozen orbitals and $\in$ is the matrix of orbital energies. The accuracy of this approach has been tested by reproducing conformational energies of small peptides in polar environments.[85, 88]

A possible limitation of this approach is that it relies on having a sufficiently large database of bonds that have been parameterized for the frozen orbital method. Currently, parameters have been obtained at both the HF and DFT levels of theory with a 6-31G* basis set. With the absence of parameters for larger basis sets, mixed basis set approaches have been tested for accuracy and performed adequately. Here, accuracy is reflected with respect to reproducing quantum mechanically calculated binding energies for peptides.

Gradients are needed to perform energy minimization and search for stationary points on the potential energy surface. The frozen orbital method uses standard procedures for calculating the gradient with respect to atomic coordinates X_A.

$$\frac{\partial E_{tot}}{\partial X_A} = \frac{\partial E_{QM}}{\partial X_A} + \frac{\partial E_{MM}}{\partial X_A} + \frac{\partial E_{QM/MM}}{\partial X_A} \qquad \text{[Eq. 3.5]}$$

Calculating the gradients for the interface terms are similar to those of the MM terms with one exception. The presence of the point charges at the middle of the frozen bond requires a modification to the gradient expression with respect to the coordinates of the interface atoms.[85]

3.3 QM/MM Method in QSite

The QM/MM package used in this study is QSite. The formulation for QSites' interface approach has been summarized in the discussion of the frozen orbital method above. With QSite, we can locate and analyze stationary points, for example, minima and transition states, on the potential energy surface. The analysis provides both structural and energetic information about the system.

One potential weakness of this QM/MM method is in locating the transition states in protein systems. Proteins have complex potential energy surfaces, so transition states are not defined by a single configuration at the saddle point, but rather as an ensemble of saddle points in the transition state region.[89, 90] Likewise, the reactant and product are not single configurations, but they are an ensemble of configurations in the reactant and product basins of attraction. With this in mind, finding the transition state(s) in a protein would mean sampling the configurational space in the region of the transition state and averaging over the collection of configurations. Since *ab initio* QM/MM methods do not involve a sampling scheme, the transition state is approximated by using chemical intuition to classify the structure based on a set of relevant donor-acceptor distances at the active site. In standard QM theory, transition states are formally defined as states where

the frequencies for 3N-6 normal mode vibrations are real except for a single imaginary vibrational frequency. To obtain these frequencies, a 3N x 3N Hessian matrix is diagonalized, and a configuration with one negative eigenvalue of the Hessian characterizes the transition state as a saddle point.[91] Calculating frequencies for the entire solvated protein system is impractical, although some QM/MM methods have overcome this limitation.[92] The partial Hessian calculated in this method reflects only the degrees of freedom of the QM subsystem in the presence of point charges from the MM region. This is an approximation of the true Hessian for the entire system and assumes that the vibrations relevant to the reaction are localized in the QM region. Thus, we rely on the fact that proton transfer transition states as having donor-acceptor distances that are shorter than the equilibrium distance, and the transferring hydrogen atom is localized approximately midway between the donor and acceptor atom.

The QM/MM code combines two separate codes: Jaguar[93] to perform the QM portion of the calculations and IMPACT[94] to perform the MM portion of the calculations. In our calculations, the geometry of the MM region is first optimized with a frozen QM region. For each subsequent QM optimization step, a full optimization of the MM region is performed, and the procedure continues iteratively, alternating between the two regions in this manner until convergence is achieved. This method has been benchmarked on several enzyme systems (triosephosphate isomerase, [95] cytochrome P450, [96-99] beta lactamase,[100] methane monooxygenase,[101, 102] and hemerythrin[103]) and has been shown to provide accuracy to within 2-5 kcal/mol for reaction energies.

3.4 QM/MM Protocol for DHOD

In this subsection, we describe a procedure, utilizing the QM/MM methodology in QSite, applied to DHOD. The procedure delivers potential energy profiles for the concerted and sequential enzyme mechanisms. The size of the QM region was varied representing four models of the active site. The smallest QM region (model QM1) includes the substrate, sidechains of the Ser215 and Thr218 residues, and a water molecule in the active site for a total of 44 atoms. Adding the backbone from Ser214 to Asn219 to QM1 gives a 76 atom QM region (model QM2). Adding the FMN cofactor to QM1 gives a 94 atom region (model QM3). Finally, adding the FMN cofactor to QM2 gives a 126 atom region (model QM4). Table 3-1 provides a summary of the active site models.

Table 3-1: Active site models and number of atoms included in each model. When the size of the QM region is varied in the calculations, they will be classified as one of these four choices.

Active Site Model	Components included	Number of Atoms
QM1	Substrate, water, Ser215 & Thr218 sidechains	44
QM2	QM1 + protein backbone from Ser214 to Asn219	76
QM3	QM1 + FMN cofactor	94
QM4	QM2 + FMN cofactor	126

The impact of the starting configuration on the reaction energy profiles was also considered. Starting configurations can be obtained from molecular dynamics trajectories representing different hydrogen bonding networks. A network where the hydrogen bonds are aligned from serine through water, to the substrate is defined as the carboxylate path (CP). Additionally, a network where the hydrogen bonds are aligned

from the serine through water, to the substrate and threonine residue simultaneously is defined as the dual path (DP). Starting configurations can also be taken from the crystal structure (CS). A tree diagram of the required calculations is shown in Figure 3-1.

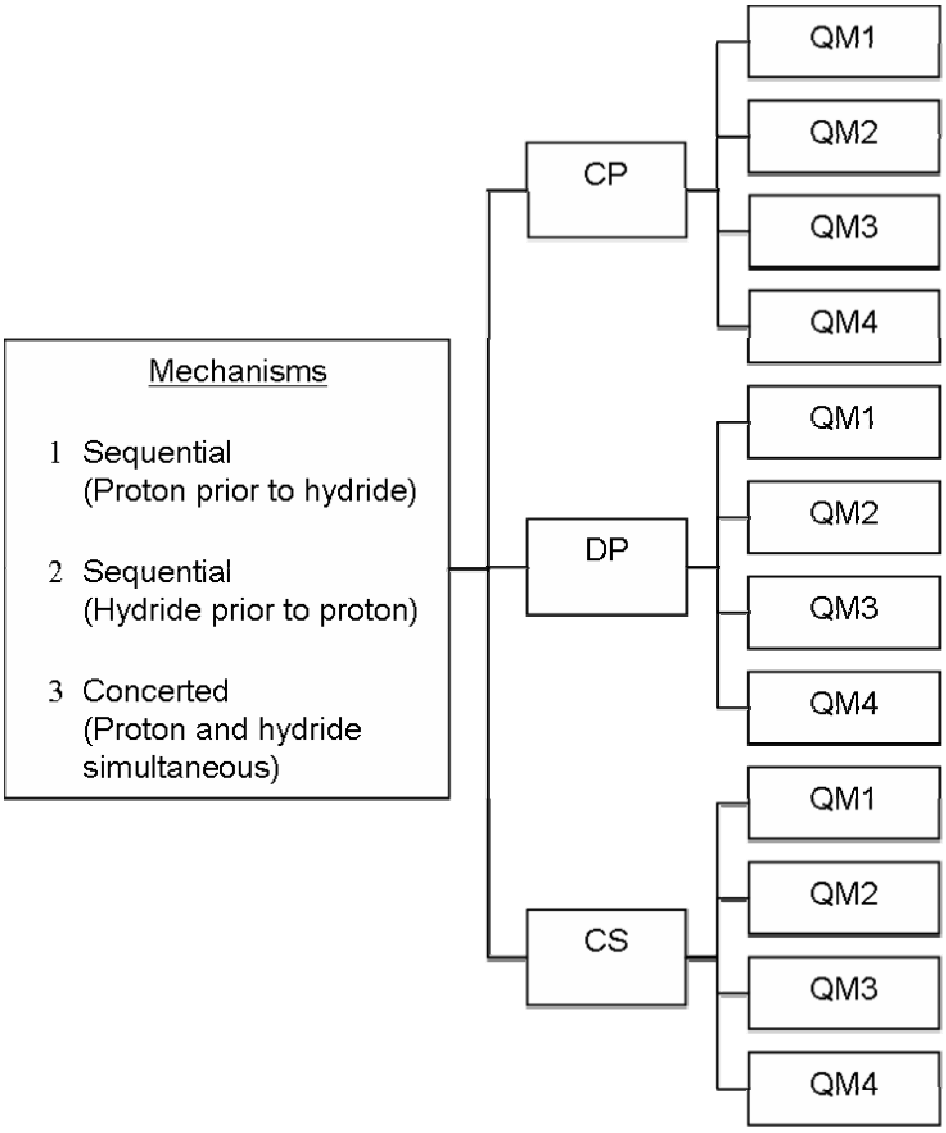

Figure 3-1: Tree diagram representing the number of independent calculations used to study three mechanisms in DHOD while varying the starting coordinates and the size of the QM region.

This approach generates 36 energy profiles that can be analyzed not only to determine the most relevant mechanism but also to spotlight the effect of varying both the starting configurations and size of the QM region. To complete each of the two sequential mechanism profiles, 3 QM/MM minimizations and 2 QM/MM transition state search calculations are needed. To complete the concerted mechanism profile, 2 QM/MM minimizations and 1 QM/MM transition state search calculations are needed. The system is being moved in the forward direction from reactants to products. However, a more thorough approach would be to repeat the procedure in the reverse direction from products to reactants. Essentially the entire approach calls for 156 independent calculations in the forward direction alone. The energy profiles are schematically represented in Figure 3-2. In the schematic, the x-axis signifies the progression of the reaction in the forward direction from reactants to products and the energies are taken relative to the reactant in all plots.

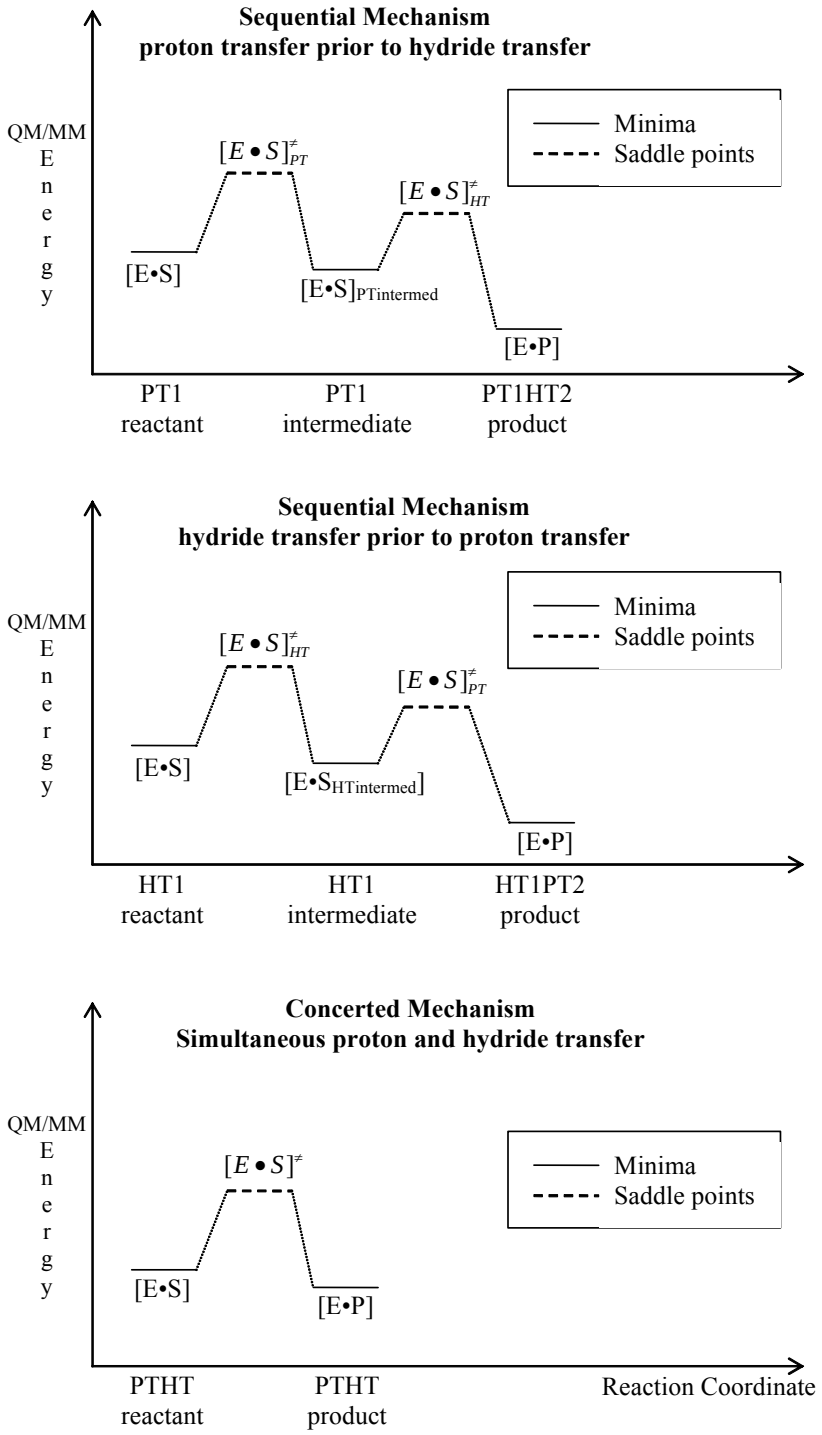

Figure 3-2: Schematic representations of the QM/MM energy profiles for three DHOD mechanisms. Straight lines represent energy minima and dashed lines represent saddle points.

Chapter 4

Hydrogen Bonding Pathways in Human Dihydroorotate Dihydrogenase

Reproduced in part with permission from Y. A. Small, V. Guallar, A. V. Soudackov, and S. Hammes-Schiffer, Journal of Physical Chemistry B **110**, 19704-19710 (2006).

4.1 Introduction

Dihydroorotate dehydrogenase (DHOD) catalyzes the only redox reaction in the pathway for pyrimidine biosynthesis. In this reaction, dihydroorotate (DHO) is oxidized to orotate (ORO) by the enzyme-bound flavin mononucleotide (FMN). Specifically, the C5 pro-*S* hydrogen of DHO is thought to be removed as a proton by an enzymatic base, and the C6 hydrogen of DHO is transferred as a hydride to the isoalloxazine ring of the flavin.[104-106] Pyrimidines are required for the supply of precursors for RNA and DNA synthesis. The diversity among DHODs of different organisms[107] enables the development of compounds that selectively inhibit pyrimidine biosynthesis in some organisms while not affecting others. As a result, these enzymes are promising targets for treating cancer, malaria, gastric ulcers, and rheumatoid arthritis.[108-111] One prominent example is the immunosuppressive drug leflunomide,[112] which inhibits human DHOD and has been used in the clinical treatment of rheumatoid arthritis.[110, 113, 114]

DHOD enzymes have been divided into two classes based on sequence homology.[115] Class 1 enzymes are soluble, located in the cytosol, and found in Gram-positive bacteria and some lower eukaryotes. Class 2 enzymes are associated with cell

membranes and are found in most eukaryotes and Gram-negative bacteria. Class 1 is further divided into 1A and 1B. The 1A enzymes are thought to use fumarate to reoxidize the flavin. In contrast, the reoxidation of the flavin in the 1B enzymes is thought to involve a series of oxidizing reactions involving an iron-sulfur center and other cofactors. The class 2 enzymes are thought to use ubiquinone to reoxidize the flavin. An important distinction is that class 1 enzymes use cysteine as the active site base for deprotonation of the substrate, while class 2 enzymes use serine for this purpose.[104, 106, 116-118] The kinetics of the hydride and proton transfer reactions in the class 1A, 1B, and 2 enzymes have been studied extensively.[117, 119-122] Structures have been determined for the class 1A and 1B enzymes from *Lactococcus lactis*[104, 105, 107, 123] and the class 2 enzyme from rat,[124] human[106] and *E. coli*.[118] The structures are all similar at the orotate binding site except for the active site base.

The present study focuses on human DHOD, which is a class 2 enzyme. The X-ray crystallographic structure of human DHOD[106] suggests that the catalytic base, Ser215, is hydrogen bonded to a tightly bound water molecule, which in turn is hydrogen bonded to Thr218. Both Ser215 and Thr218 are conserved in class 2 DHODs. Typically serine does not play the role of a catalytic base in enzymatic catalysis. Hydrogen bonding in the active site could enhance the basicity of Ser215, however, thereby facilitating the deprotonation of the substrate. Moreover, deprotonation of the substrate could involve a proton relay mechanism along a hydrogen bonding pathway in the active site. Although cysteine plays the role of the catalytic base in class 1 DHOD enzymes, mutagenesis of the active serine residue to cysteine was found to reduce the enzyme activity 4-fold in human DHOD[125] and 500-fold in *E. coli* DHOD.[115]

In this chapter, we use computational methods to investigate potential proton relay pathways in the active site of human DHOD. In addition to studying wild-type (WT) human DHOD, we also study the mutant enzyme in which Ser215 is replaced by cysteine. Despite the biological importance of this enzyme, there have not been any previous molecular dynamics studies on DHOD, although similar hydrogen bonding analyses have been performed for other enzymes.[126-129] An outline of the chapter is as follows: Section II describes the computational methods, section III presents the results and analysis, and the final section summarizes the conclusions.

4.2 Methods

The starting coordinates were obtained from the X-ray crystallographic structure for human DHOD (PDB code 1D3G).[106] In this crystal structure, the product ORO is bound to the enzyme, and the cofactor FMN is in the reactant oxidized form. The crystal structure includes the substrate and cofactor, 396 resolved residues, 274 resolved crystallographic water molecules, and seven unresolved residues that were filled using side-chain sampling and the rotamer library in the JACKAL suite of programs.[130] JACKAL was also used to add hydrogens to the protein and to determine the residue protonation states based on the local hydrogen bonding networks. Hydrogens were added to the substrate and cofactor to model the reactant state of the enzyme, in which the substrate is DHO and the cofactor is oxidized FMN, as depicted in Figure 4-1 . All of the hydrogen coordinates were optimized with IMPACT[131] using the OPLS-2001 forcefield.[132]

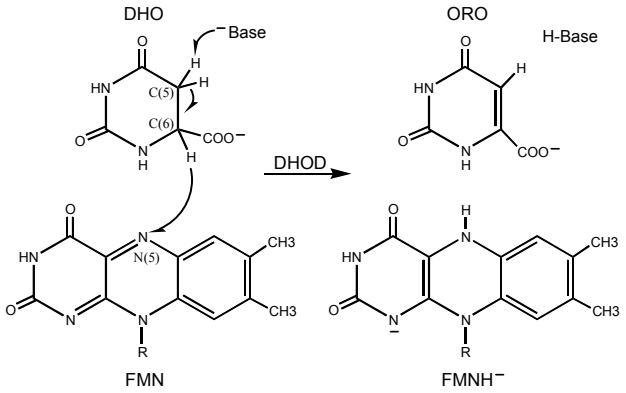

Figure 4-1: The proton and hydride transfer reactions catalyzed by DHOD. The substrate DHO is oxidized to ORO, and the cofactor FMN is reduced. The base is cysteine in class 1 DHOD enzymes and serine in class 2 DHOD enzymes.

The substrate and cofactor geometries for the reactant state were optimized with QSite[133] using a mixed quantum mechanical/molecular mechanical (QM/MM) method.[134] In these calculations, the substrate and cofactor were treated quantum mechanically at the B3LYP/6-31G** level,[135-139] and the remaining residues and crystallographic waters were treated with the OPLS-2001 forcefield.[132] The coordinates of the protein residues and crystallographic waters were fixed during the optimization of the substrate and cofactor. After this optimization, the Ser215Cys mutant enzyme was generated by replacing the serine with cysteine (i.e., replacing the oxygen with sulfur).

We performed calculations on both the WT DHOD and the Ser215Cys mutant DHOD enzymes. Both enzymes were solvated with simple point charge (SPC) water molecules[140] in a periodically replicated cubic box with sides of length 70 Å. The non-zero charge of +10 on the system was neutralized by converting solvent molecules near positively charged residues on the periphery of the enzyme to negatively charged chloride ions. Note that these chloride anions were more than 15 Å from the active site. To

ensure a smooth transition from the structures obtained with the QM/MM calculations, which used the OPLS-2001 forcefield, the system was first equilibrated with IMPACT[131] using the OPLS-2001 forcefield.[132] The charges for the substrate and cofactor were determined with the electrostatic potential (ESP) method in QSite.[141-143] The temperature was maintained with a Berendsen thermostat,[144] and the long range electrostatic interactions were treated with the Ewald method.[145] The coordinates of the substrate and cofactor were fixed to the QM/MM optimized values during the equilibration of the enzyme and solvent. In the initial preparation of the system, the enzyme was fixed, and the solvent was minimized with the steepest descent method for 200 steps. Then the enzyme backbone was fixed, and the solvent, hydrogen atoms of the enzyme, and sidechains of the enzyme were minimized with the steepest descent method for another 200 steps. Finally, all atoms in the system were minimized with the steepest descent method for 200 steps and with the conjugate gradient method for 300 steps.

After this initial preparation of the system, three phases of molecular dynamics were performed. In each phase, the system was slowly heated in 50 K increments from 0 to 300 K for 2 ps at each temperature, run at 300 K for 40 ps, and then slowly cooled in 50 K increments from 250 to 0 K for 2 ps at each temperature. New velocities were generated from a Boltzmann distribution at each temperature. The substrate and cofactor were fixed at the QM/MM optimized values throughout this simulated annealing process. In phase I, only the solvent and hydrogen atoms were unconstrained. In phase II, the solvent, hydrogen atoms, and residues within 3 Å of the substrate were unconstrained. In phase III, the solvent and all enzyme residues were unconstrained.

The trajectories used for the hydrogen bonding analysis were generated with the GROMACS molecular dynamics package.[146, 147] For these simulations, we used the GROMACS forcefield[146, 147] with the particle mesh Ewald method[148] for long-range electrostatic interactions. To be consistent with the GROMACS forcefield, the charges on the substrate and cofactor were determined with the method described in ref [149]. The protein, substrate, cofactor and solvent were each separately coupled to Nose-Hoover thermostats[150, 151] to maintain the temperature. The SHAKE algorithm[152] was used to constrain all of the X-H bonds throughout the simulation. The timestep for these simulations was 1 fs.

Starting with the final structure from the simulated annealing procedure, the system was equilibrated with GROMACS to account for the differences in forcefields. First a slow heating in 50 K increments for 2 ps at each temperature from 0 K to 300 K was performed with the coordinates of the substrate and cofactor fixed to the QM/MM optimized values. Starting with the final structure from this procedure, four independent trajectories were initiated at 300 K by choosing random velocities from a Boltzmann distribution. In these trajectories, the substrate and cofactor were no longer fixed to the QM/MM optimized values. The distance between the center of mass of the proton and hydride donor atoms and the proton acceptor was constrained to the initial values using the SHAKE algorithm[152] for the first 300 ps of molecular dynamics to prevent drifting. Then this constraint was removed, and the system was further equilibrated with an additional 300 ps of molecular dynamics. After this equilibration procedure, we generated 1 ns of molecular dynamics data for each independent trajectory, leading to 4 ns of data for the WT enzyme and 4 ns of data for the Ser215Cys mutant enzyme.

4.3 Results

We calculated the root mean square deviation (RMSD) of the alpha carbons relative to the crystal structure for all of the trajectories. The average RMSD over the four 1 ns trajectories is 1.56 Å for the WT enzyme and 1.66 Å for the mutant enzyme. We also calculated the time-averaged root mean square fluctuation (RMSF) per residue. Figure 4-2 depicts the RMSF for representative WT and mutant trajectories. As expected, the most flexible residues are located near the surface of the protein, whereas the hydrophobic core of the protein maintains a relatively rigid structure. The similarity of the RMSF for the WT and mutant enzymes suggests that the mutation does not cause major structural rearrangements.

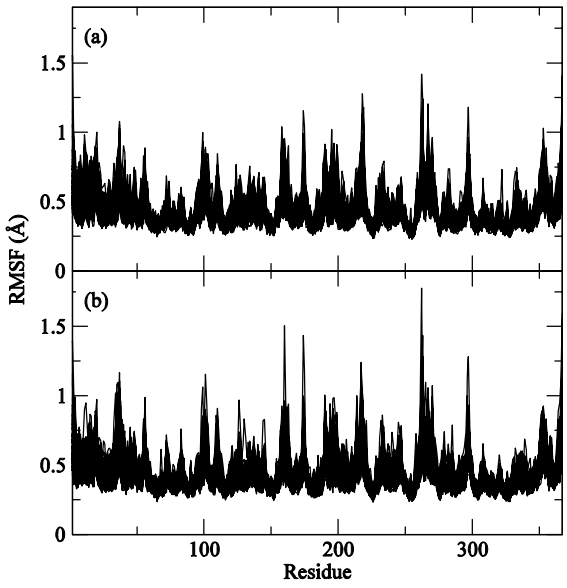

Figure 4-2: Root Mean Square fluctuations of the Cα atoms with respect to the average values for a representative (a) WT trajectory and (b) Ser215Cys mutant trajectory.

We monitored the distance between the donor and acceptor atoms for the proton and hydride transfer reactions involving the substrate. For the WT enzyme, the average proton donor-acceptor distance is 3.55 ± 0.36 Å, and the average hydride donor-acceptor distance is 3.68 ± 0.27 Å. For the mutant enzyme, the average proton donor-acceptor distance is 4.25 ± 0.36 Å, and the average hydride donor-acceptor distance is 4.18 ± 0.12 Å. These donor-acceptor distances represent average equilibrium distances, and occasionally shorter distances are sampled. Although these average donor-acceptor distances are too large for hydrogen transfer, thermal fluctuations of the system lead to sampling of smaller distances that enable proton and hydride transfer. The greater average donor-acceptor distances for the mutant are consistent with the significantly lower activity of the mutant enzyme. The greater proton donor-acceptor distance for the mutant may arise from the larger van der Waals radius of the sulfur atom in cysteine compared to the oxygen atom in serine.

We also monitored the geometry and orientation of the substrate. Figure 4-3 depicts two distinct substrate conformations that have different values of the angle α involving the carboxylate carbon atom and two ring atoms. For the WT trajectories, the average value of the angle α is 158^{o} ± 2. Moreover, a water molecule remains near Ser215 in all of the WT trajectories. In three of the four mutant trajectories, however, the carboxylate group of the substrate bends toward the cysteine base, so the average angle α is 111^{o} ± 4, and the nearby water molecule is expelled. The larger van der Waals radius of the sulfur atom may induce the expulsion of the water molecule, which in turn may cause the carboxylate to bend toward the cysteine base to stabilize the negative charge on the carboxylate through hydrogen bonding. The bent substrate conformation and absence

of the nearby water molecule widens the cavity in the vicinity of the cysteine residue. This widening may cause the interruption of the network of hydrogen bonds that are critical for the proton relay pathway and hence may contribute to the reduction in the mutant activity.

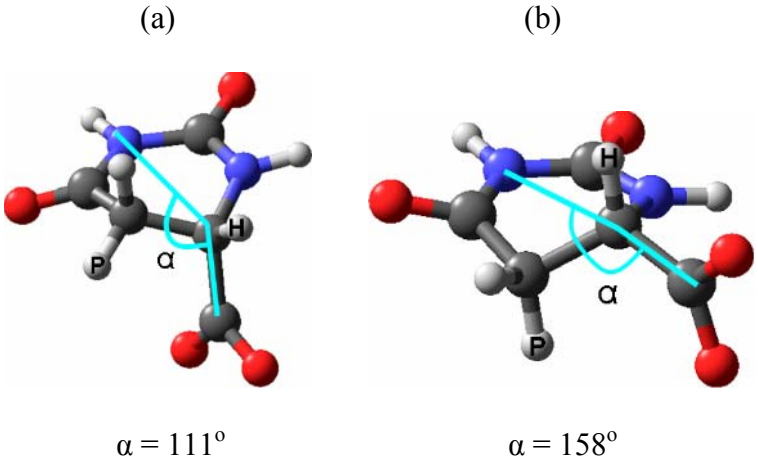

Figure 4-3: The two dominant conformations of the substrate DHO observed in the Ser215Cys mutant (left) and the WT (right) trajectories. The transferring proton is labeled as P, and the transferring hydride is labeled as H.

To assist in our analysis of hydrogen bonding pathways, we quantified the presence of a water molecule in the active site cavity. For this purpose, we defined a sphere of radius 3.3 Å centered at the center of mass of the oxygen atoms on Ser215, Thr218, and the carboxylate group of the substrate. The percentage of time that a water molecule is found in this sphere for each trajectory is given in Table 4-1 . For the WT trajectories, a water molecule is hydrogen bonded in the active site cavity virtually 100% of the time for three of the trajectories and 93% of the time for the other trajectory. For the mutant trajectories, a water molecule is absent from the active site cavity virtually

100% of the time for three of the trajectories and is present 100% of the time for the other trajectory. For the mutant trajectory with a water molecule in the active site cavity, the water molecule hydrogen bonds to Thr218 5% of the time, to the carboxylate group of the substrate 10% of the time, and to both Thr218 and the carboxylate group 4% of the time. This water molecule does not participate in the complete hydrogen bonding pathways discussed below, however, because Cys215 is too far away from the water molecule.

Table 4-1: Percentage of time that each hydrogen bonding pathway was observed for each of the four independent WT trajectories. The WT hydrogen bonding pathways are depicted in Figure 4-4 . The percentage of time that a water molecule is present in the active site cavity as defined in the text, is also given.

	Traj 1	Traj 2	Traj 3	Traj 4
Water present	100	99	93	100
Carboxylate path	98	0.79	87	73
Threonine path	0.019	93	0.05	1.8
Dual path	0.022	0.14	0.17	22

A proton relay mechanism has been hypothesized to facilitate the proton transfer reaction from the substrate to Ser215. To gain insight into this proton relay mechanism, we analyzed the hydrogen bonding networks for each WT trajectory. We identified three hydrogen bonding pathways that could facilitate the proton transfer reaction from the substrate to Ser215 in the WT trajectories. All of these pathways involve a water molecule near the substrate and Ser215. In the first pathway, the hydroxyl group of Ser215 is hydrogen bonded to a water molecule, which is hydrogen bonded to the carboxylate group of the substrate. In the second pathway, the hydroxyl group of Ser215 is hydrogen bonded to a water molecule, which is hydrogen bonded to the hydroxyl

group of Thr218. In the third pathway, the first two pathways are aligned simultaneously.

Figure 4-4 depicts these three pathways, which are denoted the carboxylate pathway, the threonine pathway, and the dual pathway, respectively. Note that the switch between the carboxylate and threonine pathways is associated with a rotation of the hydroxyl group of Thr218. For the threonine and dual pathways, the threonine is positioned to enable proton transfer to another water molecule leading to the bulk solvent.

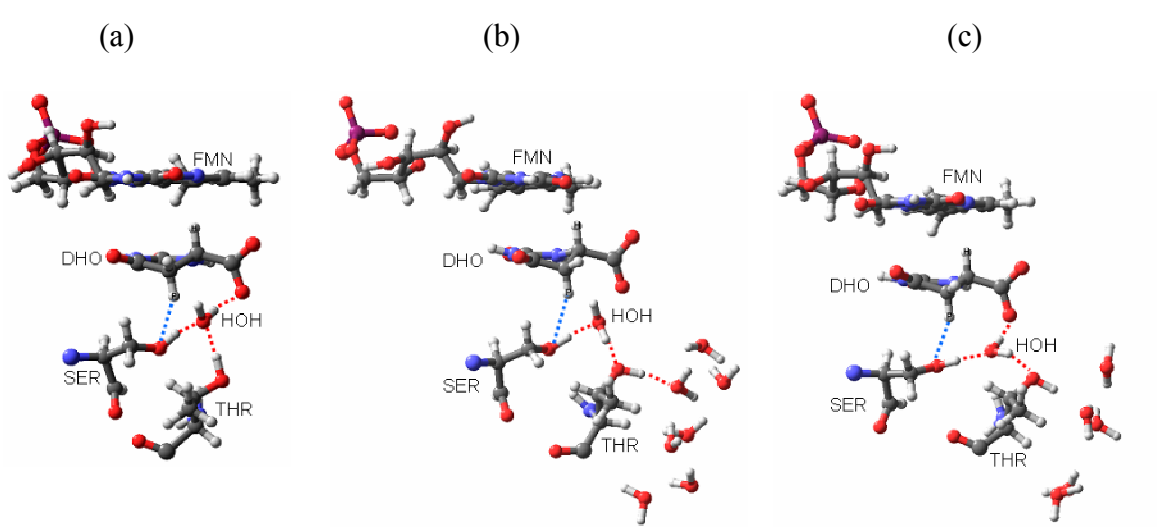

Figure 4-4: Hydrogen bonding pathways observed in the WT trajectories: (a) carboxylate pathway; (b) threonine pathway; (c) dual pathway. The transferring proton is labeled as P, and the transferring hydride is labeled as H.

The direct access of the hydroxyl group of Thr218 to the bulk solvent is illustrated in Figure 4-5. In contrast, the substrate carboxylate group was not observed to hydrogen bond to water molecules directly accessible to the bulk solvent. An advantage of the dual pathway over the threonine pathway is that the carboxylate group stabilizes the

intervening water molecule. For the threonine pathway, the substrate carboxylate group may be stabilized by surrounding residues such as Lys100.

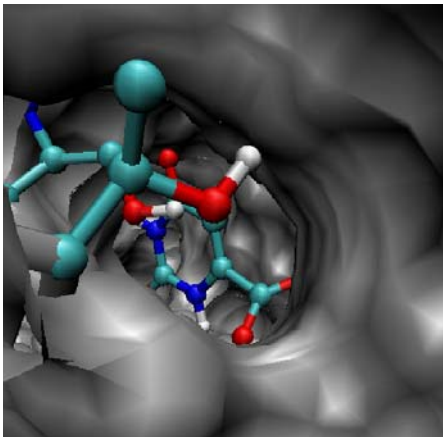

Figure 4-5: Illustration of the direct access of bulk solvent to the hydroxyl group of Thr218. The coordinates were obtained from the X-ray crystallographic structure for human DHOD. Missing residues and hydrogen atoms were added with JACKAL. The van der Waals surface created from all atoms is shown in gray. The perspective is from the exterior of the protein. Thr218 is shown in the foreground, the side chain of Ser215 is shown in the middle, and the substrate is shown in the background.

Table 4-1 gives the percentage of time that each pathway was observed for each of the four independent WT trajectories. The criteria used to define a hydrogen bond were a distance of less than 3.3 Å between the donor and acceptor and an angle of less than 35^{o} between the donor-hydrogen and donor-acceptor vectors. We did not observe a significant number of switches between the different hydrogen bonding pathways during each of the 1 ns trajectories. As a result, we are unable to provide statistically meaningful probabilities or free energies of formation for each pathway.

To further characterize the differences among the hydrogen bonding pathways, we calculated the pair distribution function for the water molecules relative to the oxygen atom of Thr218. This pair distribution function is defined as[153, 154]

Eq. 4.1

$$g(r) = \frac{1}{\rho_{\text{water}}} \left\langle \sum_{i=1}^{N_{\text{water}}} \delta(r_i - r) \right\rangle \qquad \text{[Eq. 4.1]}$$

where the summation is over the N_{water} water molecules in a sphere of radius 35 Å centered at the oxygen of Thr218, ρ_{water} is the average density of water molecules in this sphere, and r_i is the distance between the oxygen of Thr218 and the oxygen of the *i*th water molecule. The quantity in brackets is calculated in terms of the number density of water molecules in a spherical shell of radius *r* centered at the oxygen of Thr218. Thus, *g(r)* indicates the relative density of water molecules at distance *r* from the oxygen of Thr218 compared to the bulk water density.

We calculated the pair distribution function *g(r)* for two 300 ps regions of the trajectories. The first region corresponds to 96% probability of the carboxylate pathway, and the second region corresponds to 98% probability of the threonine pathway. Figure 4-6 depicts the results, which were found to be qualitatively similar for three different data sets for each pathway. In all cases, we verified that *g(r)* becomes unity at large *r*. Note that the first peak in the pair distribution function is approximately twice as high for the threonine pathway than for the carboxylate pathway. The average number of water molecules directly hydrogen bonded to Thr218 can be calculated by integrating over the first peak of the pair distribution function: Eq. 4.2

$$N_{\text{H-bond}} = \rho_{\text{water}} \int_{\text{1st peak}} 4\pi r^2 g(r) dr \qquad \text{[Eq. 4.2]}$$

$N_{\text{H-bond}}$ is 2.18 for the threonine pathway and 1.06 for the carboxylate pathway. This observation suggests that the hydrogen bonding pathway to Thr218 is associated with greater stabilization of Thr218 by hydrogen bonding to nearby water molecules.

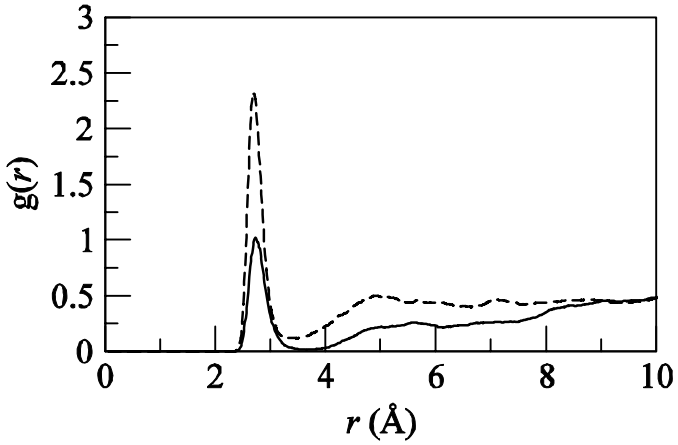

Figure 4-6: Pair distribution functions for the oxygen in Thr218 to the oxygens in the surrounding water molecules for the carboxylate pathway (solid) and the threonine pathway (dashed).

We also investigated the impact of the mutation of Ser215 to cysteine on the hydrogen bonding pathways. As discussed above, in three out of the four mutant trajectories, the carboxylate group of the substrate bends toward the cysteine base, and the nearby water molecule is expelled. These structural rearrangements prevent the formation of the hydrogen bonding pathways that were observed in the WT trajectories. We observed two hydrogen bonding pathways in these mutant trajectories. In the first pathway, the thiol group of Cys215 is directly hydrogen bonded to the carboxylate of the substrate. In the second pathway, the thiol group of Cys215 is directly hydrogen bonded to Thr218.

Figure 4-7 depicts these two hydrogen bonding pathways, and Table 4-2 gives the percentage of time that each pathway was observed for each of the four mutant trajectories. Note that the threonine pathway appears to be much less probable than the carboxylate pathway. For the three trajectories in which the substrate assumed the bent conformation and the nearby solvent molecule was expelled, the hydride transfer reaction from the substrate to the cofactor was inhibited because the transferring hydride was no longer oriented toward the cofactor. For the trajectory in which the substrate did not assume the bent conformation and the nearby solvent molecule was not expelled, we did not observe any hydrogen bonding pathways involving both Cys215 and this water molecule because they were too far apart.

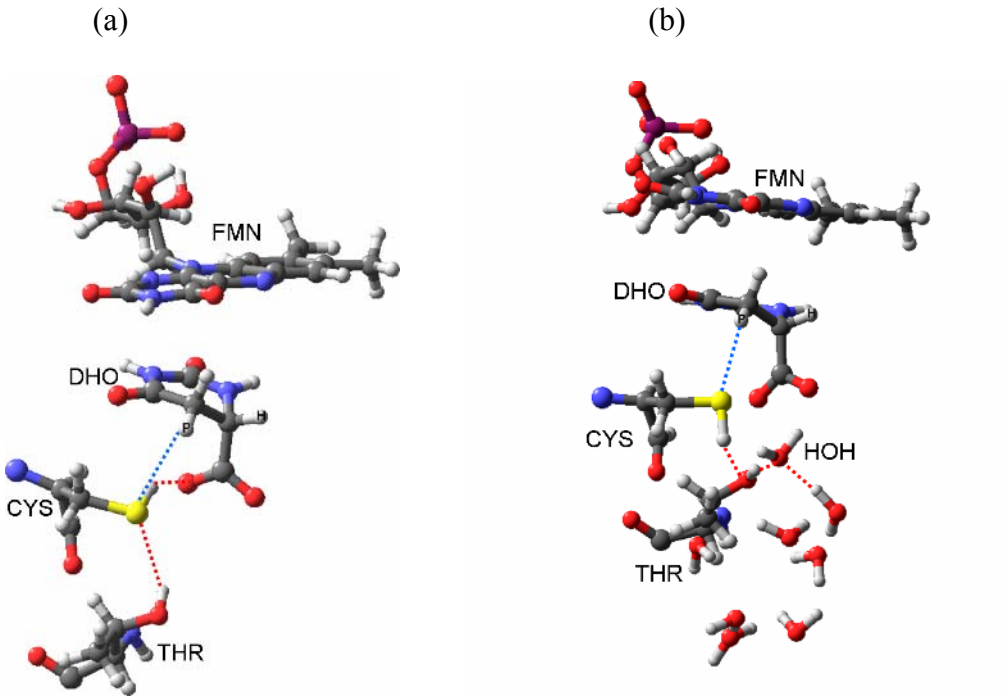

Figure 4-7: Hydrogen bonding pathways observed in the Ser215Cys mutant trajectories for which the intervening water molecule is expelled: (a) carboxylate pathway; (b) threonine pathway. The transferring proton is labeled as P, and the transferring hydride is labeled as H.

As for the WT trajectories, we were unable to obtain statistically meaningful results for the probabilities and free energies of formation of these hydrogen bonding pathways. As mentioned above, in addition to disrupting the hydrogen bonding pathways formed in the WT trajectories, the mutation of Ser215 to cysteine significantly increased the average proton and hydride donor-acceptor distances.

Table 4-2: Percentage of time that each hydrogen bonding pathway was observed for each of the four independent Ser215Cys mutant trajectories. The mutant hydrogen bonding pathways are depicted in Figure 4-7. Note that these pathways involve direct hydrogen bonding between Cys215 and the carboxylate group of the substrate or Thr218, and they do not involve an intervening water molecule. The percentage of time that a water molecule is present in the active site cavity, as defined in the text, is also given.

	Traj 1	Traj 2	Traj 3	Traj 4
Water present	2.0	0.02	100	0.00
Carboxylate path	26	44	0.04	41
Threonine path	1.92	1.73	0.00	1.42

The pK_a of cysteine indicates that this amino acid could be deprotonated in the active site. We would expect that the electrostatic repulsion between the negatively charged thiolate and carboxylate groups would render the bent conformation of the substrate energetically unfavorable. To explore this possibility, we also performed analogous simulations with deprotonated Cys215. In this case, we found that the substrate did not assume the bent conformation, but the proton donor-acceptor distance increased significantly, thereby inhibiting the proton transfer reaction from the substrate to Cys215. These simulations indicate that the deprotonated cysteine also disrupts the hydrogen bonding network in the active site.

Residues Lys100 and Lys255 are thought to play a catalytic role in this enzyme reaction through hydrogen bonding interactions with the FMN cofactor.[104, 125] Mutation of Lys100 in human DHOD has been shown experimentally to abolish enzymatic activity.[125] We analyzed the hydrogen bonding interactions of Lys100 and Lys255 to the FMN cofactor in our molecular dynamics simulations. Figure 4-8 illustrates these hydrogen bonding interactions for a representative configuration sampled during a WT trajectory.

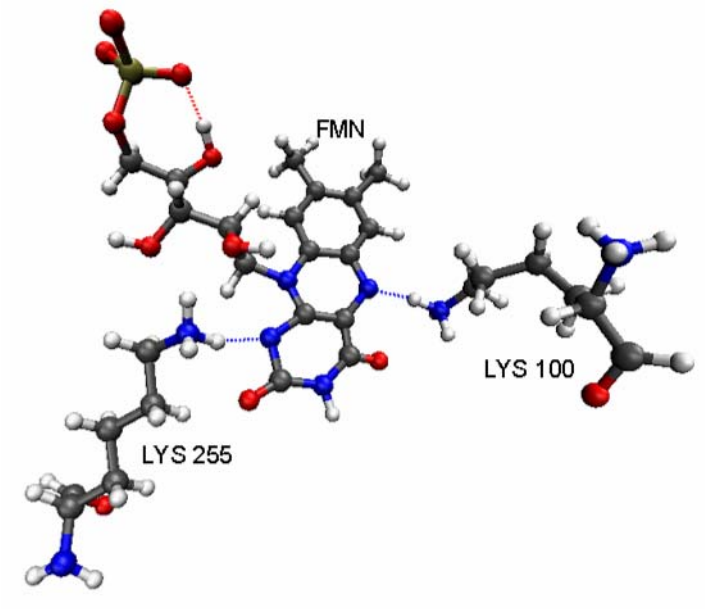

Figure 4-8: Hydrogen bonding of Lys100 and Lys255 to FMN for a representative configuration sampled in a WT trajectory. These hydrogen bonding interactions are maintained throughout virtually all of the trajectories.

We found that Lys100 and Lys255 are hydrogen bonded to the FMN cofactor for at least 99% of the time in all four WT trajectories. These hydrogen bonding interactions are also observed for at least 99% of the time in three of the mutant trajectories and at least 90% of the time in the other mutant trajectory. Thus, our simulations provide

support for the hypothesis that these lysine residues enhance the structural stability of the active site by hydrogen bonding to the FMN cofactor.

4.4 Conclusions

In this chapter, we identified and characterized potential proton relay pathways that could facilitate the redox reaction catalyzed by human DHOD. In this redox reaction, a hydride is transferred from C6 of the DHO substrate to the FMN cofactor, and a proton is transferred from C5 of the DHO substrate to a serine residue in the active site. Typically serine does not play the role of a catalytic base. Hydrogen bonding interactions within the active site have been postulated to enhance the basicity of the active Ser215, however, and deprotonation of the substrate may involve a proton relay mechanism along a hydrogen bonding pathway in the active site.

We identified three types of hydrogen bonding pathways in the active site of WT DHOD. In the first pathway, the hydroxyl group of Ser215 is hydrogen bonded to a water molecule, which is hydrogen bonded to the carboxylate group of the substrate. In the second pathway, the hydroxyl group of Ser215 is hydrogen bonded to a water molecule, which is hydrogen bonded to the hydroxyl group of Thr218. The third pathway is a dual pathway that encompasses both of these pathways. Note that a proton relay mechanism involving the substrate carboxylate group is not consistent with experiments indicating that esters of DHO are good substrates for the bovine liver DHOD enzyme.[115, 120, 155] These experiments suggest that a free carboxylate anion is not necessary for the enzyme reaction. These experimental results provide support for the

proton relay mechanism through Thr218. Our analysis indicated that the presence of the hydrogen bonding pathway to Thr218 is associated with greater stabilization of Thr218 by hydrogen bonding to nearby water molecules leading to the bulk. An advantage of the dual hydrogen bonding pathway is that the carboxylate stabilizes the intervening water molecule that is essential for proton transfer along the pathway through Thr218 to the bulk solvent.

The mutation of Ser215 to cysteine in the related *E. coli* DHOD was found experimentally to significantly decrease the enzyme activity.[115] In general, cysteine is more reactive as a catalytic base than serine. Moreover, cysteine plays the role of the catalytic base in class 1 DHOD enzymes. To elucidate the impact of this mutation of the active base, we studied the Ser215Cys mutant human DHOD by replacing Ser215 with cysteine in our simulations. The similarity of the calculated root mean square fluctuations for the WT and mutant trajectories indicates that the mutation does not significantly alter the overall structure and motion of the enzyme. The average substrate proton and hydride donor-acceptor distances, however, are significantly greater in the mutant than in the WT trajectories. Furthermore, in the majority of the mutant trajectories, the water near the substrate was expelled from the active site region, and the substrate assumed a different conformation in which the carboxylate group bends toward the cysteine base. These structural rearrangements prevent the formation of the hydrogen bonding pathways that were observed in the WT trajectories and inhibit hydride transfer from the substrate to the cofactor. Alternative hydrogen bonding pathways involving direct hydrogen bonding between Cys215 and the carboxylate or, to a lesser extent, direct hydrogen bonding between Cys215 and Thr218 were observed. Even in the mutant

trajectory for which the substrate maintained the WT conformation and the nearby water molecule remained, Cys215 did not form productive hydrogen bonding pathways. In simulations of the Ser215Cys mutant with a deprotonated Cys215, the substrate did not assume the bent conformation, but the hydrogen bonding network in the active site was disrupted and the proton donor-acceptor distance increased to an extent that would inhibit the proton transfer reaction.

The differences observed for the mutant DHOD trajectories may arise from the larger van der Waals radius of the sulfur atom in cysteine compared to the oxygen atom in serine. The disruption of the hydrogen bonding pathways and the greater donor-acceptor distances could lead to significant decrease in activity. The class 1 DHOD enzymes could utilize an alternative proton relay mechanism involving, for example, direct proton transfer from the cysteine base to the substrate carboxylate. For both the WT and the mutant trajectories, however, the substrate carboxylate group was not observed to hydrogen bond to water molecules directly accessible to the bulk solvent.

The inhibition of DHODs has been found to be an effective way to block pyrimidine nucleotide biosynthesis. As a result, these enzymes are promising targets for drugs treating a wide range of diseases. Thus, the elucidation of the mechanism of DHOD is important for the design of inhibitors that selectively impact the activity of only certain members of the enzyme family. Computational studies on parasitic DHOD enzymes, as well as human DHOD, will provide additional insights for drug design.

Chapter 5

QM/MM Findings on Dihydroorotate Dehydrogenase

5.1 Introduction

The possibility of a stepwise or concerted proton and hydride transfer in the DHOD active site is at the root of the ongoing dialogue about DHOD. Published Kinetic Isotope Effect (KIE) studies show rates that represent both oxidation and reduction half-reactions. Studies propose that the class I proton and hydride transfer is sequential, but the order of these transfers is still unresolved. Class II DHOD currently has two conflicting proposals for the proton and hydride transfer reaction. Earlier KIE studies point to a concerted transfer,[65] while later single molecule studies point to a sequential mechanism.[74] Even when the two halves of the redox process can be isolated, these studies are unable to isolate the rate of the proton and/or hydride transfer steps during the timeframe of FMN switching from oxidized to reduced states. More importantly, for the argument of a stepwise mechanism these studies are unable to establish the order of transfers, making it unclear whether proton proceeds before hydride or vice versa. The open questions of sequential or concerted mechanisms in DHOD leave a lot of room for computational methods to address the issues and offer predictive contributions to the ongoing discussion around DHOD and enzyme mechanisms in general.

In this chapter, a computational approach to address the sequential versus concerted mechanistic question will be described. A defined computational protocol

utilizing a quantum mechanical/molecular mechanical (QM/MM) approach was carried out with three mechanisms in mind. The mechanisms are: proton transfer prior to hydride transfer (PT1HT2), hydride transfer prior to proton transfer (HT1PT2) and concerted proton and hydride transfer (PTHT). The objective is to compare the reaction energies for each of the three charge transfers in order to quantify which mechanism is most likely to occur in class II DHOD. This QM/MM approach also allows for the characterization of features near the active site that may contribute to reducing the reaction energy, such as the effect of changing the size of the QM region, varying the initial coordinates, changing protonation states and altering the initial donor-acceptor distances.

5.2 Methods

The QM/MM method of choice is the *ab initio* QM/MM approach implemented in QSite (ref). The QM region is treated with density functional theory using the B3LYP functional and a 6-31G** basis set. The MM region is treated with the OPLS-2001[156] all-atom force field using a non-bonded cut-off of 50 Å. The shell of atoms beyond 20 Å of the substrate was frozen throughout all of the calculations. If covalent bonds separate the QM and MM regions, the method utilizes a frozen orbital approach, where the frozen bonds have been adequately parameterized and benchmarked for all protein residues. Technical details of the implementation have been described in chapter 3 and more robustly in reference 85. The protocol utilized in this study features a series of constrained and unconstrained QM/MM optimization calculations illustrated

schematically in Figure 5-1 . Two key considerations of this protocol are reproducibility and stability of the reactant, product and intermediate states defined by coordinates and energies. To ensure reproducibility, this protocol was executed starting with three initial configurations taken from crystal structures and MD trajectories of our previous study.[157]

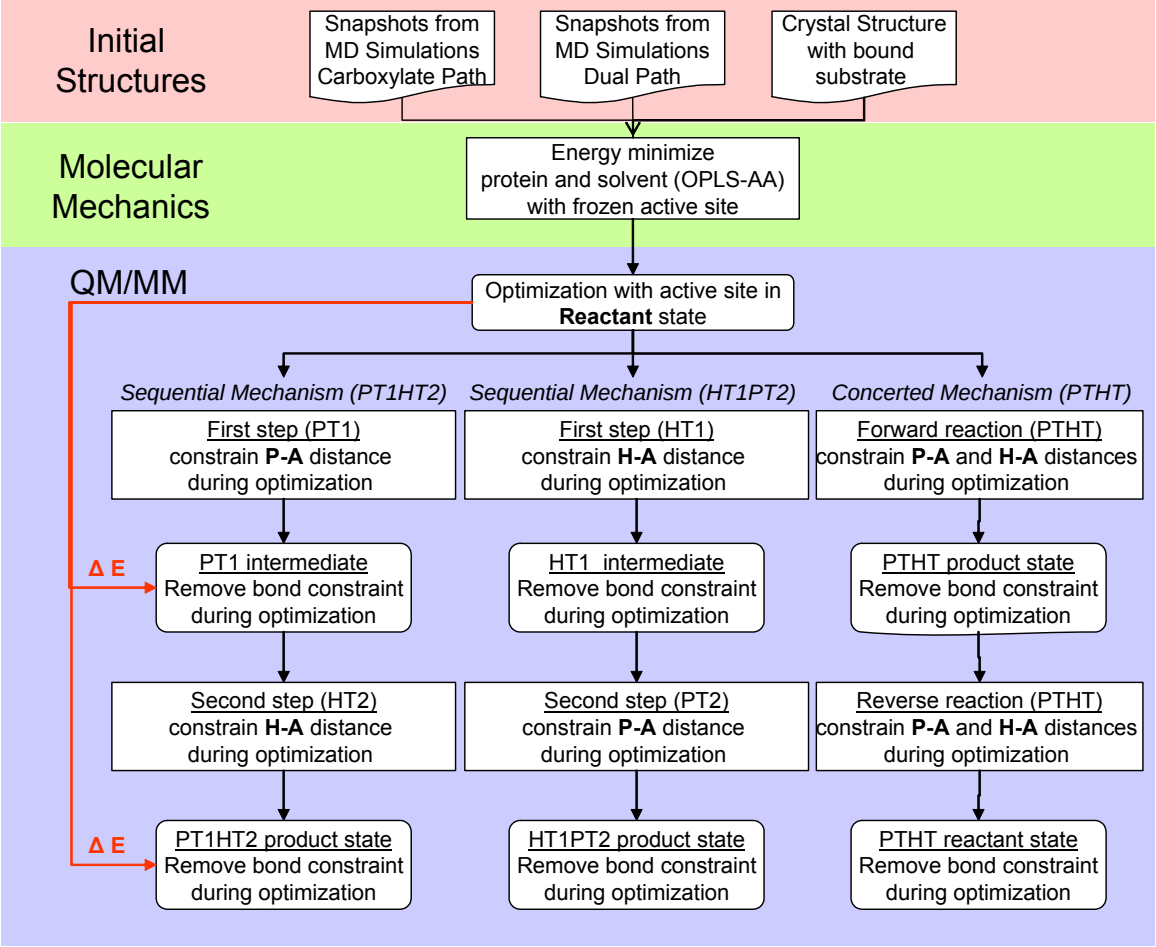

Figure 5-1: Protocol of constrained and unconstrained QM/MM calculations to study three mechanisms in DHOD.

Using multiple starting configurations is one attempt to overcome the absence of configurational sampling in this QM/MM method, albeit not the most statistically relevant sampling scheme given the low number of initial configurations. Similarly, to ensure the stability of the reaction energy of states classified as initial, final or intermediates states, the size of the QM region was varied. In earlier generations of QM/MM ideology, treatment of the interface between QM/MM regions was thought to introduce inconsistencies in the QM/MM reaction energies described as artifactual effects introduced when one region polarizes the other.[158, 159] As such, this protocol is capable of spotlighting any inconsistencies introduced by interface artifacts.

The starting coordinates for the QM/MM calculations selected from the molecular dynamics (MD) trajectories of our previous study represented key hydrogen bonding networks relevant to proton transfer in DHOD. The selected coordinates are defined as the carboxylate path and dual path. Solvent molecules beyond 25 Å of the substrate were removed, and hydrogens were added using the Maestro graphical user interface. Pure molecular mechanics energy minimizations were performed for 10000 steps with the conjugate gradient algorithm. Coordinates of the active site components including the substrate, cofactor, water molecule, SER215 and THR218 were fixed to maintain the hydrogen bonding network for the carboxylate and dual paths during this minimization.

From the energy minimized structures, two QM/MM models were created to mimic the reactant state for the forward reaction. In the reactant state model, the proton and hydride are bound to the C5 and C6 carbon atoms of substrate respectively. In the product state model, the proton is bound to the carboxylate group of the substrate, thereby neutralizing the molecule, and the hydride is bound to the N5 position of the FMN

cofactor. The smallest QM region included the substrate, water, SER215 sidechain and THR218 sidechain (44 total atoms). The size of the QM region was systematically increased by adding the following components to the smallest QM region: protein backbone from SER215 to ASN219 (76 total atoms), FMN cofactor (94 total atoms), and both backbone and cofactor simultaneously (126 total atoms).

From the reactant state models, three sets of constrained QM/MM optimizations were performed to represent the PT1HT2, HT1PT2 and PTHT mechanisms. For the proton transfer, an intermolecular bond constraint between the proton and the acceptor oxygen was employed to maintain a distance of 0.98 Å. For the hydride transfer, an intermolecular bond constraint between the hydride and the acceptor nitrogen was used, maintaining a distance of 1.023 Å. Essentially, at the active site there was a single bond constraint for the two sequential mechanisms and two bond constraints for the concerted mechanism. After energy convergence with the constrained active site, the bond constraints were removed and the system was re-optimized to produce intermediates for the PT1HT2 and HT1PT2 cases and a final product state for the PTHT case. The PT1 intermediate refers to the structure where the proton has been transferred to the acceptor but the hydride still remains on the donor. The HT1 intermediate refers to the structure where the hydride has been transferred to the acceptor but the proton is still bound to the donor. Subsequently, a second constrained optimization was performed from the unconstrained intermediate states to the final product state. States representing the second step in the sequential proton and hydride transfer process are defined as HT2 and PT2 corresponding to the PT1 and HT1 cases, respectively, as shown in Figure 5-1 . In

this protocol, energies obtained after unconstrained optimizations are the values used to calculate the reaction energy with respect to the reactant state.

The study investigated reaction minima prior to transition states to get an estimate of the time needed to calculate these stationary points. The length of time needed for transition state search calculations are much longer but can be reduced if reactant and product minima already exist and are used as TS search inputs where the algorithm can find a saddle point by interpolating a reaction path between the reactant and product starting configurations.

To quantify the catalytic effect of the enzyme, usually a reference reaction is defined. In some instances, this reference reaction has been defined by the reaction in water, but in this study, the reference reaction is that of the gas phase reaction so that the contribution of the protein and solvent environment can be isolated. Gas phase electronic structure calculations were performed on the isolated substrate and isolated cofactor molecules, with the assumption that the proton relay terminates on the carboxylate. The Jaguar[135-139] program was used for the calculations employing density functional theory with the B3LYP functional and a 6-31G** basis set.[135-139] The charges on the substrate and cofactor for the reactant state are -1 and -2, respectively, which come from the carboxylate group on the substrate and the two negatively charged oxygen atoms on the phosphate tail of the FMN cofactor. Likewise, the charges on the substrate and cofactor for the product state are 0 and -3, respectively. Reaction energies (ΔE) were calculated as:

$$\Delta E_{reaction} = E_{(ORO+reducedFMN)} - E_{(DHO+oxidizedFMN)}$$

where $E_{(ORO+reducedFMN)}$ is the sum of the energy of the isolated ORO and reduced FMN (product) and $E_{(DHO+oxidizedFMN)}$ is the sum of the energy of the isolated DHO and oxidized FMN (reactant).

5.3 Results

Reaction energies were obtained for the sequential and concerted mechanisms. The plot in figure 5-2 illustrates the first sequential case where the proton is transferred before the hydride. For this mechanism, the dual path (DP) starting coordinates were used to compare the effect of changing the size of the quantum region from smallest to largest. The energy differences between the reactant and intermediate states obtained were similar for all sizes of the QM region. However, the QM regions that included the FMN cofactor initially gave lower reaction energies by 5 kcal/mol, which may indicate that it would not be sufficient to model the proton transfer without the hydride transfer components, but by including the hydride transfer components better energy estimates could be obtained. To test whether including FMN in QM region consistently resulted in lower reaction energies, the QM3 calculations were repeated.

As shown in the figure 5-2 , the dual path structure with the QM3 active site fits the range previously seen for the small QM regions, thus the size of the QM region does not sufficiently impact the calculations. More importantly, the reaction energies for all sizes of the QM region ranged from 25-33 kcal/mol, which implies that the transition state barrier between these states would be even higher and thereby unlikely to be relevant to the reaction catalyzed by this enzyme.

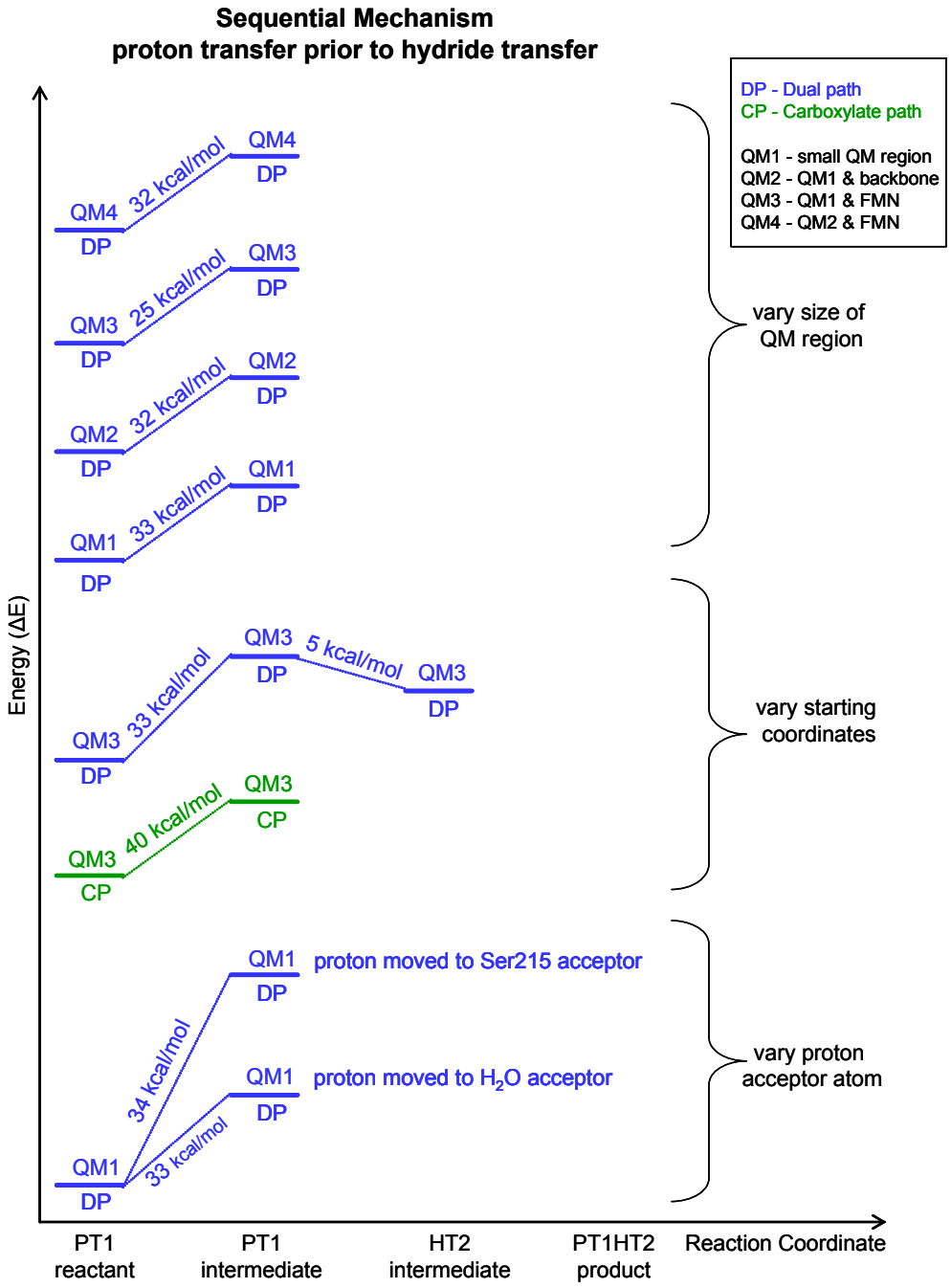

Figure 5-2: Schematic of QM/MM energy for the sequential mechanism where the proton transfers prior to the hydride. Three sets of data are shown illustrating the effect of varying size of the QM region, starting coordinates and proton acceptor atom. Energies are taken relative to the PT1 reactant state and are not drawn to scale.

The dual path initial coordinates were used to estimate the effect of changing the acceptor atom for proton transfer. The hypothesis for this mechanism states that the proton proceeds through a relay pathway but it is not clear which atom functions as the agent to abstract the proton. The oxygens on Ser215, water and the carboxylate of the substrate were all tested as likely candidates to abstract the proton. The reaction energies for the candidates ranged from 33-34 kcal/mol for the Ser215 and water acceptors implying that they are both equally likely to perform this function for the enzyme. However, for each of these choices of acceptor atoms, the largest contributor to the energy difference was the QM region. Carboxylate path initial coordinates were also used to obtain reaction energies for this mechanism. Figure 5-2 shows that similarly high relative estimates were obtained, implying that the starting coordinates may be relatively similar in both the carboxylate and dual path configurations.

Figure 5-3 illustrates the second sequential case where the proton is transferred after the hydride. For this mechanism, the QM region containing the substrate DHO, FMN cofactor, water, and sidechains of Ser215 and Thr218 (QM3) was used to compare how the reaction energy was impacted by variability of the initial coordinates. The results from the carboxylate path indicate high reaction energies on the order of 51 kcal/mol for the hydride transfer step and an additional 15 kcal/mol for the proton transfer step. The 51 kcal/mol decomposes to 45.6 kcal/mol from the QM contribution and 5.81 kcal/mol from the MM contribution. Starting with the dual path coordinates, the first step was 49 kcal/mol and the second step even higher at 11 kcal/mol. Here, the 49 kcal/mol decomposes to a 52.8 kcal/mol contribution from the QM region and a -3.644 kcal/mol contribution from the MM region. The consistent feature of these results is that

from both sets of starting coordinates, the primary source of the endothermicity comes from the QM region which is similar to the observation from the previous sequential mechanism. Given the similarity in the energetic trends between both sequential mechanisms, varying the size of the QM region was not pursued for the second sequential mechanism.

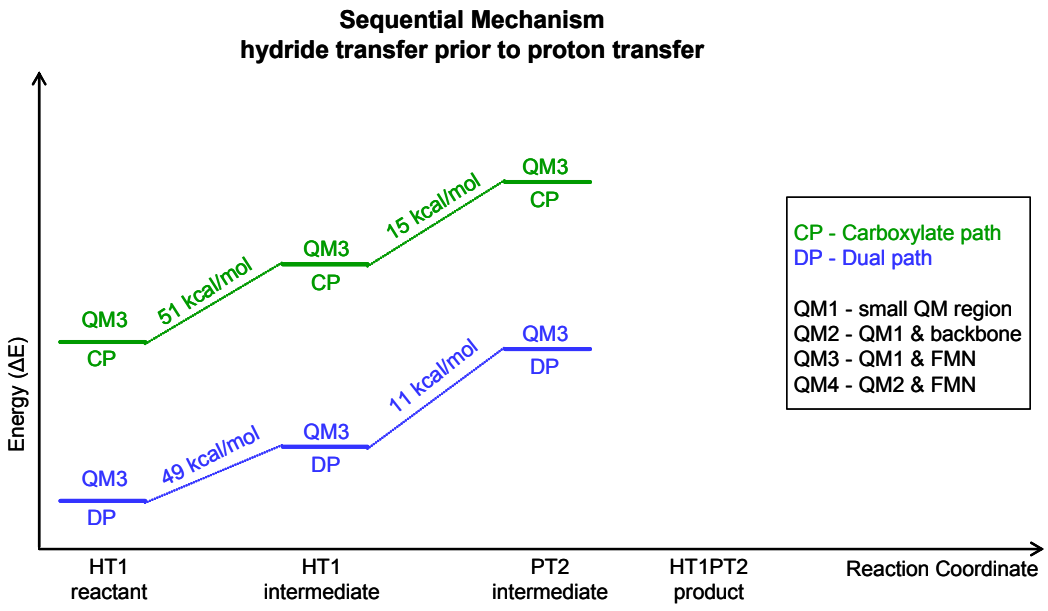

Figure 5-3: Schematic of the QM/MM reaction energy profiles for the sequential mechanism (HT1PT2). Energies are taken relative to the HT1 reactant state and are not drawn to scale. The QM region was QM3 and the focus is on the influence of varying the starting coordinates CP and DP.

Figure 5-4 illustrates the concerted case where the proton and hydride are transferred simultaneously (PTHT). For this mechanism, the QM region size QM3 was chosen to provide comparative estimates to the two sequential mechanisms previously calculated. Here, the starting coordinates were varied to see how initial structures impacted the reaction energy of the concerted mechanism. The reaction energy from the

crystal structure coordinates, carboxylate path coordinates, and dual path coordinates ranged between 50-54 kcal/mol, where the QM energy contributes approximately 56.3 kcal/mol and the MM energy contributes approximately -1.89 kcal/mol. This energetic trend is similar to both of the previously described mechanisms in that the primary source of endothermicity comes from the QM contribution to the overall QM/MM energy. Again, the consistency between the reaction energies from various starting coordinates reflects a general similarity in all of the starting structures. Additionally, the high endothermic energies imply that this mechanism is virtually impossible in the enzyme.

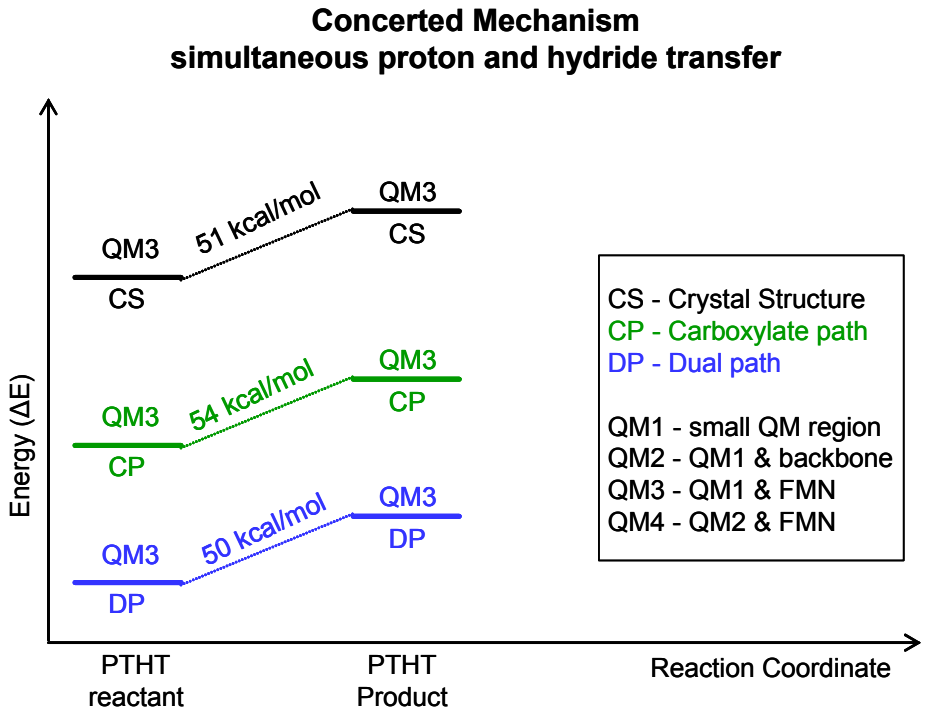

Figure 5-4: Schematic of QM/MM reaction energy comparison for the concerted mechanism starting from three different initial coordinates. Energies are taken to be relative to the PTHT reactant and are not drawn to scale.

The general energetic trends observed from the QM/MM results indicate that an unfortunate set of conditions exist for this DHOD system. One possible technique that can be used to isolate the source of these problematic conditions is to compare the system to a reference reaction. Calculations were also performed on the *E. coli* DHOD structure which is a similar class II structure.[71]

In this study, the gas phase reaction energies are calculated as an estimate of the reference reaction. The tail of the FMN cofactor is truncated with a methyl group, bringing the overall charge of the system to -1. Figure 5-5 shows the results from the gas phase calculations.

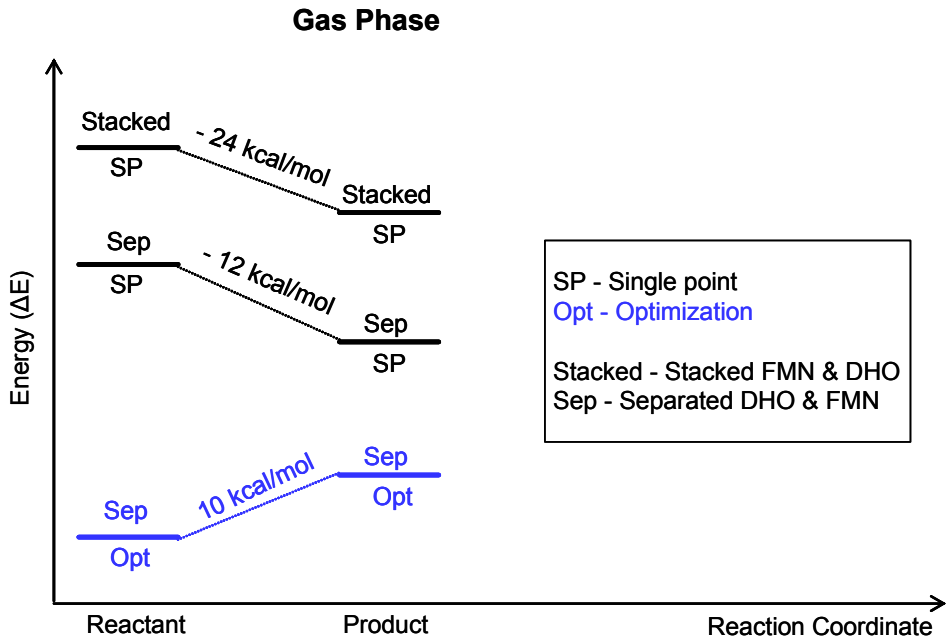

Figure 5-5: Single point gas phase energies of the stacked species show an exothermicity of 24 kcal/mol. Single point calculations of the individual components show and exothermicity of 12 kcal/mol. Optimizations of the individual components show an endothermicity of 10 kcal/mol. All energies shown are schematic representations and are not drawn to scale.

In the first gas phase calculation, the QM region was taken from the fully optimized QM/MM structure of the CP model, and a single point QM calculation was performed on only the stacked substrate and cofactor configuration at the same level of theory as the QM/MM calculation (B3LYP/6-31G*). This single point calculation reveals an exothermicity of -24 kcal/mol for the stacked substrate and cofactor. To quantify how much the stacking interaction between the substrate and cofactor contributes the single point energy, a single point calculation was performed on the independent, separated substrate and cofactor. The result of -12 kcal/mol upheld the exothermic trend observed for the stacked configuration and may imply that either stacking the species or better charge delocalization, doubles the energy difference. This information becomes relevant when the species are optimized in the gas phase. Optimizing the substrate and cofactor in the gas phase destroys the stacking configuration which is observed at the active site of the crystal structure. In fact, the two molecules fly apart in an attempt to adopt the least sterically hindered configuration, as was observed in our earlier gas phase electronic structure calculations. The result of optimizing the independent species of the substrate and cofactor gave a reaction energy of 10 kcal/mol, which completely reverses the exothermic trend of the QM/MM optimized QM components. It should be noted that this reaction energy of 10 kcal/mol reflects a higher energy conformation of the substrate where the proton on the carboxylate is closest to the ring as in Figure 5-6a.

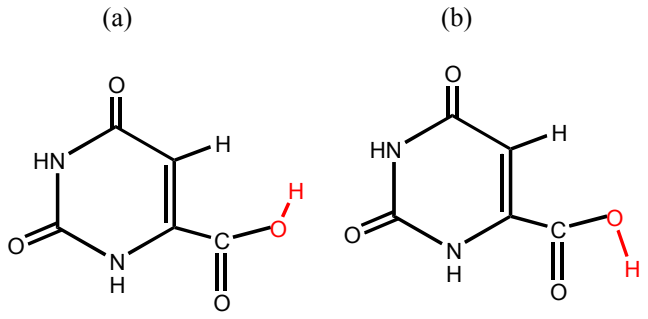

Figure 5-6: (a) The higher energy structure of DHO substrate which is immediately formed after proton relay from the C5 position through a water molecule, then on to the carboxylate. (b) The lower energy structure of the DHO substrate.

A lower energy conformation was calculated with the proton being closest to the other oxygen of the carboxylate as in Figure 5-6b. The reaction energy for this conformation is 3.47 kcal/mol, but this is not the state formed immediately after proton relay. The enzyme and solvent would have to adjust such that it leaves room for this proton to reorient itself to the lower energy substrate conformation.

The exothermicity of the gas phase results in comparison to the high endothermicity in the QM/MM results reflects the change of the QM charge distribution in the presence of the MM point charges. Recall that in the QM/MM reactant state, the total negative charge of -3 is delocalized with -2 on the FMN cofactor phosphate tail and -1 on the substrate carboxylate tail. In the QM/MM product state the charge distribution localizes the negative charge of -3 on the FMN cofactor, which may contribute to the observed penalty to the QM energy in the environment of the MM point charges. Unlike the gas phase example where the FMN phosphate tail has been truncated, the -1 charge on the substrate carboxylate in the reactant state is delocalized on the larger isoalloxazine

ring of the FMN and may provide the stability of the product states seen in the gas phase results.

In looking at the energy terms contributing to the total QM/MM energy, the largest source of endothermicity is from the QM* energy. The QM* energy can be further analyzed where the electronic energy difference is exothermic in most cases and the nuclear repulsion energy is strongly endothermic. The electronic energy relates to the electron density, coulomb and exchange of the QM region as indicated by Eq. 3.2 . The nuclear repulsion energy relates to the charges on the nuclei of the QM region and points to the geometries as a source of the irregularities. By definition of the total QM/MM energy, the contribution of the MM point charges on the QM region is embedded in the QM* energy terms. Thus, as a means of separating the pure QM energy from the field of MM point charges, additional single point QM calculations were performed on the active site of the QM optimized geometries. Table 5-1 shows the results for the three mechanisms.

In most cases, the relative energy of the pure QM region is much less endothermic, and in one case, exothermic. This may indicate that the field of MM point charges is strongly polarizing the QM region resulting in these non-reactive QM/MM energies. Additional single point calculations were performed on the QM/MM optimized active site where the phosphate tail was removed from the FMN. The majority of these calculations gave exothermic relative energies, implying that the electrostatic environment of the MM region near the phosphate tail may be affecting the stabilization of the product geometries.

Table 5-1: Energy comparison of QM* and pure QM energy. Single point energy taken at the optimized QM/MM geometry. QM* is the QM component of the QM/MM energy with the influence of the MM point charges.

Mechanism	**Total QM/MM (ΔE) (kcal/mol)**	**QM* (ΔE) (kcal/mol)**	**QM (sub+cof) (ΔE) (kcal/mol)**	**QM Single point (ΔE) (kcal/mol)**
CP_PT1	46.58	37.2	(QM1)	27.42
CP_HT2	--	--	--	--
CP_HT1	51.4	45.6	-24.6	-24.86
CP_PT2	15.5	9.08	6.8	-2.74
DP_Concerted	50.15	49.9	44.2	50.44
CP_Concerted	54.4	56.3	-7.18	-0.3

The comparison of the QM* relative energy to the pure QM relative energy raises two additional hypotheses. (1) Perhaps the MM charges near the active site need to be modified to reduce the observed polarization effect. (2) Perhaps the MM environment is in the incorrect reactive state, and the MM environment does not have the opportunity to significantly adjust to the product structure. For both hypotheses', additional sampling may be needed to see a significant change in the current energetic trends.

5.4 Benchmark on Dihydrofolate Reductase

Analyzing the reaction in DHFR with QSites' QM/MM approach provided the first benchmark of this protocol on a hydride transfer system as shown in Figure 5-7. DHFR is a well-studied system[160, 161] and has an experimentally measured reaction energy of -4 kcal/mol.[162] Utilizing the same protocol as the DHOD system, QM/MM calculations on the hydride transfer in DHFR yielded a reaction energy of -18 kcal/mol. Upon closer inspection of the individual components contributing to the overall QM/MM energy in DHFR, the MM energy terms contribute -11 kcal/mol (-5.44 kcal/mol from Lennard Jones energy and -4.74 kcal/mol from electrostatic energy) and the QM energy terms contribute the remaining -7 kcal/mol. An important realization gained from this DHFR result is that since the primary source of energy variability is caused by changes in the MM region, mitigating this disparity of the QM/MM result with respect to the experimental value is a straightforward technical manipulation of certain parameters in the MM terms. The QM contribution to the energy difference of -7 kcal/mol is in good agreement with the experimental energy difference of -4 kcal/mol and shows great support for the accuracy of the method when it is coupled with this protocol. Another important realization from this result is that the method was able to reproduce the exothermic characteristic of the system, unlike the observation in the DHOD system. In a good number of protein systems, the exothermicity shows that the reaction is thermodynamically favorable where the products are lower in energy than the reactants. This idea is the thrust behind proposals emphasizing that the main catalytic effect of the enzyme is reactant state destabilization.

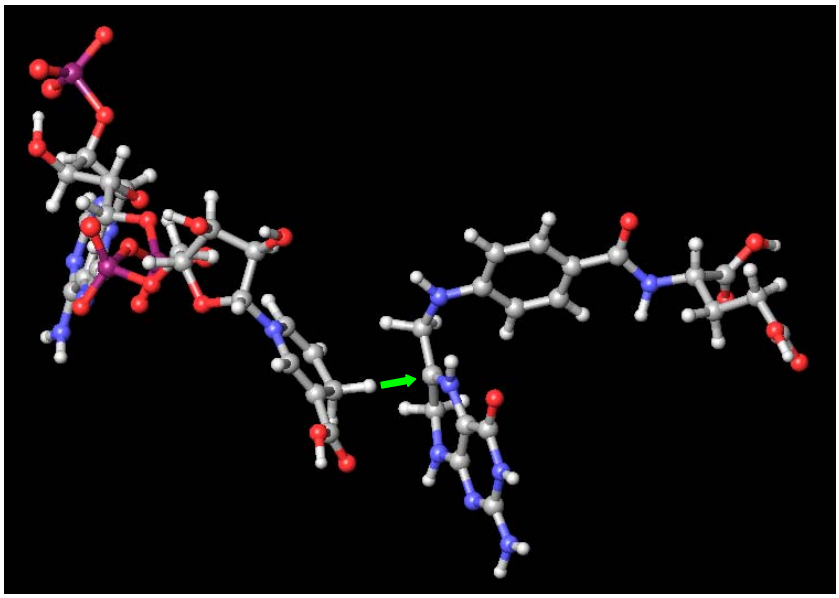

Figure 5-7: Schematic representation active site model used in QM/MM calculations on Dihydrofolate Reductase (DHFR). Green arrow indicates hydride transfer from nicotinamine adenine dinucleotide phosphate (NADPH) to folate. The representation assumes that proton transfer occurred in a previous step.

To reduce the energy of the MM region, thereby bringing the overall QM/MM reaction energy to within good agreement of the experimental value, the constrained and unconstrained QM/MM protocol should be repeated, moving the system back and forth between reactant and product states until the MM energy terms stop changing between the two states. Additionally, if the disparity persists, altering the cut-off parameter for the MM region should be considered. The basis for altering the cut-off parameter comes from benchmarks used to test this QM/MM methodology. The tests indicate that in some systems smaller cut-off distances can cause jumps in the MM energy of greater than 20 kcal/mol, and larger cut-off distances can cause energy jumps of less than 10 kcal/mol. Thus it is prudent to perform a series of tests, for each new system, using various cut-off

distances to determine the exact distance that invokes the least amount of MM energy fluctuation. For these studies, a cut-off distance of 50 Å was used.

5.5 Conclusions

In this study, the focus was on reaction energies as an approach to identifying essential features of the reaction that may become key points of consideration for the next steps in searching for transition states that lead to activation energies. The high endothermic reaction energies from all of the DHOD cases inspired a litany of tests aimed at isolating the sources of error. One hypothesis for the high energy estimates is that the ab initio QM/MM protocol used on this system may be a poor choice when combined with certain crystal structures or for certain systems where entropic effects play a large role in catalysis. Two main reasons form the basis of this hypothesis. The first stems from the fact that this specific protocol has not been tested on more well understood enzymes, and the second stems from the exclusion of entropy contributions to the energy inherent to this ab initio QM/MM method. In this QM/MM methodology, there is an iterative procedure in which the energy of the MM region is minimized with a conjugate gradient or steepest decent algorithm after each QM optimization step. This scheme proceeds self consistently until convergence is achieved. As such, it is likely that the system can become trapped in a local minimum on the potential energy surface, resulting in the unfortunate high energies observed for the DHOD system.

Given the absence of sampling of configurational space, the energies gained are the enthalpic contributions, and the entropic effects are ignored. There are several

examples of protein systems where the primary contributions to the reactive event are governed by the change in enthalpy, and the entropic effects are believed to be small.[163-165] This premise can usually be verified by comparison to experimental data.

The DHFR study, chosen to test the QM/MM protocol, delivered results that illustrate the ability of this method to come within good agreement of the experimental reaction energies for a hydride transfer system. This finding is recognizably not transferable to all enzyme systems, but in light of these DHFR results, more critical attention was turned to the starting structures for DHOD in an effort to sort out the reason for the unusually high reaction energies.

In assessing how the size of the QM region contributes to the reaction energy, the energy profile for different sized QM regions were generated. There was no noticeable difference in the energy for various sizes of QM region, so the recommended size contains around 44 atoms, equivalent to QM3. This should provide reasonable energy estimates in a relatively short amount of computational time.

Moving the proton from the donor to the oxygens of Ser215 and water were both calculated with this method. Even though the Ser215 and water reported similar reaction energies, the high pKa of serine (~13) makes it a weak base and thus a poor candidate for proton abstraction. Based on the structural integrity of the hydrogen bonded networks in all of these QM/MM calculations, it is likely to conceive that Ser215 is not directly responsible for proton abstraction but that serine's primary role is to maintain a hydrogen bond to the water which serves to abstract the proton.

Changing the protonation states of residues near the active site was also considered as a way to lower the energy difference. The nearby Lys100 residue was

neutralized, and the entire profile was regenerated. The resulting reaction energy was endothermic at 25 kcal/mol. While being less endothermic by around 5 kcal/mol compared to the protonated (+1) Lys, neutralizing the lysine may need to be included in a recipe with other protonation state manipulations and should not be considered independently. As such, the N1 position of the FMN cofactor was also protonated. Similar to the Lys case, this alteration reduced the endothermicity by around 3 kcal/mol. However, when combined with neutralizing the Lys, the overall energy difference reduced the endothermicity by 5 kcal/mol, which is clearly not an additive effect of the two alterations. These modest changes in the energy do not show a significant departure from the original result and may be within the same degree of error.

Chapter 6

Conclusions

6.1 General Conclusions

One of the main areas of focus in the biochemistry community is to attempt to resolve the issue of enzyme catalysis. How enzymes are able to catalyze reactive events as efficiently as they do remains a subject of debate. Kinetics and structural analysis of enzymes offer relevant information about reaction rates and the identity of residues directly involved in catalysis, but these methods are unable to pinpoint the exact source of catalysis. Many researchers have come to the forefront with theories to complement experimental studies. These theories promise to dissect the energy contributions to the rate in an attempt to pinpoint the most prominent effect that contributes to the experimentally observed reaction rates. The usefulness of being able to quantify specific contributions to the rate is well understood by the drug design community. Quantum mechanics, molecular dynamics simulations and mixed quantum/classical methods are currently the popular approaches to answer the questions about which factors contribute to enzyme catalysis. However, the issue of accuracy of these methods in terms of representing the entire condensed phase enzyme system leaves plenty of room for many combinations of these methods to propose adequate explanations of catalysis. In the current state of this field, there is still no clearly defined answer to what contributes to

catalysis, but a number of competing proposals can all be deemed as plausible explanations of catalysis.

This study took a systematic approach to investigating the proton and hydride transfer mechanism in DHOD. Many combinations of computational methods were used to try to explain three main factors in DHOD catalysis: concerted or sequential nature of the mechanism, role of the enzyme's environment in the mechanism, and changes in the mechanism or environment when mutations are introduced.

To tackle the sequential versus concerted mechanism, both QM and mixed QM/MM methods were used to build reaction profiles for each mechanism. The objective was to be able to decipher the most likely reaction mechanism by ruling out energy barriers that were too large. Unfortunately, in these calculations all of the results were too similar to be able to distinguish the most relevant mechanism. However, a great deal was learned about the computational methods and how certain advantages, limitations and approximations may be utilized for the greater understanding of enzyme catalysis. Additionally, much understanding was gained in terms of how experimental data are interpreted and which measurements are comparable to computationally calculated results.

Along the route of elucidating the mechanism in DHOD, the role of the immediate active site environment revealed the impact of certain hydrogen bonding networks on catalysis. This part of the study utilized molecular dynamics simulations to probe different networks that may facilitate proton transfer. With all of the computational methods used in the study, the issue of how mutations impact catalysis was addressed. The study demonstrated that mutations had structural impact by

lengthening donor-acceptor distances and disrupting key hydrogen bonding networks. While the energetic impact could not be quantified from the results, the qualitative effect paralleled those of the experimental results.

DHOD is a system rich with possibilities for future study, and the experimental contributions are still in the very early stages in attempting to answer questions about catalysis. One consideration would be to obtain reaction energy estimates for all known crystal structures. This will perhaps illustrate problematic starting coordinates at the outset before further refinement of the structures is pursued. Furthermore, ideas about solvent insertion, side chain sampling and enthalpy-entropy compensation should be investigated.

6.2 Crystal Structure Considerations

An essential ingredient to studying protein systems with computational methods is the availability of crystal structures. For DHOD, crystal structures exist for the bacterial (class I) and mammalian (class II) enzymes.[67, 68, 166] This study focuses on the class II enzyme because of the size and simplicity of the crystal structure compared to the class I DHOD system. The assumption was that a monomeric, globular enzyme with only one set of active site components would be sufficient to provide good initial estimates of the reaction and activation energies for a system where there is little to no other computational results published. In contrast, the class I DHOD enzymes range from the homodimeric to heterotetrameric structures with multiple active site regions. The larger

number of degrees of freedom can complicate an effort aimed at isolating a finite set of contributions that influence catalysis.

For the crystal structures that are available for human and rat class II DHOD enzymes, it should be noted that all of these structures have an inhibitor bound in the crystallized molecule.[167] The inherent definition of an inhibitor is to interfere with the chemical action of the system, leading to an obvious disadvantage at the onset of these computational studies. Essentially, all crystal structures are manipulated to stop the reaction so that the crystals can be formed, but some manipulation procedures are more intrusive than others. For example, sometimes a structure is given with an analogue of the substrate bound to the crystal. The molecule is still reactive albeit at a much slower rate to sufficiently allow the crystals to form. However, in the case of DHOD with a molecule of the Brequinar inhibitor bound in the crystal[168-171], it is unclear how much the structure was adversely affected. In principle, minimizing the energy and properly equilibrating the system with molecular dynamics for long time periods (10-20 ns) should put the system in a more favorable reactive state.

More importantly, when studying systems with questionable starting coordinates, high emphasis should be placed on ensuring proper sampling of phase space before any reasonable conclusions can be drawn about enzyme catalysis. A far reaching example of how reasonable starting coordinates can lead to reasonable energy estimates can be seen by comparing the QM/MM results between DHFR and DHOD. The crystal structure used in the DHFR test (1RX2)[172] does not have any inhibitors bound to the structure, unlike the DHOD crystal structure (1D3G),[168] where the Brequinar inhibitor is bound to the system. In all of these DHOD studies, this inhibitor is removed before commencing

any calculations. Published findings about the Brequinar inhibitor reflect a fair amount of uncertainty about the impact of the inhibitor on the molecular level,[173] and this uncertainty is inherently transferred to these QM/MM studies. This can be one reason that the high relative energies observed in these QM/MM calculations may indeed be good estimates for the inhibited starting structures.[174]

6.3 Solvent Insertion Considerations

A second consideration, after good initial coordinates are secured, would be to pursue solvent insertion schemes which monitor the structural behavior as solvent is added to hydrophilic interior parts of the protein. The idea here is that if the solvent assisted hydrogen bonding networks observed in the molecular dynamics study are of any merit, then longer solvent assisted chains that are relevant for DHOD catalysis may exist. This idea is not completely foreign, as it has been studied extensively in the carbonic anhydrase II protein. Solvent insertion algorithms have been designed to execute this work[175]. Additionally, the hydrogen bonding analysis code in appendix C.1 can be readily extended to calculate probabilities of the existence of longer hydrogen bonding networks. Measuring long hydrogen bonding networks from the active site of DHOD to the bulk may prove to be an interesting addition to the knowledge on DHOD.

6.4 Side Chain Sampling Considerations

A third consideration is that more robust sampling schemes can be included in the QM/MM protocol described herein to generate initial configurations for the QM/MM calculations. A number of Loop prediction[176-178] and emerging Monte Carlo[179] based side-chain sampling schemes are coming to the forefront to make this possible. The idea is that while a large rotamer library is scanned, the energy of the system is analyzed until structures meeting a predefined set of conditions is obtained. These conditions may include the existence of certain hydrogen bonding networks or certain electrostatic environments. The most relevant feature of this idea is that in the absence of comparable experimental data, a reasonable amount of configurations can be sampled with the same QM/MM protocol described herein.

Bibliography

[1] A. Warshel, M. Levitt, Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme, J Mol Biol 103 (1976) 227-249.

[2] T.Z. Mordasini, W. Thiel, Combined quantum mechanical and molecular mechanical approaches, Chimia 52 (1998) 288-291.

[3] P. Sherwood, A.H. de Vries, M.F. Guest, G. Schreckenbach, C.R.A. Catlow, S.A. French, A.A. Sokol, S.T. Bromley, W. Thiel, A.J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S.C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjovoll, A. Fahmi, A. Schafer, C. Lennartz, QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis, J Mol Struc-Theochem 632 (2003) 1-28.

[4] R.A. Marcus, Schrödinger Equation for Strongly Interacting Electron-Transfer Systems, J Phys Chem-Us 96 (1992) 1753-1757.

[5] A. Warshel, Computer simulations of enzyme catalysis: Methods, progress, and insights, Annu Rev Bioph Biom 32 (2003) 425-443.

[6] A. Warshel, P.K. Sharma, M. Kato, Y. Xiang, H.B. Liu, M.H.M. Olsson, Electrostatic basis for enzyme catalysis, Chem Rev 106 (2006) 3210-3235.

[7] A. Yadav, R.M. Jackson, J.J. Holbrook, A. Warshel, Role of Solvent Reorganization Energies in the Catalytic Activity of Enzymes, J Am Chem Soc 113 (1991) 4800-4805.

[8] J.A. Gascon, S.S.F. Leung, E.R. Batista, V.S. Batista, A self-consistent space-domain decomposition method for QM/MM computations of protein electrostatic potentials, J Chem Theory Comput 2 (2006) 175-186.

[9] M.P. Repasky, C.R.W. Guimaraes, J. Chandrasekhar, J. Tirado-Rives, W.L. Jorgensen, Investigation of solvent effects for the Claisen rearrangement of chorismate to prephenate: Mechanistic interpretation via near attack conformations, J Am Chem Soc 125 (2003) 6663-6672.

[10] L.Y. Li, C.J. Falzone, P.E. Wright, S.J. Benkovic, Functional-Role of a Mobile Loop of *Escherichia coli* Dihydrofolate-Reductase in Transition-State Stabilization, Biochemistry-Us 31 (1992) 7826-7833.

[11] P.K. Agarwal, S.R. Billeter, P.T.R. Rajagopalan, S.J. Benkovic, S. Hammes-Schiffer, Network of coupled promoting motions in enzyme catalysis, P Natl Acad Sci USA 99 (2002) 2794-2799.

[12] N.M. Antikainen, R.D. Smiley, S.J. Benkovic, G.G. Hammes, Conformation coupled enzyme catalysis: Single-molecule and transient kinetics investigation of dihydrofolate reductase, Biochemistry-Us 44 (2005) 16835-16843.

[13] C.E. Cameron, S.J. Benkovic, Evidence for a functional role of the dynamics of glycine-121 of *Escherichia coli* dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant, Biochemistry-Us 36 (1997) 15792-15800.

[14] J.A. Gascon, V.S. Batista, QM/MM study of energy storage and molecular rearrangements due to the primary event in vision, Biophys J 87 (2004) 2931-2941.

[15] M.S. Gordon, M.A. Freitag, P. Bandyopadhyay, J.H. Jensen, V. Kairys, W.J. Stevens, The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry, J Phys Chem A 105 (2001) 293-307.

[16] H. Li, C.S. Pomelli, J.H. Jensen, Continuum solvation of large molecules described by QM/MM: a semi-iterative implementation of the PCM/EFP interface, Theor Chem Acc 109 (2003) 71-84.

[17] P.K. Agarwal, S.P. Webb, S. Hammes-Schiffer, Computational studies of the mechanism for proton and hydride transfer in liver alcohol dehydrogenase, J Am Chem Soc 122 (2000) 4803-4812.

[18] S.R. Billeter, S.P. Webb, T. Iordanov, P.K. Agarwal, S. Hammes-Schiffer, Hybrid approach for including electronic and nuclear quantum effects in molecular dynamics simulations of hydrogen transfer reactions in enzymes, J Chem Phys 114 (2001) 6925-6936.

[19] Y.T. Chang, C. Minichino, W.H. Miller, Classical Trajectory Studies of the Molecular Dissociation Dynamics of Formaldehyde - H2co-]H-2+Co, J Chem Phys 96 (1992) 4341-4355.

[20] M. Garcia-Viloca, C. Alhambra, D.G. Truhlar, J.L. Gao, Hydride transfer catalyzed by xylose isomerase: Mechanism and quantum effects, J Comput Chem 24 (2003) 177-190.

[21] E. Rosta, M. Klahn, A. Warshel, Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions, J Phys Chem B 110 (2006) 2934-2941.

[22] U.W. Schmitt, G.A. Voth, Multistate empirical valence bond model for proton transport in water, J Phys Chem B 102 (1998) 5547-5551.

[23] U.W. Schmitt, G.A. Voth, Quantum properties of the excess proton in liquid water, Israel J Chem 39 (1999) 483-492.

[24] U.W. Schmitt, G.A. Voth, The computer simulation of proton transport in water, J Chem Phys 111 (1999) 9361-9381.

[25] S.P. Webb, P.K. Agarwal, S. Hammes-Schiffer, Combining electronic structure methods with the calculation of hydrogen vibrational wavefunctions: Application to hydride transfer in liver alcohol dehydrogenase, J Phys Chem B 104 (2000) 8884-8894.

[26] Q. Cui, M. Elstner, M. Karplus, A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH), J Phys Chem B 106 (2002) 2721-2740.

[27] K.P. Eurenius, D.C. Chatfield, B.R. Brooks, M. Hodoscek, Enzyme mechanisms with hybrid quantum and molecular mechanical potentials .1. Theoretical considerations, Int J Quantum Chem 60 (1996) 1189-1200.

[28] J.A. Gascon, E.M. Sproviero, V.S. Batista, QM/MM study of the NMR spectroscopy of the retinyl chromophore in visual rhodopsin, J Chem Theory Comput 1 (2005) 674-685.

[29] J.A. Gascon, E.M. Sproviero, V.S. Batista, Computational studies of the primary phototransduction event in visual rhodopsin, Accounts Chem Res 39 (2006) 184-193.

[30] H. Lin, J.C. Schoneboom, S. Cohen, S. Shaik, W. Thiel, QM/MM study of the product-enzyme complex in P450(cam) catalysis, J Phys Chem B 108 (2004) 10083-10088.

[31] P.D. Lyne, M. Hodoscek, M. Karplus, A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region, J Phys Chem A 103 (1999) 3462-3471.

[32] R.B. Murphy, D.M. Philipp, R.A. Friesner, A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments, J Comput Chem 21 (2000) 1442-1457.

[33] J.C. Schoneboom, S. Cohen, H. Lin, S. Shaik, W. Thiel, Quantum mechanical/molecular mechanical investigation of the mechanism of C-H

hydroxylation of camphor by cytochrome P450(cam): Theory supports a two-state rebound mechanism, J Am Chem Soc 126 (2004) 4017-4034.

[34] J.C. Schoneboom, F. Neese, W. Thiel, Toward identification of the Compound I reactive intermediate in cytochrome P450 chemistry: A QM/MM study of its EPR and Mossbauer parameters, J Am Chem Soc 127 (2005) 5840-5853.

[35] J.C. Schoneboom, W. Thiel, The resting state of p450(cam): A QM/MM study, J Phys Chem B 108 (2004) 7468-7478.

[36] H.M. Senn, D. O'Hagan, W. Thiel, Insight into enzymatic C-F bond formation from QM and QM/MM calculations, J Am Chem Soc 127 (2005) 13643-13655.

[37] J.J. Zheng, D.Q. Wang, W. Thiel, S. Shaik, QM/MM study of mechanisms for compound I formation in the catalytic cycle of cytochrome P450cam, J Am Chem Soc 128 (2006) 13204-13215.

[38] S.R. Billeter, C.F.W. Hanser, T.Z. Mordasini, M. Scholten, W. Thiel, W.F. van Gunsteren, Molecular dynamics study of oxygenation reactions catalysed by the enzyme p-hydroxybenzoate hydroxylase, Phys Chem Chem Phys 3 (2001) 688-695.

[39] S.R. Billeter, A.J. Turner, W. Thiel, Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates, Phys Chem Chem Phys 2 (2000) 2177-2186.

[40] J. Chandrasekhar, S. Shariffskul, W.L. Jorgensen, QM/MM simulations for Diels-Alder reactions in water: Contribution of enhanced hydrogen bonding at the transition state to the solvent effect, J Phys Chem B 106 (2002) 8078-8085.

[41] W. Thiel, Computational methods for large molecules, Theochem-J Mol Struc 398 (1997) 1-6.

[42] J.L. Gao, D.G. Truhlar, Quantum mechanical methods for enzyme kinetics, Annu Rev Phys Chem 53 (2002) 467-505.

[43] D.T. Major, M. Garcia-Viloca, J.L. Gao, Path integral simulations of proton transfer reactions in aqueous solution using combined QM/MM potentials, J Chem Theory Comput 2 (2006) 236-245.

[44] B.Y. Ma, S. Kumar, C.J. Tsai, Z.J. Hu, R. Nussinov, Transition-state ensemble in enzyme catalysis: Possibility, reality, or necessity?, J Theor Biol 203 (2000) 383-397.

[45] J.K. Hwang, A. Warshel, How important are quantum mechanical nuclear motions in enzyme catalysis?, J Am Chem Soc 118 (1996) 11745-11751.

[46] G.Y. Hong, M. Strajbl, T.A. Wesolowski, A. Warshel, Constraining the electron densities in DFT method as an effective way for ab initio studies of metal-catalyzed reactions, J Comput Chem 21 (2000) 1554-1561.

[47] M. Eichinger, P. Tavan, J. Hutter, M. Parrinello, A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields, J Chem Phys 110 (1999) 10452-10467.

[48] H.L. Woodcock, M. Hodoscek, P. Sherwood, Y.S. Lee, H.F. Schaefer, B.R. Brooks, Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase, Theor Chem Acc 109 (2003) 140-148.

[49] P. Rowland, S. Norager, K.F. Jensen, S. Larsen, Crystallization and preliminary X-ray studies of membrane-associated *Escherichia coli* dihydroorotate dehydrogenase, Acta Crystallogr D 56 (2000) 659-661.

[50] P. Rowland, S. Norager, K.F. Jensen, S. Larsen, Structure of dihydroorotate dehydrogenase B: Electron transfer between two flavin groups bridged by an iron-sulphur cluster, Structure 8 (2000) 1227-1238.

[51] P. Rowland, F.S. Nielsen, K.F. Jensen, S. Larsen, The crystal structure of the flavin containing enzyme dihydroorotate dehydrogenase A from *Lactococcus lactis*, Structure 5 (1997) 239-252.

[52] P. Rowland, F.S. Nielsen, K.F. Jensen, S. Larsen, Crystallization and preliminary X-ray diffraction analysis of the heterotetrameric dihydroorotate dehydrogenase B of *Lactococcus lactis*, a flavoprotein enzyme system consisting of two PyrDB subunits and two iron-sulfur cluster containing PyrK subunits, Acta Crystallogr D 53 (1997) 802-804.

[53] P. Rowland, O. Björnberg, F.S. Nielsen, K.F. Jensen, S. Larsen, The crystal structure of *Lactococcus lactis* dihydroorotate dehydrogenase A complexed with the enzyme reaction product throws light on its enzymatic function, Protein Sci 7 (1998) 1269-1279.

[54] S. Norager, S. Arent, O. Björnberg, M. Ottosen, L. Lo Leggio, K.F. Jensen, S. Larsen, *Lactococcus lactis* dihydroorotate dehydrogenase A mutants reveal important facets of the enzymatic function, J Biol Chem 278 (2003) 28812-28822.

[55] A.T. Cordeiro, P.R. Feliciano, M.C. Nonato, Crystallization and preliminary X-ray diffraction analysis of *Leishmania major* dihydroorotate dehydrogenase, Acta Crystallographica Section F-Structural Biology and Crystallization Communications 62 (2006) 1049-1051.

[56] F.S. Nielsen, P.S. Andersen, K.F. Jensen, The B form of dihydroorotate dehydrogenase from *Lactococcus lactis* consists of two different subunits, encoded by the pyrDb and pyrK genes, and contains FMN, FAD, and [FeS] redox centers, J Biol Chem 271 (1996) 29359-29365.

[57] E. Zameitat, W. Knecht, J. Piskur, M. Loffler, Two different dihydroorotate dehydrogenases from yeast *Saccharomyees kluyveri*, Febs Letters 568 (2004) 129-134.

[58] E. Zameitat, Z. Gojkovic, W. Knecht, J. Piskur, M. Loffler, Biochemical characterization of recombinant dihydroorotate dehydrogenase from the opportunistic pathogenic yeast *Candida albicans*, Febs Journal 273 (2006) 3183-3191.

[59] F.S. Nielsen, P. Rowland, S. Larsen, K.F. Jensen, Purification and characterization of dihydroorotate dehydrogenase A from *Lactococcus lactis*, crystallization and preliminary X-ray diffraction studies of the enzyme, Protein Sci 5 (1996) 852-856.

[60] A. Argyrou, M.W. Washabaugh, C.M. Pickart, Dihydroorotate dehydrogenase from *Clostridium oroticum* is a class 1B enzyme and utilizes a concerted mechansim of catalysis, Biochemistry-Us 39 (2000) 10373-10384.

[61] P.S. Andersen, P.J.G. Jansen, K. Hammer, 2 Different Dihydroorotate Dehydrogenases in *Lactococcus lactis*, J Bacteriol 176 (1994) 3975-3982.

[62] T.D. Poulsen, M. Garcia-Viloca, J.L. Gao, D.G. Truhlar, Free energy surface, reaction paths, and kinetic isotope effect of short-chain Acyl-CoA dehydrogenase, J Phys Chem B 107 (2003) 9567-9578.

[63] B.J. Bahnson, J.P. Klinman, Hydrogen Tunneling in Enzyme Catalysis, Method Enzymol 249 (1995) 373-397.

[64] B.J. Bahnson, D.H. Park, K. Kim, B.V. Plapp, J.P. Klinman, Unmasking of Hydrogen Tunneling in the Horse Liver Alcohol-Dehydrogenase Reaction by Site-Directed Mutagenesis, Biochemistry-Us 32 (1993) 5503-5507.

[65] V. Hines, M. Johnston, Mechanistic Studies on the Bovine Liver Mitochondrial Dihydroorotate Dehydrogenase Using Kinetic Deuterium-Isotope Effects, Biochemistry-Us 28 (1989) 1227-1234.

[66] R.A. Pascal, C.T. Walsh, Mechanistic Studies with Deuterated Dihydroorotates on the Dihydroorotate Oxidase from *Crithidia-Fasciculata*, Biochemistry-Us 23 (1984) 2745-2752.

[67] O. Bjornberg, P. Rowland, S. Larsen, K.F. Jensen, Active site of dihydroorotate dehydrogenase A from *Lactococcus lactis* investigated by chemical modification and mutagenesis, Biochemistry 36 (1997) 16197-16205.

[68] O. Bjornberg, A.C. Gruner, P. Roepstorff, K.F. Jensen, The activity of *Escherichia coli* dihydroorotate dehydrogenase is dependent on a conserved loop identified by sequence homology, mutagenesis, and limited proteolysis, Biochemistry 38 (1999) 2899-2908.

[69] W.J. Jiang, G. Locke, M.R. Harpel, R.A. Copeland, J. Marcinkeviciene, Role of Lys100 in human dihydroorotate dehydrogenase: Mutagenesis studies and chemical rescue by external amines, Biochemistry 39 (2000) 7990-7997.

[70] S. Ghisla, V. Massey, Mechanisms of Flavoprotein-Catalyzed Reactions, Eur J Biochem 181 (1989) 1-17.

[71] B.A. Palfey, O. Bjornberg, K.F. Jensen, Insight into the chemistry of flavin reduction and oxidation in *Escherichia coli* dihydroorotate dehydrogenase obtained by rapid reaction studies, Biochemistry-Us 40 (2001) 4381-4390.

[72] J. Shi, B.A. Palfey, J. Dertouzos, D. Steel, A. Gafni, Single molecule study of the reaction kinetics of dihydroorotate dehydrogenase (DHOD), Biophys J 84 (2003) 121a-121a.

[73] J. Shi, B.A. Palfey, J. Dertouzos, K.F. Jensen, A. Gafni, D. Steel, Multiple states of the Tyr318Leu mutant of dihydroorotate dehydrogenase revealed by single-molecule kinetics, J Am Chem Soc 126 (2004) 6914-6922.

[74] J. Shi, B.A. Palfey, J. Dertouzos, A. Gafni, D. Steel, Single molecule studies of half-site reactivity in dihydroorotate dehydrogenase A, Biophys J 88 (2005) 167a-167a.

[75] J. Shi, B.A. Palfey, J. Dertouzos, A. Gafni, D. Steel, Subunit activity of dihydroorotate dehydrogenase a studied by single molecule spectroscopy, Biophys J 86 (2004) 493a-493a.

[76] J. Shi, J. Dertouzos, A. Gafni, D. Steel, B.A. Palfey, Single-molecule kinetics reveals signatures of half-sites reactivity in dihydroorotate dehydrogenase A catalysis, Proc Nat Acad Sci of the United States of America 103 (2006) 5775-5780.

[77] J.R. Brender, J.M. Dertouzos, J. Shi, E. Burmeister, B. Palfey, D.P. Ballou, V. Massey, A. Gafni, D.G. Steel, Conformational regulation and heterogeneity of para-hydroxybenzoate hydroxylase studied at the single molecule level, Biophys J 84 (2003) 124a-124a.

[78] J.R. Brender, J.M. Dertouzos, B. Palfey, D.P. Ballou, V. Massey, A. Gafni, D.G. Steel, Kinetics of conformational regulation of para-hydroxybenzoate hydroxylase at the single molecule level, Biophys J 86 (2004) 474a-474a.

[79] J.R. Brender, J. Dertouzos, B. Palfey, D.P. Ballou, V. Massey, B. Entsch, A. Gafni, D.G. Steel, Conformational equilibria in para-hydroxybenzoate hydroxylase studied through single molecule fluorescence spectroscopy, Biophys J 88 (2005) 167a-167a.

[80] J.R. Brender, J. Dertouzos, D.P. Ballou, V. Massey, B.A. Palfey, B. Entsch, D.G. Steel, A. Gafni, Conformational dynamics of the isoalloxazine in substrate-free p-hydroxybenzoate hydroxylase: Single-molecule studies, J Am Chem Soc 127 (2005) 18171-18178.

[81] W.L. Jorgensen, D.S. Maxwell, J. TiradoRives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J Am Chem Soc 118 (1996) 11225-11236.

[82] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules, Comput Phys Commun 91 (1995) 1-41.

[83] Y.K. Zhang, T.S. Lee, W.T. Yang, A pseudobond approach to combining quantum mechanical and molecular mechanical methods, J Chem Phys 110 (1999) 46-54.

[84] J.L. Gao, P. Amara, C. Alhambra, M.J. Field, A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations, J Phys Chem A 102 (1998) 4714-4721.

[85] R.B. Murphy, D.M. Philipp, R.A. Friesner, Frozen orbital QM/MM methods for density functional theory, Chem Phys Lett 321 (2000) 113-120.

[86] S.F. Boys, Electronic Wave Functions .3. Some Theorems on Integrals of Antisymmetric Functions of Equivalent Orbital Form, Proc R Soc Lon Ser-A 206 (1951) 489-509.

[87] A.E. Cho, V. Guallar, B.J. Berne, R. Friesner, Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach, J Comput Chem 26 (2005) 915-931.

[88] D.M. Philipp, R.A. Friesner, Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide, J Comput Chem 20 (1999) 1468-1494.

[89] F.S. Csajka, D. Chandler, Transition pathways in a many-body system: Application to hydrogen-bond breaking in water, J Chem Phys 109 (1998) 1125-1133.

[90] C. Dellago, P.G. Bolhuis, F.S. Csajka, D. Chandler, Transition path sampling and the calculation of rate constants, J Chem Phys 108 (1998) 1964-1977.

[91] S. Fischer, M. Karplus, Conjugate Peak Refinement - an Algorithm for Finding Reaction Paths and Accurate Transition-States in Systems with Many Degrees of Freedom, Chem Phys Lett 194 (1992) 252-261.

[92] Q. Cui, M. Karplus, Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations, J Chem Phys 112 (2000) 1133-1149.

[93] D. Chasman, M.D. Beachy, L.M. Wang, R.A. Friesner, Parallel pseudospectral electronic structure: I. Hartree-Fock calculations, J Comput Chem 19 (1998) 1017-1029.

[94] J.L. Banks, H.S. Beard, Y.X. Cao, A.E. Cho, W. Damm, R. Farid, A.K. Felts, T.A. Halgren, D.T. Mainz, J.R. Maple, R. Murphy, D.M. Philipp, M.P. Repasky, L.Y. Zhang, B.J. Berne, R.A. Friesner, E. Gallicchio, R.M. Levy, Integrated modeling program, applied chemical theory (IMPACT), J Comput Chem 26 (2005) 1752-1780.

[95] V. Guallar, M. Jacobson, A. McDermott, R.A. Friesner, Computational modeling of the catalytic reaction in triosephosphate isomerase, J Mol Biol 337 (2004) 227-239.

[96] A. Altun, V. Guallar, R.A. Friesner, S. Shaik, W. Thiel, The effect of heme environment on the hydrogen abstraction reaction of camphor in P450(cam) catalysis: A QM/MM study, J Am Chem Soc 128 (2006) 3924-3925.

[97] A. Altun, S. Shaik, W. Thiel, Systematic QM/MM investigation of factors that affect the cytochrome P450-catalyzed hydrogen abstraction of camphor, J Comput Chem 27 (2006) 1324-1337.

[98] A. Altun, W. Thiel, Combined quantum mechanical/molecular mechanical study on the pentacoordinated ferric and ferrous cytochrome P450(cam) complexes, J Phys Chem B 109 (2005) 1268-1280.

[99] R.A. Friesner, V. Guallar, Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis, Annu Rev Phys Chem 56 (2005) 389-427.

[100] B.F. Gherman, S.D. Goldberg, V.W. Cornish, R.A. Friesner, Mixed quantum mechanical/molecular mechanical (QM/MM) study of the deacylation reaction in a penicillin binding protein (PBP) versus in a class C beta-lactamase, J Am Chem Soc 126 (2004) 7652-7664.

[101] R.A. Friesner, M.H. Baik, B.F. Gherman, V. Guallar, M. Wirstam, R.B. Murphy, S.J. Lippard, How iron-containing proteins control dioxygen chemistry: a detailed atomic level description via accurate quantum chemical and mixed quantum mechanics/molecular mechanics calculations, Coordin Chem Rev 238 (2003) 267-290.

[102] B.F. Gherman, S.J. Lippard, R.A. Friesner, Substrate hydroxylation in methane monooxygenase: Quantitative modeling via mixed quantum mechanics/molecular mechanics techniques, J Am Chem Soc 127 (2005) 1025-1037.

[103] M. Wirstam, S.J. Lippard, R.A. Friesner, Reversible dioxygen binding to hemerythrin, J Am Chem Soc 125 (2003) 3980-3987.

[104] P. Rowland, F.S. Nielsen, K.F. Jensen, S. Larsen, The crystal structure of the flavin containing enzyme dihydroorotate dehydrogenase A from *Lactococcus lactis*, Structure 5 (1997) 239-252.

[105] P. Rowland, O. Bjornberg, F.S. Nielsen, K.F. Jensen, S. Larsen, The crystal structure of *Lactococcus lactis* dihydroorotate dehydrogenase A complexed with the enzyme reaction product throws light on its enzymatic function, Protein Science 7 (1998) 1269-1279.

[106] S. Liu, E.A. Neidhardt, T.H. Grossman, T. Ocain, J.J. Clardy, Structures of human dihydroorotate dehydrogenase in complex with antiproliferative agents, Structure 8 (2000) 25-33.

[107] P. Rowland, S. Norager, K.F. Jensen, S. Larsen, Structure of dihydroorotate dehydrogenase B: Electron transfer between two flavin groups bridged by an iron-sulphur cluster, Structure 8 (2000) 1227-1238.

[108] W.L. Elliott, D.P. Sawick, S.A. DeFrees, P.F. Heinstein, J.M. Cassady, D.J. Morre, Cyclic modulation of enzymes of pyrimidine nucleotide biosynthesis

precedes sialoglycoconjugate changes during 2-acetylaminofluorene-induced hepatocarcinogenesis in the rat, Biochimica et Biophysica Acta 800 (1984) 194-201.

[109] G. Weber, Biochemical strategy of cancer cells and the design of chemotherapy: G. H. A. Clowes memorial lecture, Cancer Research 43 (1983) 3466-3492.

[110] F.C. Breedveld, J.-M. Dayer, Leflunomide: mode of action in the treatment of rheumatoid arthritis, Ann. Rheum. Dis. 59 (2000) 841-849.

[111] B.A. Palfey, O. Bjornberg, K.F. Jensen, Specific inhibition of a family 1A dihydroorotate dehydrogenase by benzoate pyrimidine analogues, Journal of Medicinal Chemistry 44 (2001) 2861-2864.

[112] F.C. Breedveld, J.M. Dayer, Leflunomide: mode of action in the treatment of rheumatoid arthritis, Ann Rheum Dis 59 (2000) 841-849.

[113] H.M. Cherwinski, N. Byars, S.J. Ballaron, G.M. Nakano, J.M. Young, J.T. Ransom, Leflunomide interferes with pyrimidine nucleotide biosynthesis, Inflammation Research 44 (1995) 317-322.

[114] H.M. Cherwinski, R.G. Cohn, P. Cheung, D.J. Webster, Y.Z. Xu, J.P. Caulfield, J.M. Young, G.M. Nakano, J.T. Ransom, The immunosuppressant leflunomide inhibits lymphocyte-proliferation by inhibiting pyrimidine biosynthesis, Journal of Pharmacology and Experimental Therapeutics 275 (1995) 1043-1049.

[115] O. Bjornberg, A.-C. Gruner, P. Roepstorff, K.F. Jensen, The activity of *Escherichia coli* dihydroorotate dehydrogenase is dependent on a conserved loop identified by sequence homology, mutagenesis, and limited proteolysis, Biochemistry 38 (1999) 2899-2908.

[116] O. Bjornberg, P. Rowland, S. Larsen, K.F. Jensen, Active site of dihydroorotate dehydrogenase A from *Lactococcus lactis* investigated by chemical modification and mutagenesis, Biochemistry 36 (1997) 16197-16205.

[117] B.A. Palfey, O. Bjornberg, K.F. Jensen, Insight into the chemistry of flavin reduction and oxidation in *Escherichia coli* dihydroorotate dehydrogenase obtained by rapid reaction studies, Biochemistry 40 (2001) 4381-4390.

[118] S. Norager, K.F. Jensen, O. Bjornberg, S. Larsen, *E. coli* dihydroorotate dehydrogenase reveals structural and functional distinctions between different classes of dihydroorotate dehydrogenases, Structure 10 (2002) 1211-1223.

[119] R.A. Pascal, Jr., C.T. Walsh, Mechanistic studies with deuterated dihydroorotates on the dihydroorotate oxidase from *Crithidia fasciculata*, Biochemistry 23 (1984) 2745-2752.

[120] V. Hines, M. Johnston, Mechanistic studies on the bovine liver mitochondrial dihydroorotate dehydrogenase using kinetic deuterium isotope effects, Biochemistry 28 (1989) 1227-1234.

[121] A. Argyrou, M.W. Washabaugh, C.M. Pickart, Dihydroorotate dehydrogenase from *Clostridium oroticum* is a class 1B enzyme and utilizes a concerted mechanism of catalysis, Biochemistry 39 (2000) 10373-10384.

[122] J. Shi, B.A. Palfey, J. Dertouzos, K.F. Jensen, A. Gafni, D. Steel, Multiple states of the Tyr318Leu mutant of dihydroorotate dehydrogenase revealed by single-molecule kinetics, Journal of the American Chemical Society 126 (2004) 6914-6922.

[123] S. Norager, S. Arent, O. Bjornberg, M. Ottosen, L. Lo Leggio, K.F. Jensen, S. Larsen, *Lactococcus lactis* dihydroorotate dehydrogenase A mutants reveal important facets of the enzymatic function, Journal of Biological Chemistry 278 (2003) 28812-28822.

[124] M. Hansen, J. Le Nours, E. Johansson, T. Antal, A. Ullrich, M. Loffler, S. Larsen, Inhibitor binding in a Class 2 dihydroorotate dehydrogenase causes variations in the membrane-associated N-terminal domain, Protein Science 13 (2004) 1031-1042.

[125] W. Jiang, G. Locke, M.R. Harpel, R.A. Copeland, J. Marcinkeviciene, Role of Lys100 in human dihydroorotate dehydrogenase: Mutagenesis studies and chemical rescue by external amines, Biochemistry 39 (2000) 7990-7997.

[126] V. Guallar, D.L. Harris, V.S. Batista, W.H. Miller, Proton-transfer dynamics in the activation of cytochrome P450efyF, Journal of the American Chemical Society 124 (2002) 1430-1437.

[127] D. Lu, G.A. Voth, Molecular dynamics simulations of human carbonic anhydrase II: Insight into experimental results and the role of solvation, PROTEINS: Structure, Function, and Genetics 33 (1998) 119-134.

[128] S. Toba, G. Colombo, K.M. Merz, Jr., Solvent dynamics and mechanism of proton transfer in human carbonic anhydrase II, Journal of the American Chemical Society 121 (1999) 2290-2302.

[129] R.I. Cukier, A molecular dynamics study of water chain formation in the proton-conducting K channel of cytochrome c oxidase, Biochimica et Biophysica Acta 1706 (2005) 134-146.

[130] J.Z. Xiang, B. Honig, JACKAL: A Protein Structure Modeling Package, Columbia University & Howard Hughes Medical Institute, New York, 2002.

[131] IMPACT, Schrödinger, Inc., Portland, OR, 2000.

[132] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, Journal of the American Chemical Society 118 (1996) 11225-11236.

[133] QSite, Schrödinger, Inc., Portland, OR, 2005.

[134] R.B. Murphy, D.M. Philipp, R.A. Friesner, A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments, Journal of Computational Chemistry 21 (2000) 1442-1457.

[135] R. Ditchfield, W.J. Hehre, J.A. Pople, Journal of Chemical Physics 54 (1971) 724-728.

[136] W.J. Hehre, R. Ditchfield, J.A. Pople, Self-consistent molecular orbital methods: Further extensions of Gaussian-type basis sets for use in molecular-orbital studies of organic molecules, Journal of Chemical Physics 56 (1972) 2257-2261.

[137] M.M. Francl, W.J. Pietro, W.J. Hehre, J.S. Binkley, M.S. Gordon, D.J. DeFrees, J.A. Pople, Self-consistent molecular-orbital methods: A polarization-type basis set for 2nd-row elements, Journal of Chemical Physics 77 (1982) 3654-3665.

[138] C. Lee, W. Yang, P.G. Parr, Development of the Colle-Salvetti correlation energy formula into a functional of the electron density, Physical Review B 37 (1988) 785-789.

[139] A.D. Becke, Density-functional thermochemistry: The role of exact exchange, Journal of Chemical Physics 98 (1993) 5648-5652.

[140] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), Intermolecular Forces, D. Reidel Publishing Company, Dordrecht, 1981.

[141] C.M. Breneman, K.B. Wiberg, Determining atom-centered monopoles from molecular electrostatic potentials -- the need for high sampling density in

formamide conformational analysis, Journal of Computational Chemistry 11 (1990) 361-373.

[142] L.E. Chirlian, M.M. Francl, Atomic charges derived from electrostatic potentials -- a detailed study, Journal of Computational Chemistry 8 (1987) 894-905.

[143] R.J. Woods, M. Khalil, W. Pell, S.H. Moffat, V.H. Smith, Derivation of net atomic charges from molecular electrostatic potentials, Journal of Computational Chemistry 11 (1990) 297-310.

[144] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, Journal of Chemical Physics 81 (1994) 3684-3690.

[145] P.P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, Ann. Phys. 64 (1921) 253-287.

[146] E. Lindahl, B. Hess, D. Van der Spoel, GROMACS 3.0: A package for molecular simulation and trajectory analysis, Journal of Molecular Modelling 7 (2001) 306-317.

[147] H.J.C. Berendsen, D. Van der Spoel, R. Van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comp. Phys. Comm. 91 (1995) 43-56.

[148] T. Darden, D. York, L. Pedersen, Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems, Journal of Chemical Physics 98 (1993) 10089-10092.

[149] A.W. Schuettelkopf, D.M.F. Van Aalten, PRODRG - a tool for high-throughput crystallography of protein-ligand complexes, Acta Crystallography D60 (2004) 1355-1363.

[150] S. Nose, A molecular dynamics method for simulations in the canonical ensemble, Molecular Physics 52 (1984) 255-268.

[151] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Physical Review A 31 (1985) 1695-1697.

[152] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Journal of Computational Physics 23 (1977) 327.

[153] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, H.J.C. Berendsen, GROMACS, 2004.

[154] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Clarendon Press, Oxford, 1989.

[155] V. Hines, L.D. Keys, III, M. Johnston, Purification and properties of the bovine liver mitochondrial dihydroorotate dehydrogenase, Journal of Biological Chemistry 261 (1986) 1386-1392.

[156] N.A. McDonald, W.L. Jorgensen, Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles, J Phys Chem B 102 (1998) 8049-8059.

[157] Y.A. Small, V. Guallar, A.V. Soudackov, S. Hammes-Schiffer, Hydrogen bonding pathways in human dihydroorotate dehydrogenase, J Phys Chem B 110 (2006) 19704-19710.

[158] U. Rothlisberger, P. Carloni, K. Doclo, M. Parrinello, A comparative study of galactose oxidase and active site analogs based on QM/MM Car Parrinello simulations, J Biol Inorg Chem 5 (2000) 236-250.

[159] S. Marti, V. Moliner, Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems, J Chem Theory Comput 1 (2005) 1008-1016.

[160] M. Garcia-Viloca, D.G. Truhlar, J.L. Gao, Reaction-path energetics and kinetics of the hydride transfer reaction catalyzed by dihydrofolate reductase, Biochemistry-Us 42 (2003) 13558-13575.

[161] P.T.R. Rajagopalan, Z.Q. Zhang, L. McCourt, M. Dwyer, S.J. Benkovic, G.G. Hammes, Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics, P Natl Acad Sci USA 99 (2002) 13481-13486.

[162] P.K. Agarwal, S.R. Billeter, S. Hammes-Schiffer, Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis, J Phys Chem B 106 (2002) 3283-3293.

[163] V.M. Krishnamurthy, B.R. Bohall, V. Semetey, G.M. Whitesides, The paradoxical thermodynamic basis for the interaction of ethylene glycol, glycine, and sarcosine chains with bovine carbonic anhydrase II: An unexpected manifestation of enthalpy/entropy compensation, J Am Chem Soc 128 (2006) 5802-5812.

[164] T.A. Ozal, N.F.A. van der Vegt, Confusing cause and effect: Energy-entropy compensation in the preferential solvation of a nonpolar solute in dimethyl sulfoxide/water mixtures, J Phys Chem B 110 (2006) 12104-12112.

[165] N.F.A. van der Vegt, M.E. Lee, D. Trzesniak, W.F. van Gunsteren, Enthalpy-entropy compensation in the effects of urea on hydrophobic interactions, J Phys Chem B 110 (2006) 12852-12855.

[166] O. Bjornberg, D.B. Jordan, B.A. Palfey, K.F. Jensen, Dihydrooxonate is a substrate of dihydroorotate dehydrogenase (DHOD) providing evidence for involvement of cysteine and serine residues in base catalysis, Arch Biochem Biophys 391 (2001) 286-294.

[167] M. Hansen, J. Le Nours, E. Johansson, T. Antal, A. Ullrich, M. Loffler, S. Larsen, Inhibitor binding in a class 2 dihydroorotate dehydrogenase causes variations in the membrane-associated N-terminal domain, Protein Sci 13 (2004) 1031-1042.

[168] S.P. Liu, E.A. Neidhardt, T.H. Grossman, T. Ocain, J. Clardy, Structures of human dihydroorotate dehydrogenase in complex with antiproliferative agents, Struct Fold Des 8 (2000) 25-33.

[169] D.Y. Leger, B. Liagre, J.L. Beneytout, Low dose leflunomide activates PI3K/Akt signalling in erythroleukemia cells and reduces apoptosis induced by anticancer agents, Apoptosis 11 (2006) 1747-1760.

[170] D.E. Hurt, J. Widom, J. Clardy, Structure of Plasmodium falciparum dihydroorotate dehydrogenase with a bound inhibitor, Acta Crystallogr D 62 (2006) 312-323.

[171] D.E. Hurt, A.E. Sutton, J. Clardy, Brequinar derivatives and species-specific drug design for dihydroorotate dehydrogenase, Bioorganic & Medicinal Chemistry Letters 16 (2006) 1610-1615.

[172] M.R. Sawaya, J. Kraut, Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence, Biochemistry-Us 36 (1997) 586-603.

[173] D.V. Cramer, F.A. Chapman, B. Jaffee, E.A. Jones, L. Makowka, Inhibition of the Pyrimidine Biosynthetic-Pathway with S-8660, an Analog of Brequinar Sodium, Prolongs Cardiac Allograft Survival in Rats, J Heart Lung Transpl 12 (1993) 140-146.

[174] B.A. Palfey, O. Bjornberg, F. Jensen, Specific inhibition of a family 1A dihydroorotate dehydrogenase by benzoate pyrimidine analogues, J Med Chem 44 (2001) 2861-2864.

[175] S.A. Seibold, D.A. Mills, S. Ferguson-Miller, R.I. Cukier, Water chain formation and possible proton pumping routes in Rhodobacter sphaeroides cytochrome c

oxidase: A molecular dynamics comparison of the wild type and R481K mutant, Biochemistry-Us 44 (2005) 10475-10485.

[176] M.P. Jacobson, D.L. Pincus, C.S. Rapp, T.J.F. Day, B. Honig, D.E. Shaw, R.A. Friesner, A hierarchical approach to all-atom protein loop prediction, Proteins 55 (2004) 351-367.

[177] Z.X. Xiang, B. Honig, Evaluating ensemble energies in sparsely sampled space: The colony energy concept and its application to the problem of loop prediction, Biophys J 82 (2002) 135a-135a.

[178] Z.X. Xiang, C.S. Soto, B. Honig, Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction, P Natl Acad Sci USA 99 (2002) 7432-7437.

[179] K.W. Borrelli, A. Vitalis, R. Alcantara, V. Guallar, PELE: Protein energy landscape exploration. A novel Monte Carlo based technique, J Chem Theory Comput 1 (2005) 1304-1311.

Appendix A

Preliminary Results: Gas Phase QM and QM/MM calculations

A.1 Gas phase QM results

Gas phase electronic structure calculations were carried out on the proton transfer components of the active site of DHOD as a first attempt to study the mechanism in DHOD. The active site was taken from two crystal structures; 2DOR is the structure for a class I DHOD where cysteine is the native base residue of interest and 1D3G is the structure for a class II DHOD where serine is the base residue of interest. The active site model contained only the base residue and the substrate DHO. Energies were calculated for each of the native enzyme configurations. Each native configuration was also mutated by replacing the native base with either serine or cysteine to form the mutant. Minima and transition states were calculated at the B3LYP/6-31G* level of theory using the Gaussian 98 suite of programs.(ref) The results for the energy are summarized in figure A-1 . The results illustrated that for the class I DHOD enzyme, the *in-silico* mutation from cysteine to serine produced the same transition state barrier of 20 kcal/mol. For the class II DHOD enzyme, the in-silico mutation from serine to cysteine, the native enzyme barrier was 12 kcal/mol and the mutated enzymes' barrier was lower by around 6 kcal/mol.

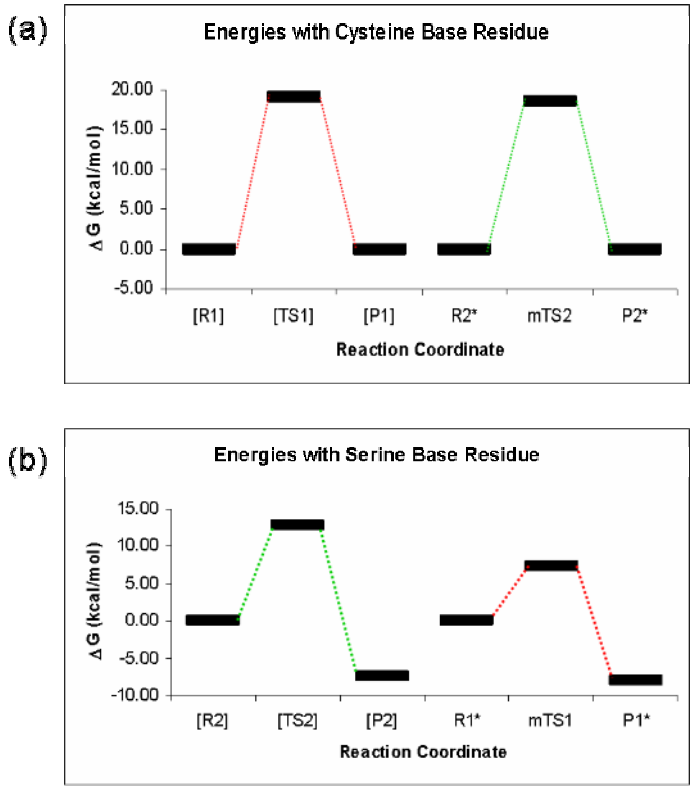

Figure A-1: Gas phase plots of the energy for the reactant (R), transition state (TS) and product (P). The native enzyme is represented with [ ], and the mutated enzyme is represented by *. (a) From the 2DOR crystal structure, the energy for the native enzyme is in red and the mutant is in green. (b) From the 1D3G crystal structure, the native enzyme is in red and the mutant is in green.

The structural comparison of comparison is the data is summarized in table A-1 . The structural results show that, at the transition state, the structures with cysteine as the base residue have a TS donor-acceptor distance of around 3.1 Å and the structures with a serine as the base residue have a TS donor-acceptor distance of around 2.6 Å. This could be due to the longer van der Waals radii of the sulfer atom over the oxygen atom.

Table A-1: Transition state distances for the class I (2DOR) and class II (1D3G) DHOD proton transfer components and their corresponding mutants. Class I is represented by TS1 and class II is represented by TS2. Distances where cysteine is the base residue are represented in yellow. Distances where serine is the base residue are represented in red. The donor-acceptor distance is R, the proton-acceptor distance is r, the donor-proton distance is r'. All distances are shown in units of Angstroms (Å)

	R	r	r'
TS1	3.11	1.74	1.36
mut_TS2	3.13	1.52	1.61
TS2	2.57	1.27	1.32
mut_TS1	2.57	1.18	1.41

Figure A-2 shows the TS structures of all four calculations where the TS have been properly characterized by a frequency calculation. The single negative frequency obtained was verified to be the vibrational mode of the transferring proton.

Figure A-2

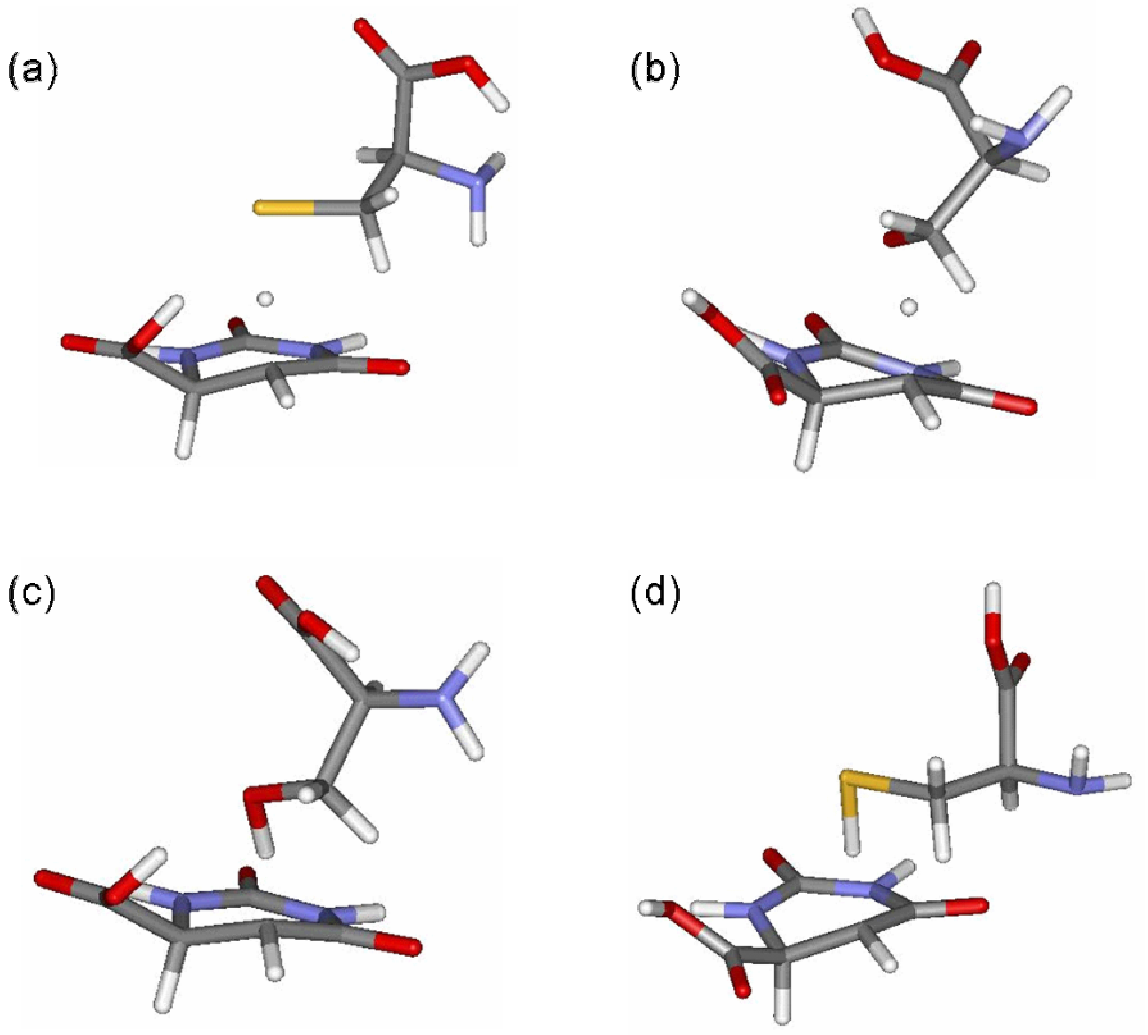

Figure A-2: Proton transfer transition states of class I and class II DHOD. The four transition states reflect (a) Native class I, (b) Mutated class I , (c) Native class II, (d) Mutated class II. In all figures the proton is approximately midway between the donor-acceptor distance as show in table A-1 .

A.2 QM/MM calculations with earlier version of QSite code

A.2.1 Methods

QM/MM calculations were performed using an earlier version of the code, v. 3.5. The objective of the study was to compare the wild type and mutant energy profiles where five possible reaction mechanisms were considered for each wild type and mutant model. The information that follows will describe the model of the active site, the strategy used for these calculations and the results obtained. This data has been relegated to the appendix because the results are inconclusive.

The model of the active site includes a truncated FMN cofactor (where the phosphate tail was replaced with a methyl group), DHO substrate, and residues Ser214 to Asn219, which includes the reactive base of interest. In the wild type, the base residue is Ser215 and in the mutant the oxygen on serine has been changed to sulfur to mimic the mutant's cysteine residue. The reacting base was initially deprotoned in these calculations.

The sequential and concerted mechanisms were considered. Figure A-3 is a schematic representation of the three cases. Five sets of calculations were carried out on the wild type and another five on the mutant using the following protocol to generate the energy profiles. Starting with the crystal structure, optimizations were performed with the proton and hydride frozen at a distance of 1.5 Å from their respective donor and acceptor atoms. Upon convergence, the frozen restrictions were removed and a transition state (TS) search calculation was performed.

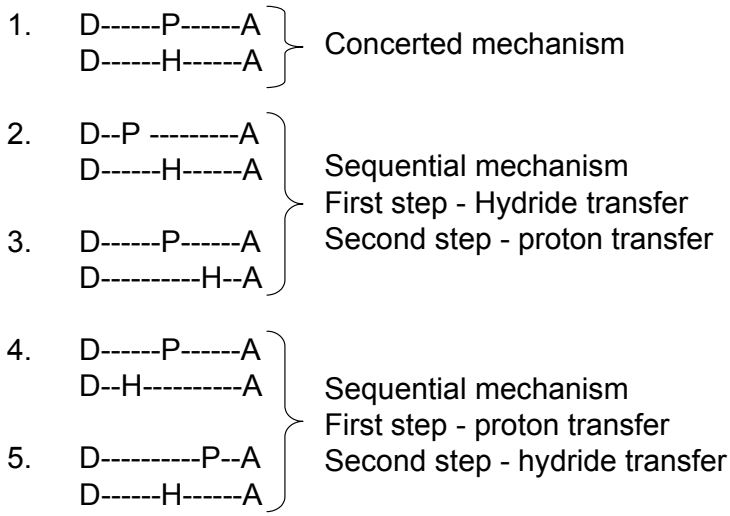

Figure A-3: Concerted and sequential mechanisms for reaction in DHOD. When the proton (P) or hydride (H) is located midway between the donor (D) atom and the acceptor (A) atoms, the state is either an intermediate or transition state.

When the TS calculation converged, the proton remained in the middle of its' donor and acceptor atoms but the hydride returned to it's donor atom. From this TS, the proton was nudged slightly towards its' donor to capture the nearest reactant minima. In another simultaneous calculation, the proton was nudged slightly towards its acceptor to capture the closest intermediate minima (the proton transfer product state). From the intermediate, the hydride was frozen for a second time at a distance of 1.5 Å from it's donor and acceptor atoms. Upon convergence, the frozen restrictions were removed and a transition state calculation was executed. The TS converged and the hydride remained in the middle of the donor and acceptor atoms while the proton was still on it's acceptor

atom. From the hydride transfer TS, the hydride was nudged slightly towards it's donor and acceptor atoms to capture the nearest reactant and product minima respectively.

A.2.2 Results

This method provided energy profiles for the sequential case where the proton transfers before the hydride. Energy profiles and structural results for mechanism 4 and 5 from figure A-3 are plotted below. Results for mechanisms 1, 2 and 3 could not be obtained because the calculations had convergence difficulties. Even though energy profiles could be obtained for one of the three mechanisms, the other two mechanisms cannot be ruled out in the absence of this data. Figure A-4 represents the proton transfer energy profile for mechanism 4.

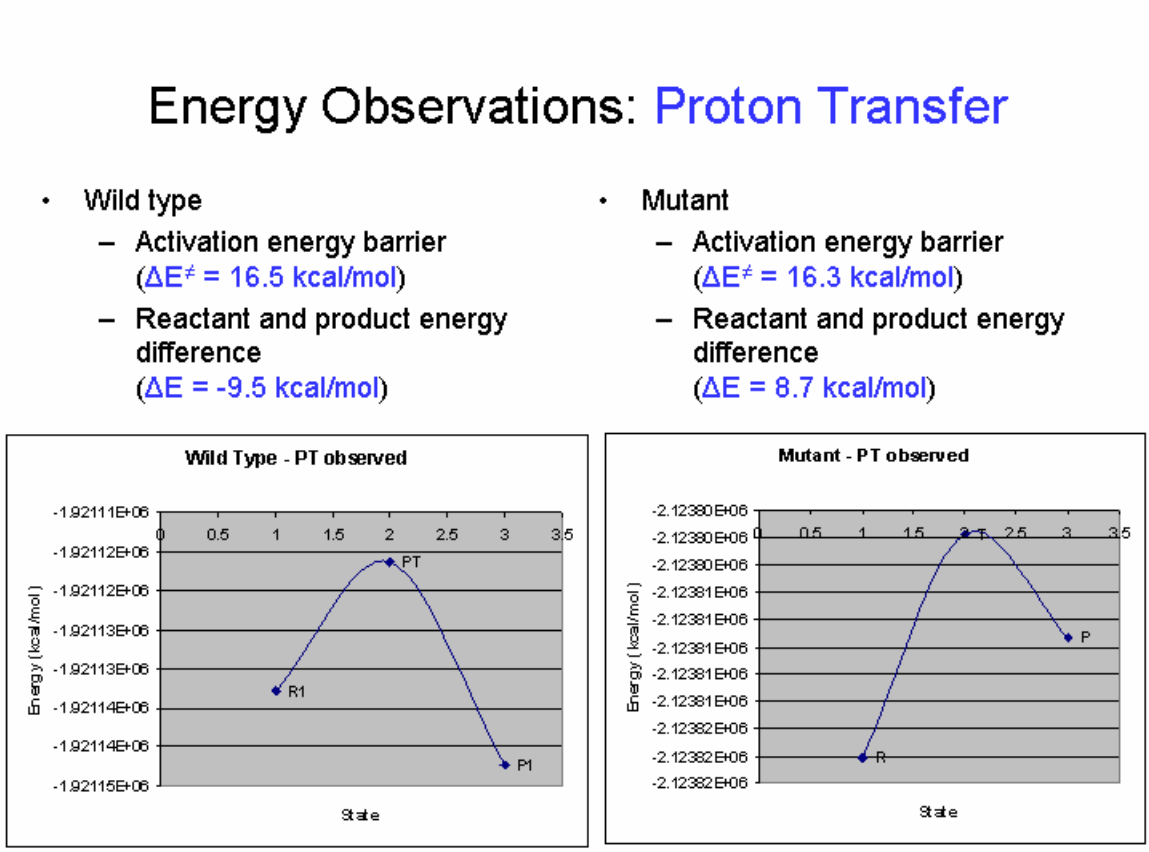

Figure A-4: Energy profile for wild type compared to mutant for proton transfer in DHOD.

For the proton transfer, the activation energy barriers for both wild type and mutant were similar at around 16 kcal/mol. However, the reaction energies were markedly different in that the wild type was exothermic by approximately 9 kcal/mol and the mutant was endothermic by around the same amount. One argument can be that since the natural pKa of cysteine is lower than serine, then a protonated serine is more stable (i.e. lower in energy) than a protonated cysteine. This could possible reflect that the mutant proceeds by a different mechanism than the wild type but this data is insufficient to confirm.

Figure A-5 illustrates the structural results for proton transfer, comparing the wild type to the mutant for mechanism 4.

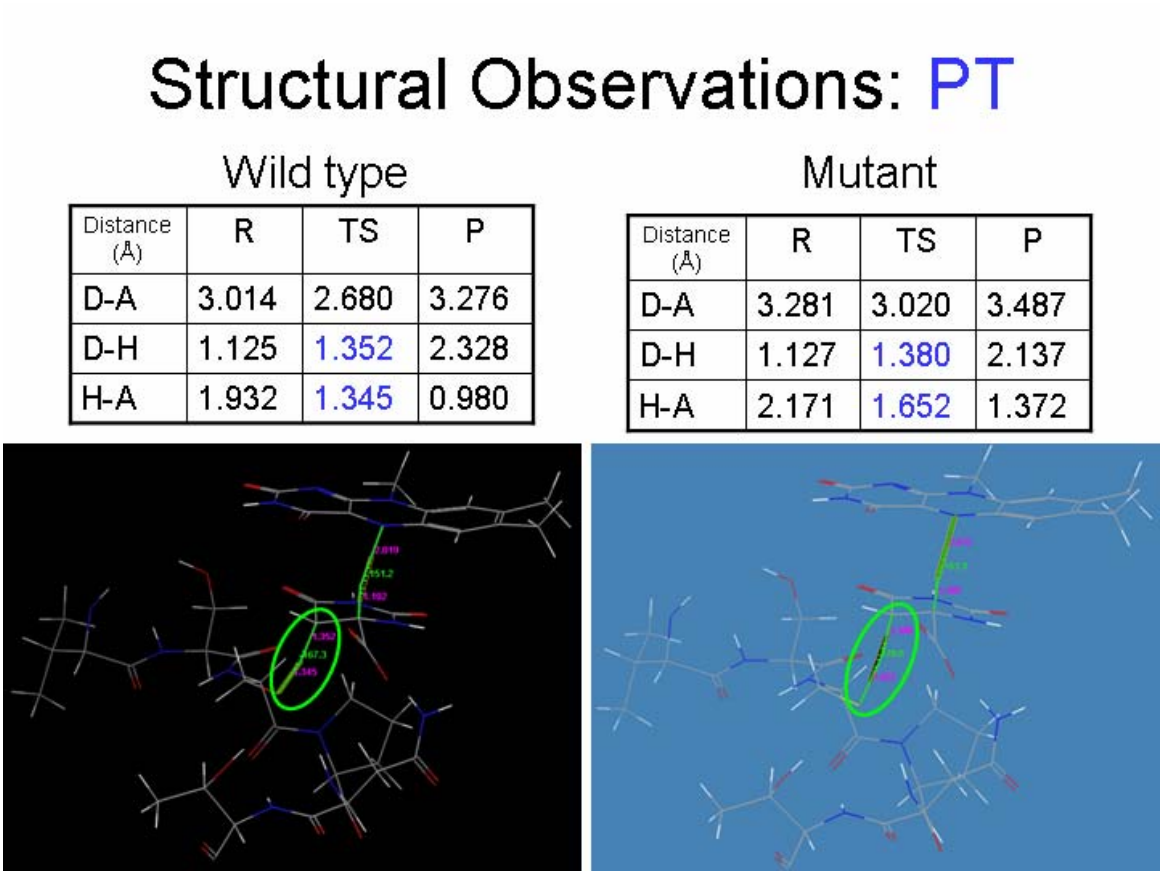

Figure A-5: Proton donor acceptor distances for wild type and mutant DHOD. The donor acceptor (D-A) distance shrinks at the transition state (TS), from the values at the reactant (R) and product (P) states for both wild type and mutant. The donor-proton (D-H) and acceptor proton (H-A) distance is symmetric about the position of the proton for the wild type but asymmetric for the mutant. All distance are shown in units of Angstroms.

The structural results indicate that the mutant maintained longer donor-acceptor distances than the wild type which may be an obvious result from the larger van der Waals radii of sulfur over oxygen. The TS distances for the wild type were more symmetric than in the

mutant which may again imply a different mechanism. Both wild type and mutant results show a compression of the donor acceptor distance at the active site allowing for the identification of the state as a transition state. The second step in this sequential process relates to mechanism 6. Figure A-6 represents the hydride transfer energy profile for mechanism 5.

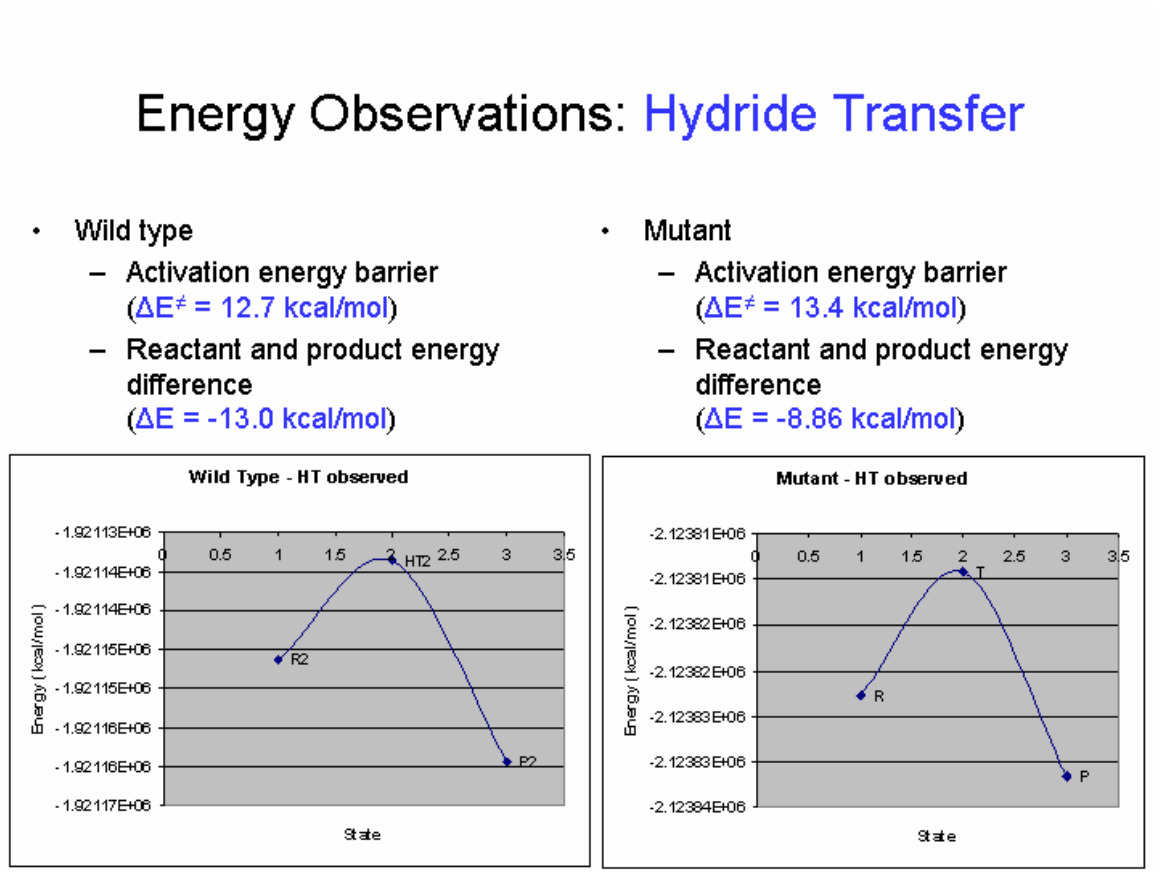

Figure A-6: Energy profile for wild type compared to mutant for hydride transfer in DHOD. Energy is measured relative to the reactant.

For the hydride transfer, both reaction energies and barriers were similar for the wild type and mutant, approximately -10.5 kcal/mol and 13 kcal/mol respectively. The

exothermicity of both profiles and consistency of the barriers may be a reflection of having the same components involved in this step for both wild type and mutant. Figure A-7 illustrates the structural results for hydride transfer, comparing the wild type to the mutant for mechanism 5.

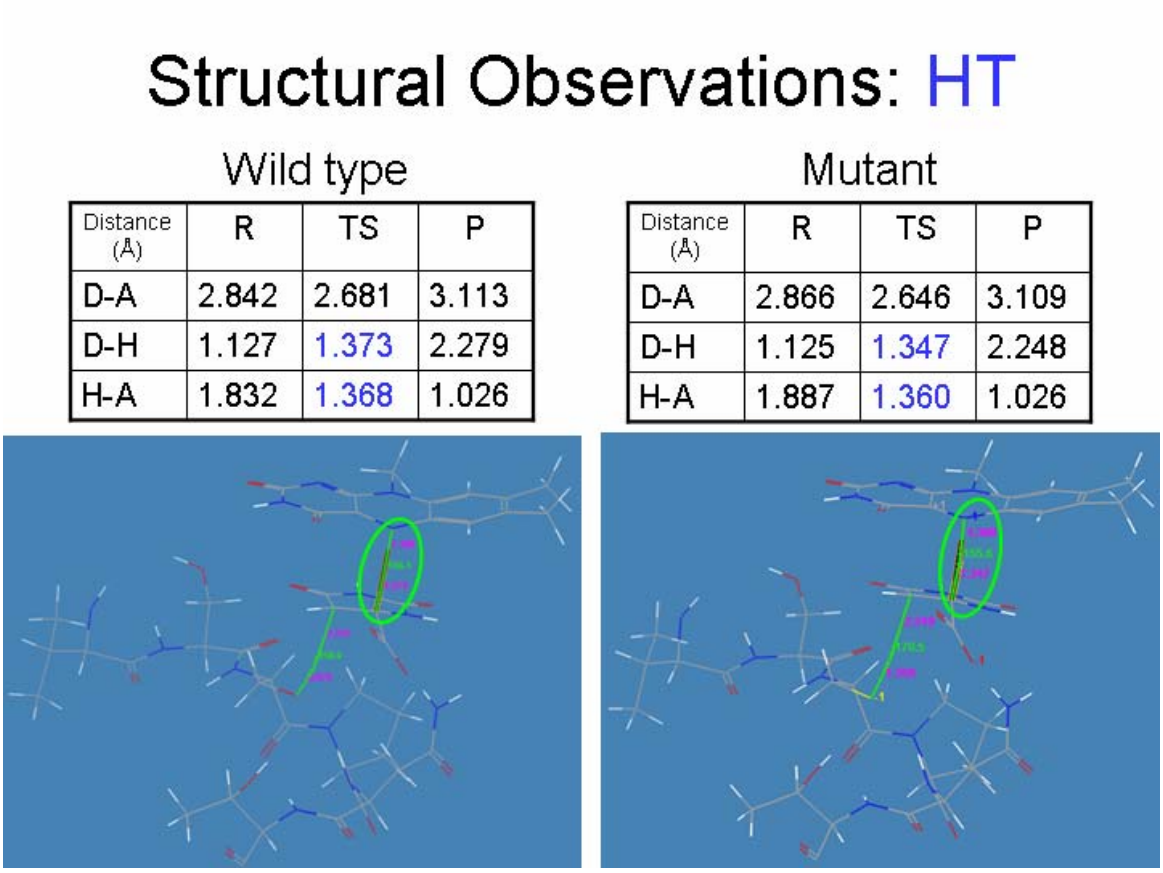

Figure A-7: Hydride donor acceptor distance for wild type and mutant DHOD. The donor acceptor (D-A) distance shrinks at the transition state (TS), from the values at the reactant (R) and product (P) states for both wild type and mutant. The donor-hydride (D-H) and acceptor-hydride (H-A) distance is symmetric about the position of the hydride for both the wild type and mutant. All distance are shown in units of Angstroms.

The structural donor-acceptor distances for the hydride transfer were also similar between the wild type and mutant. A far reaching conclusion is that the first proton transfer step is rate limiting due to inherent differences between SER and CYS.

The design of this approach was to construct an initial set of conditions with will comply with all three mechanisms by placing the both proton and hydride in the middle of their respective atoms. For unexplained reasons, the system migrated towards a sequential mechanism in which the proton transfers first but in the absence of comparable barriers for the other two mechanisms, no clear conclusions can be drawn. Obtaining barriers for the other two cases were extremely problematic but this does not indicate that those other states do not exist on the potential energy surface.

Appendix B

Supplementary Materials

B.1 Supplementary materials from 4 nanosecond long molecular dynamics analysis

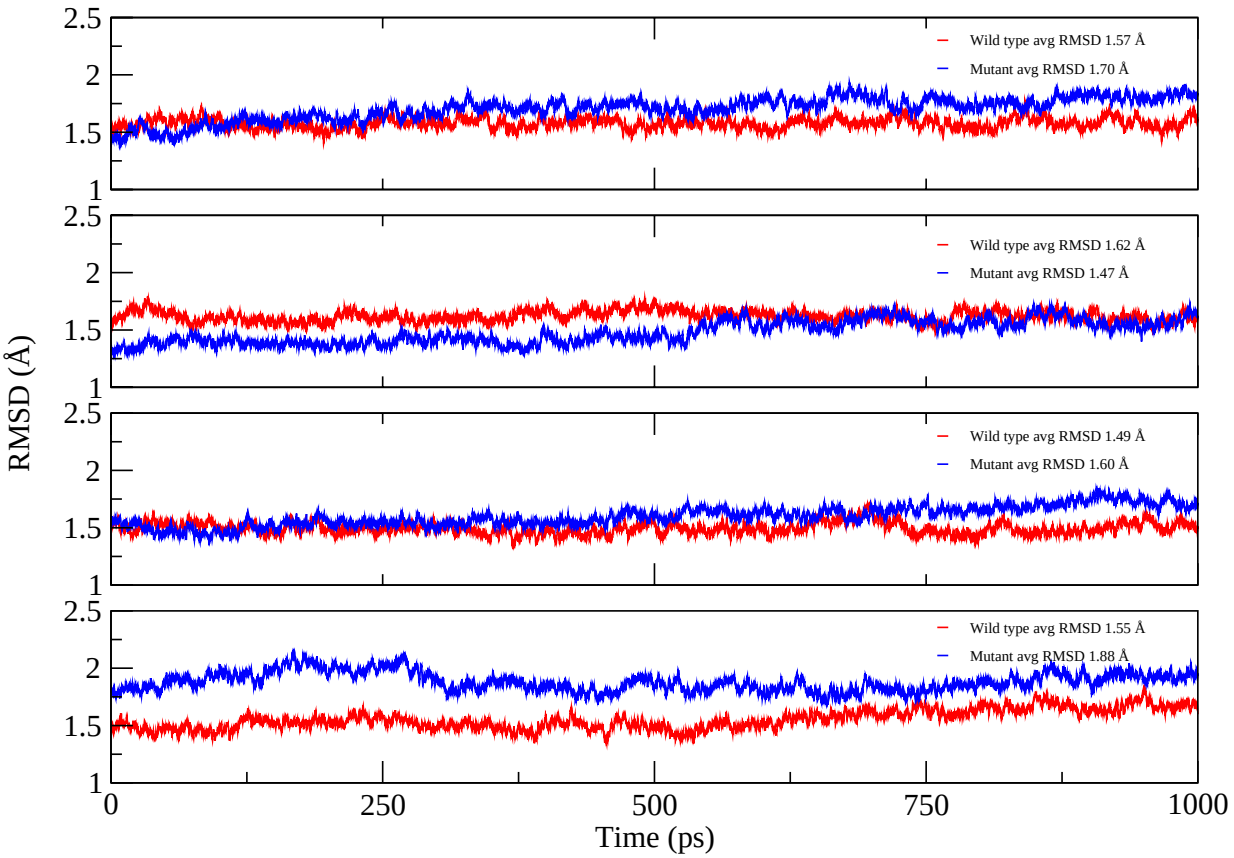

Figure B-1: Time evolution of RMS deviations with respect to the crystal structure C-alpha atoms for the four MD simulations in this work. Wild type trajectories are shown in red and mutant trajectories are shown in blue.

Figure B-2 Proton and hydride donor-acceptor distances

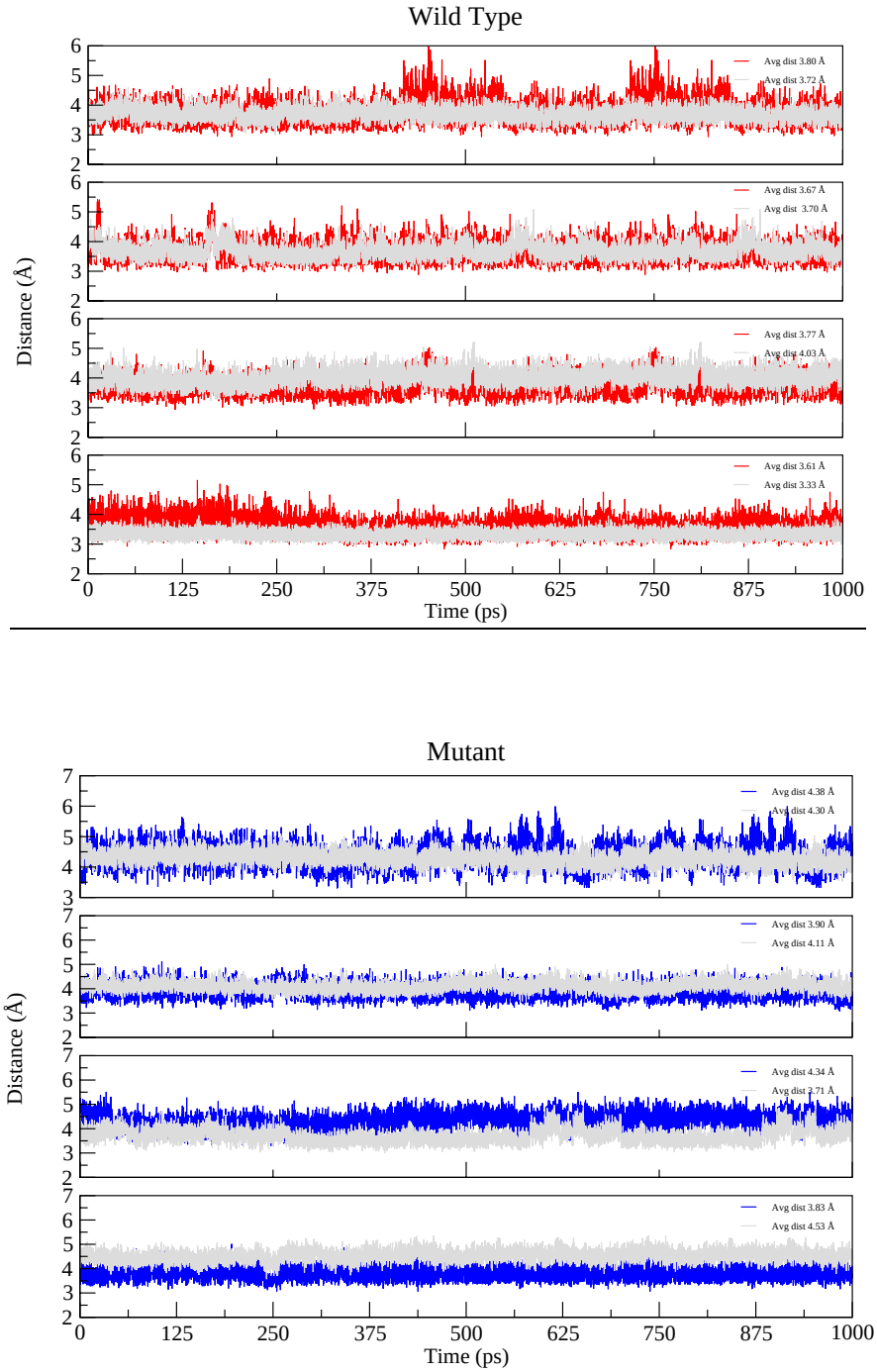

Figure B-2: Time evolution of donor-acceptor distances for the four MD simulations of this work. The proton donor-acceptor distance of the wild type is shown in red and the mutant is shown in blue. The hydride donor-acceptor distance is shown in gray.

Figure B-3 Substrate angle as a function of time for the wild type and mutant trajectories

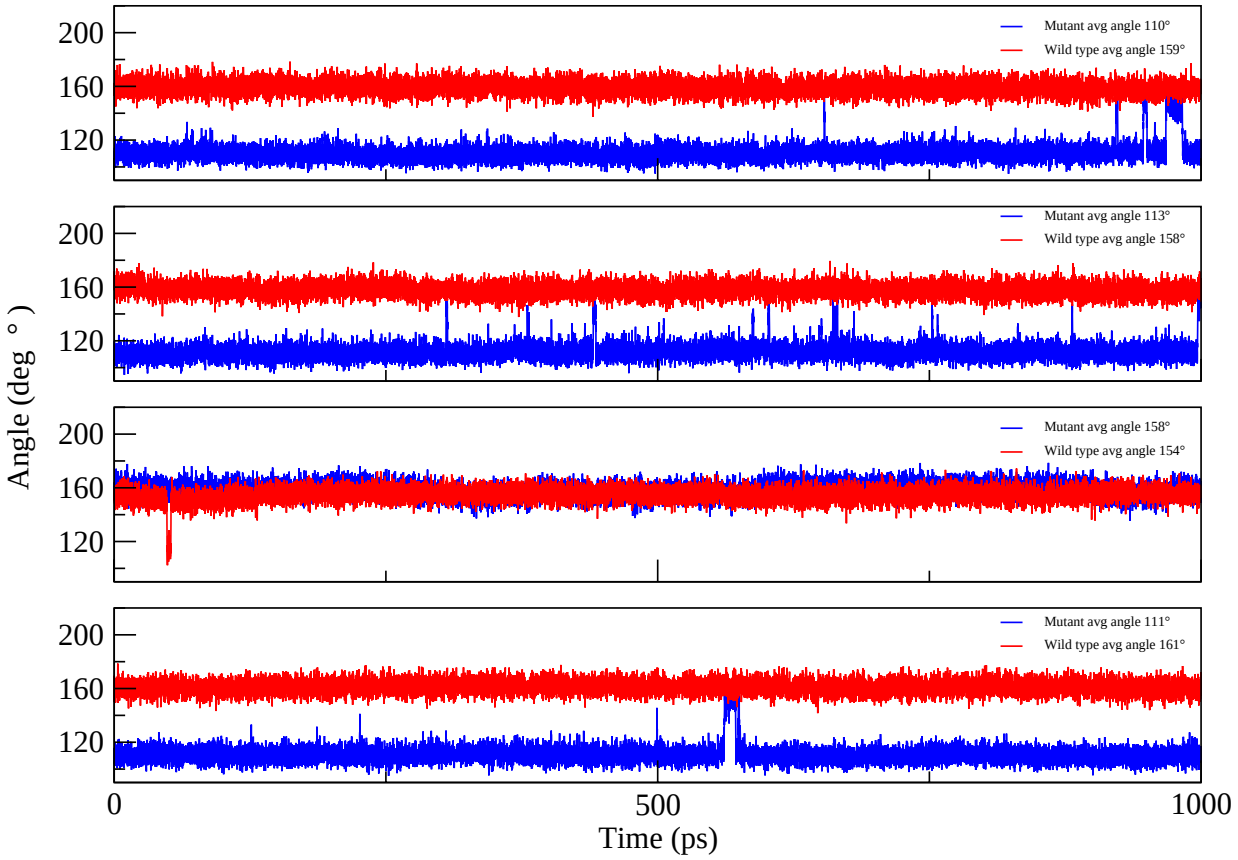

Figure B-3: Time evolution of the planar angle sampled by the substrate (DHO). The wild type angle is shown in red and the mutant angle is shown in blue.

Figure B-4 Hydrogen bonding Histogram

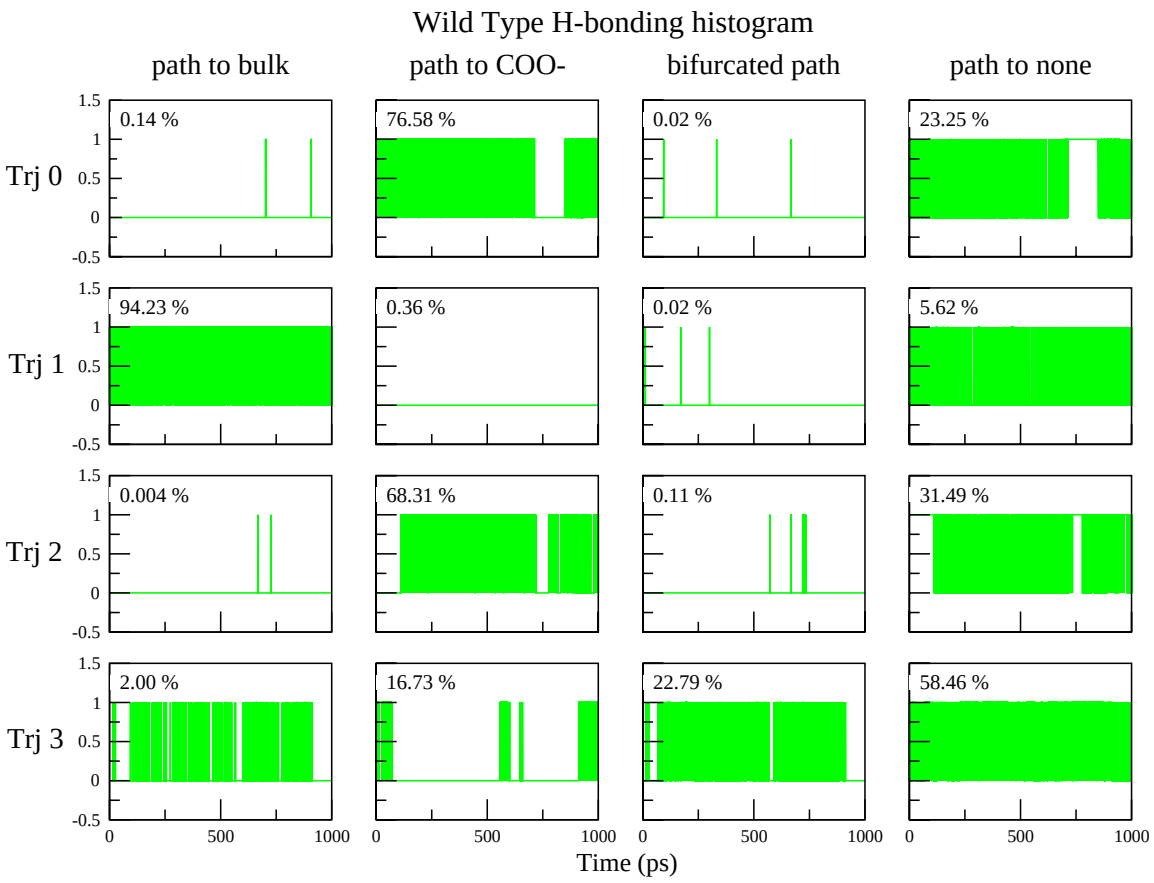

Figure B-4: Histogram of hydrogen bond formation with respect to time for the wild type simulations.

Figure B-5 Radial distribution functions

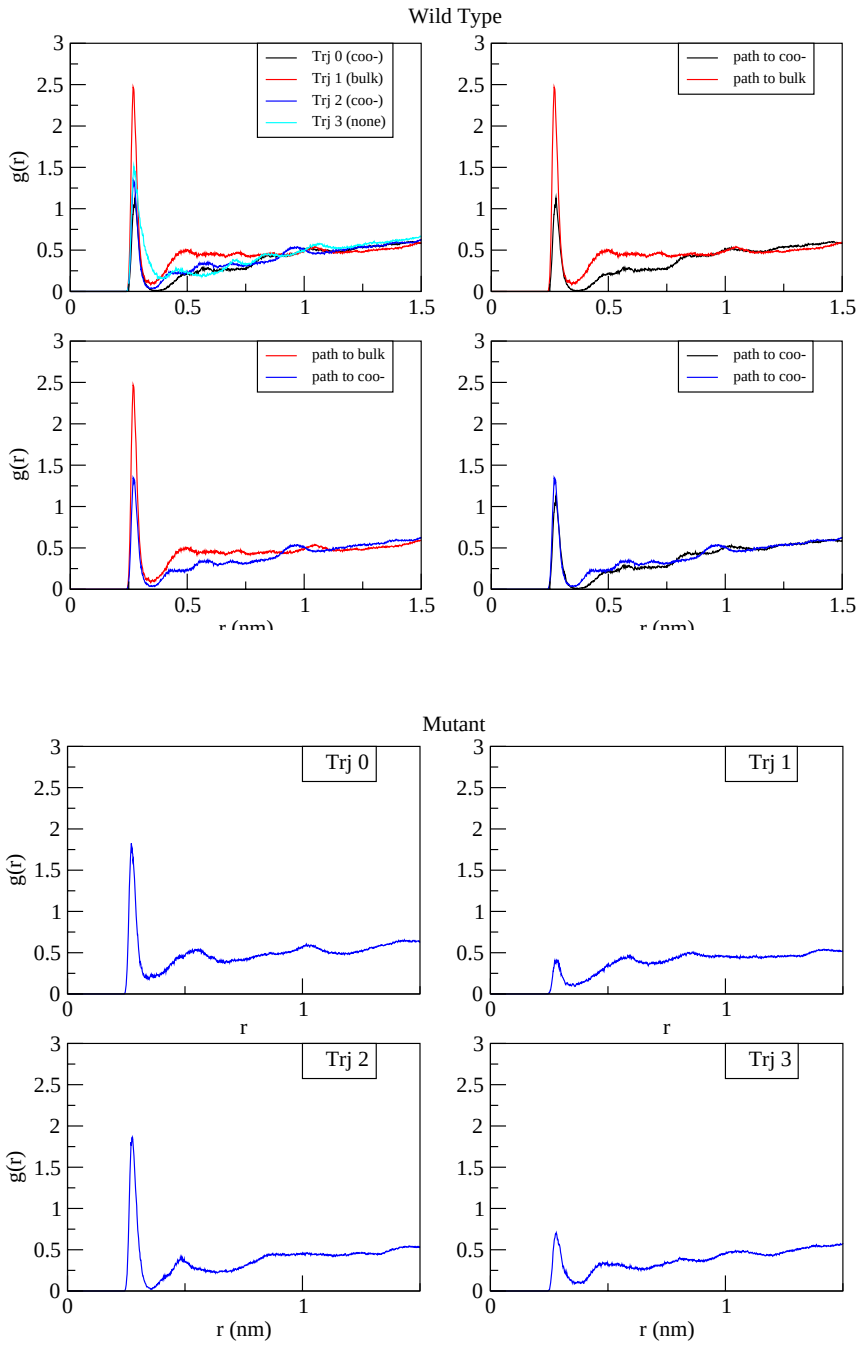

Figure B-5: Pair distribution functions for the oxygen in THR to the oxygens in the surrounding molecules. (a) Wild type, (b) Mutant

B.2 Supplementary materials from QM/MM calculations

(a) (b)

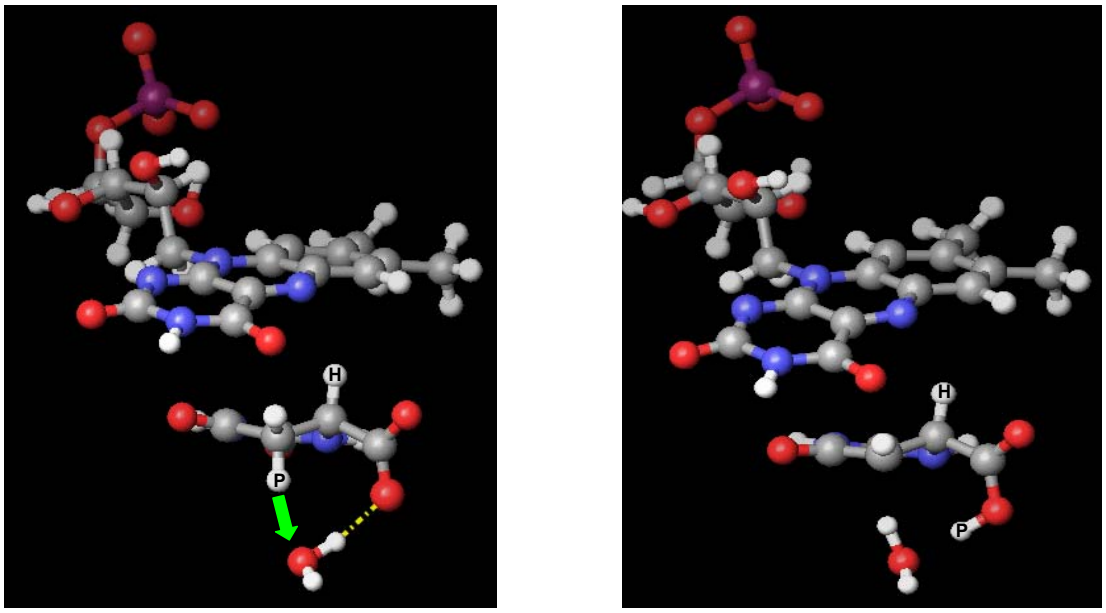

Figure B-6: Structural representation of sequential transfer where proton moves prior to hydride. Green arrow indicates direction of proton movement. Energy difference is approximately 36 kcal/mol. (a) Reactant state (b) Proton transfer intermediate.

(a) (b)

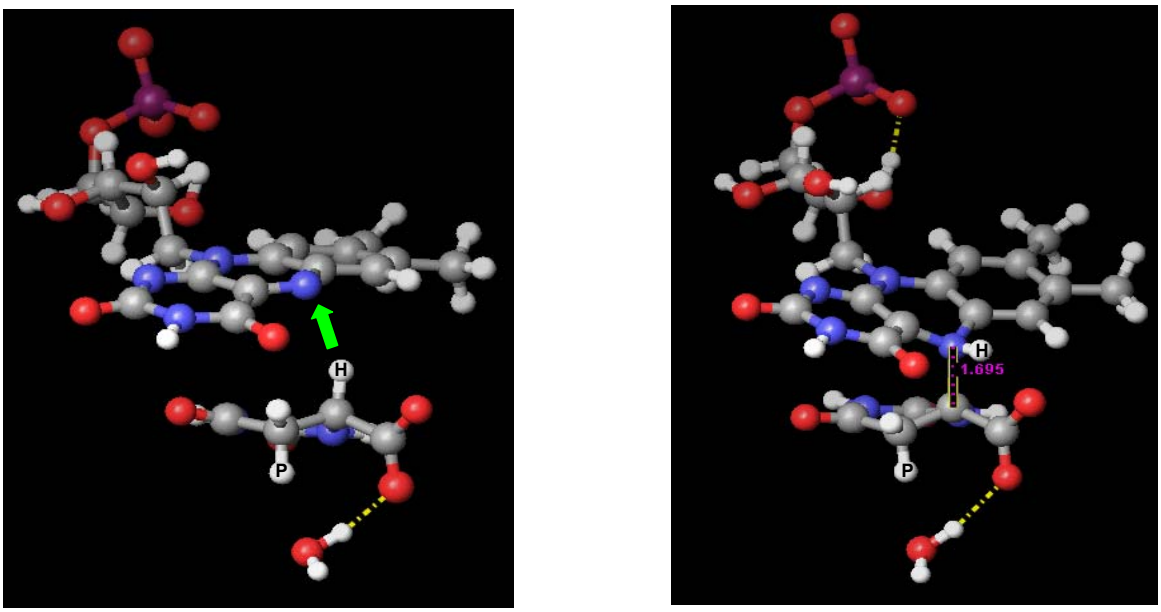

Figure B-7: Structural representation of sequential transfer where hydride moves prior to proton. Green arrow indicates direction of hydride movement. Energy difference is approximately 33 kcal/mol. (a) Reactant state (b) Hydride transfer intermediate.

(a) (b)

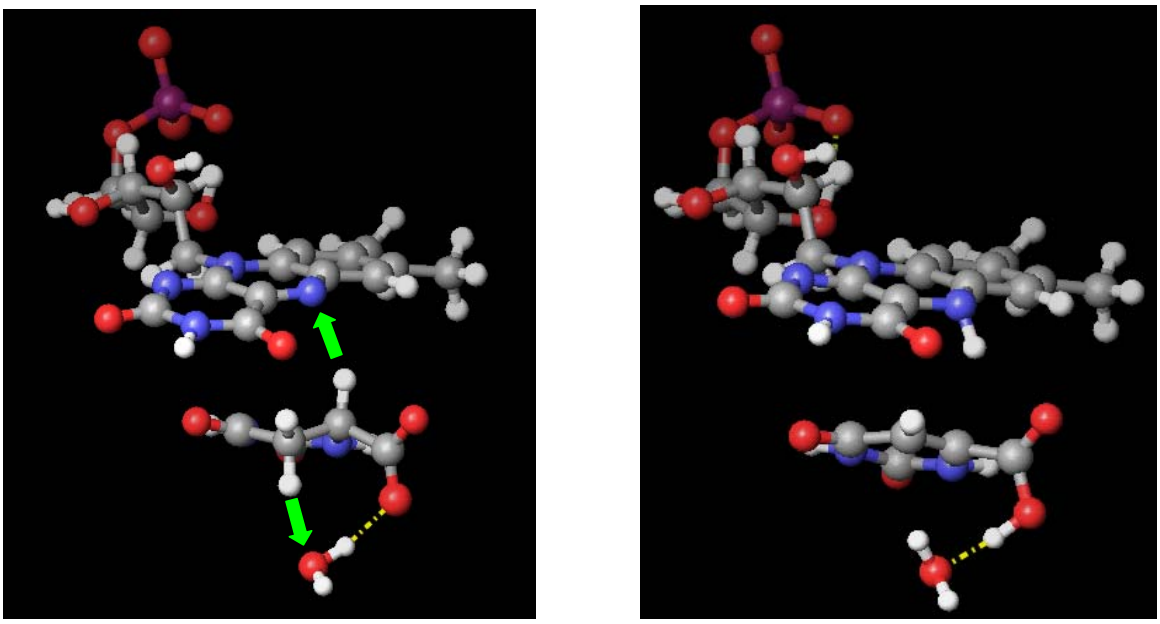

Figure B-8: Structural representation of concerted transfer where proton and hydride moves simultaneously. Green arrows indicate direction of hydrogen atom movement. Energy difference is approximately 50 kcal/mol. (a) Reactant state (b) Product state.

Appendix C

Hydrogen bond analysis code and helpful scripts

C.1 Sample Input file for hydrogen bonding analysis program

Note: Data extracted from GROMACS trajectories

Column 1: Time
Column 2: Residue number
Column 3: Residue name
Column 4: Atom number
Column 5: Atom type
Column 6: Distance from center of mass of 3 atoms (oxygens on Ser, Thr, COO-)
Column 7: units in nanometers
Column 8: x-coordinate of atom
Column 9: y-coordinate of atom
Column 10: z-coordinate of atom

```
    0   186   SER  1838    OG  0.279688 (nm) 3.16915  2.3122  3.28673
    0   186   SER  1839    HG  0.190319 (nm) 3.26261  2.29041  3.31487
    0   189   THR  1864   OG1  0.355721 (nm) 3.31983  1.95504  3.40707
    0   189   THR  1865   HG1  0.259048 (nm) 3.33506  2.03918  3.35521
    0   369   DHO  3593   O71  0.167696 (nm) 3.54394  2.32793  3.1495
    0   369   DHO  3595   O72  0.344596 (nm) 3.73457  2.42747  3.20159
    0   386   SOL  3644    OW  0.108906 (nm) 3.40237  2.20723  3.35041
    0   386   SOL  3645   HW1  0.0253343 (nm) 3.45513  2.25844  3.28263
    0   386   SOL  3646   HW2  0.183231 (nm) 3.44519  2.21685  3.44027
0.025   186   SER  1838    OG  0.278569 (nm) 3.16729  2.30445  3.28064
0.025   186   SER  1839    HG  0.189869 (nm) 3.25939  2.2781  3.30936
0.025   189   THR  1864   OG1  0.344251 (nm) 3.32149  1.96404  3.40753
0.025   189   THR  1865   HG1  0.249482 (nm) 3.34219  2.05901  3.38402
0.025   369   DHO  3593   O71  0.166322 (nm) 3.54285  2.32256  3.14788
0.025   369   DHO  3595   O72  0.334077 (nm) 3.72547  2.40756  3.18667
0.025   386   SOL  3644    OW  0.111425 (nm) 3.40568  2.19949  3.34933
0.025   386   SOL  3645   HW1  0.0380871 (nm) 3.46625  2.23445  3.27785
0.025   386   SOL  3646   HW2  0.18909 (nm) 3.45541  2.19498  3.43597
 0.05   186   SER  1838    OG  0.282845 (nm) 3.16268  2.2967  3.28169
 0.05   186   SER  1839    HG  0.200295 (nm) 3.24875  2.2561  3.31241
 0.05   189   THR  1864   OG1  0.343832 (nm) 3.33015  1.9659  3.41393
```

The following programs used in the analysis portion of this work have been attached for reference. **NOTE**: These programs are open-source and are not bound by the copyright rules of the rest of this document. Working versions of these programs can be copied from /HOME/Public/ysmall on the shsgroup network.

C.2 hbond_ysmall.f90

C.3 get_doc.f90

C.4 get_doc2.f90

C.5 get_PercHBond.f90

C.6 com_dist.f90

C.7 lifetime.f90

C.8 Sample_simulated_annealing.inp

C.9 vmd_hbond.tcl

C.10 vmd_rgyr.tcl

C.11 vmd_rmsd.tcl

```
 1: program build_triads
 2:
 3: ! program description
 4: ! This program reads lines with x columns of variables (time, index,
 5: ! and atomtype, x, y, z coordinates)
 6: ! calculates valid hydrogen bond triads
 7: ! r = rows
 8: ! c = columns
 9: !-------------------------------------------------------------------------
10: ! NOTE: THIS CODE HAS BEEN MANIPULATED FOR PRINTING PURPOSES
11: ! AND THEREFORE DOES NOT REFECT THE FORM OF IDEAL CODING PRACTICES
12: ! LINES HAVE BEEN SHORTENED TO FIT WITHIN THE PAGE
13: ! THIS CODE IS OPEN-SOURCE AND IS NOT SUBJECT TO THE COPYRIGHT
14: ! RULES OF THE REST OF THIS THESIS
15:
16: !-------------------------------------------------------------------------
17: ! Declarations
18:
19:   use hb_dlprotein_mod
20:
21:   implicit none
22:   integer,parameter :: MAXDATA=50
23:   character (LEN=30) :: input_file, output_HBpaths, output_ari
24:   character (LEN=30) :: out_percent, out_hist
25:   integer :: i, j, k, kk, pp, ios, ierr, r, p, q, iostat
26:   integer :: n_lines, HOH, HOH2
27:   integer :: records, bins, timesteps
28:   integer :: NCOL
29:   integer :: pathBULK, pathCOO, pathDUAL, dirCOO, dirBULK
30:   logical :: valid_line
31:   real (8) :: time_last, time_this, rcut_dist, rcut_angle, perc_COO
32:   real (8) :: perc_dirCOO,perc_dirBULK, perc_BULK, perc_DUAL
33:
34:   integer :: lresnum, latom_index, tresnum, tatom_index
35:   integer :: iresnum(MAXDATA),iatom_index(MAXDATA), input_triad(MAXDATA,3)
36:   integer :: input_coord(3,MAXDATA)
37:
38:   character(LEN=5)  :: lresname, latom_type, lunit, tunit, tresname
39:   character(LEN=5)  :: atom_type, iunit(MAXDATA),iresname(MAXDATA)
40:   character(LEN=5)  :: iatom_type(MAXDATA)
41:
42:
43:   real (8) :: lcomdist, tcomdist, icomdist
44:   real(8), dimension(3) :: lcoordXYZ, tcoordXYZ
45:   real(8), dimension(MAXDATA,3) ::  icoordXYZ
46:
47:   integer                  :: n_A, n_B, n_C, n_D, n_E, n_F, n_H, n_J, n_K
48:   integer                  :: n_input_triad
49:   integer, DIMENSION(40,3) :: arrayA, arrayB, arrayC, arrayD, arrayE, arrayF
50:   integer, DIMENSION(40,3) :: arrayH, arrayK, arrayJ
51:
52: !-------------------------------------------------------------------------
53: ! Initializations
54:   valid_line = .TRUE.
55:
56: !-------------------------------------------------------------------------
57: ! Ask for input files
58: ! Maximum of 3 attempts allowed
59:   ReadFile:
60:   DO i=1,3
61:    PRINT *, "Enter name of INPUT file only and 3 output filenames"
62:    READ (*,*) input_file, output_HBpaths, output_ari, out_percent, out_hist
63:    OPEN(unit=11,status='old',file=input_file,action='read',IOSTAT=ios)
64:    rewind(11)
65:    OPEN(unit=20,position='append',file=out_HBpaths,action='write',IOSTAT=ios)
66:    OPEN(unit=30,position='append',file=out_ari,action='write',IOSTAT=ios)
67:    OPEN(unit=40,position='append',file=out_percent,action='write',IOSTAT=ios)
68:    OPEN(unit=50,position='append',file=out_hist,action='write',IOSTAT=ios)
69:
70:    IF (ios==0) EXIT
71:    PRINT *, "Error opening file - try again"
72:   ENDDO ReadFile
73: !-------------------------------------------------------------------------
74: ! Initializations
75:
76:    HOH = 0
77:    HOH2 = 0
78:    pathBULK = 0
79:    pathCOO = 0
80:    pathDUAL = 0
81:    dirCOO = 0
82:    dirBULK = 0
83:
84:    r = 0
85:    p = 0
86:    q = 0
87:    records = 0
88:    bins = 0
89:    timesteps = 0
90:    n_lines = 20
91:    rcut_dist = 0.33d0 ! (nm)
92:    rcut_angle = 35    ! (degrees)
93:
94: !-------------------------------------------------------------------------
95: ! Main body of program
96:
97: ! Read first row increment bin and number of records
98: ! set the first line to the data arrays
99:
100:   read(11,*,iostat=iostat) time_last, lresnum, lresname, latom_index,
       latom_type, lcomdist, lunit, lcoordXYZ(:)
101:       if ( iostat /= 0 ) then
```

```
102:          STOP 'Cannot READ file'
103:       end if
104:   records = records + 1
105:   bins = bins + 1
106:   iresnum(bins) = lresnum
107:   iresname(bins) = lresname
108:   iatom_index(bins) = latom_index
109:   iatom_type(bins) = latom_type
110:   icoordXYZ(bins,:) = lcoordXYZ(:)
111:
112: ! Read remaining rows and analyze arrays
113:
114:    ReadData: do
115:
116:       read(11,*,iostat=iostat) time_this, tresnum, tresname, tatom_index,
        tatom_type, tcomdist, tunit, tcoordXYZ(:)
117:
118:            if ( iostat /= 0 ) then
119:                 valid_line = .FALSE.
120:            else
121:                 records = records + 1
122:            end if
123:
124:            if ( ( time_last == time_this ) .AND. valid_line ) then
125:
126:                 bins = bins + 1
127:                 iresnum(bins) = tresnum
128:                 iresname(bins) = tresname
129:                 iatom_index(bins) = tatom_index
130:                 iatom_type(bins) = tatom_type
131:                 icoordXYZ(bins,:) = tcoordXYZ(:)
132:
133:            else
134:
135:                 write (*,*)
136:                 write (20,*) 'END OF TIME STEP: ', time_last
137:                 write (*,*) 'NUMBER OF STEPS:  ', bins
138:                 write (*,*) 'INPUT DATA........'
139:
140:                 timesteps = timesteps + 1
141:
142:                 Checking input arrays
143:                if ( bins >= 15 ) then
144:                 WriteInput:
145:                 do i = 1, bins
146:                   if ( iatom_type(i) == "OW" ) then
147:                      HOH = HOH + 1
148:                      write (50,*) records, iresnum(i),iresname(i)
149:                      write (50,*) iatom_index(i),iatom_type(i), icoordXYZ(i,:)
150:                      write (50,*) records, iresnum(i),iresname(i)
151:                      write (50,*) iatom_index(i),iatom_type(i), HOH
152:                   end if
153:                 enddo WriteInput
154:                end if
155:
156:                 write (*,*) 'Performing hbond analysis on timestep ', time_last
157:
158:                 ! COUNT NUMBER OF WATERS per TIMESTEP
159:                 do i=1, bins
160:                    if ( iatom_type(i) == "OW" ) then
161:                       HOH = HOH + 1
162:                    end if
163:                 end do
164:                       write (20,*) 'No. of waters per timestep: ', HOH
165:                       if ( HOH <= 2 .AND. HOH > 0 ) then
166:                          HOH2 = HOH2 + 1
167:                       end if
168:                 ! FINISHED COUNTING NUMBER OF WATERS per TIMESTEP
169:
170:                 ! BUILD TRIADS
171:                 write (*,*) 'start hb_sort '
172:                 CALL hb_sort(bins,iatom_type,iatom_index,rcut_dist,rcut_angle)
173:
174:                 write (*,*) 'start hb_anal '
175:                 do i = 1, bins-1
176:                    do j = i+1, bins
177:                       CALL hb_anal(i,j,bins,icoordXYZ(1:bins,1), &
178:                                    & icoordXYZ(1:bins,2),icoordXYZ(1:bins,3))
179:                     end do
180:                  end do
181:
182:                  write (*,*) 'print triads '
183:                  write (*,*) 'number of triads ', nhbf2
184:                  do k = 1, nhbf2
185:                     write (*,*) 'TRIADS ihbf2', (ihbf2(kk,k), kk=1,3)
186:                     write (*,*) 'TRIADS atom_index', iatom_index(ihbf2(1:3,k))
187:                     do i = 1,3
188:                        input_triad(k,i) = iatom_index(ihbf2(i,k))
189:                     end do
190:                  end do
191:                  ! FINISHED BUILDING TRIADS
192:                  do i=1,nhbf2
193:                     write (*,*) 'TRIADS input_triad', input_triad(i,1:3)
194:                  enddo
195:
196:                  ! DETERMINE PROTON TRANSFER PATHS
197:                  n_input_triad = size(input_triad,1)
198:                  CALL path_arrays(nhbf2, n_input_triad, input_triad, n_A, &
199:                                   & arrayA, n_B, arrayB, n_C, arrayC, n_D, &
200:                                   & arrayD, n_E, arrayE, n_F, arrayF, n_H, &
201:                                   & arrayH, n_K, arrayK, n_J, arrayJ)
202:                   ! FINISHED DETERMINING PROTON TRANSFER PATHS
```

```
203:
204:
205:        ! COUNT PATHS
206:        ! Count path with COO character
207:        ! (at each time, arrayA, arrayE but not arrayD must exist)
208:        do r = 1,n_A
209:           do p = 1,n_E
210:              if (n_D /= 0) then
211:                 do q = 1,n_D
212:                    if (arrayA(r,3)==arrayE(p,1) .AND. &
213:                    & arrayA(r,3)/=arrayD(q,1))then
214:                       pathCOO = pathCOO + 1
215:                       write (*,*) arrayA(r,:), arrayE(p,:)
216:                    end if
217:                 end do
218:              else
219:                  if ( arrayA(r,3) == arrayE(p,1)) then
220:                     pathCOO = pathCOO + 1
221:                     write (*,*) arrayA(r,:), arrayE(p,:)
222:                  end if
223:              endif
224:           end do
225:        end do
226:
227:        do r = 1, n_H
228:           dirCOO = dirCOO + 1
229:           write(50,*) "dirCOO ", dirCOO
230:           write(50,*) "arrayH ", time_last, n_H, arrayH(r,:)
231:        end do
232:
233:        ! Count path with dual character
234:        ! (at each time, arrayA, arrayE & arrayD must exist)
235:        do r = 1,n_A
236:           do p = 1,n_E
237:              do q = 1,n_D
238:                 if (arrayA(r,3)==arrayE(p,1).AND.arrayA(r,3)==arrayD(q,1))then
239:                    pathDUAL = pathDUAL + 1
240:                    write (*,*) "pathDUAL    ", pathDUAL
241:                    write (*,*) arrayA(r,:), arrayE(p,:), arrayD(q,:)
242:                 end if
243:              end do
244:           end do
245:        end do
246:
247:        ! TEST Count path in MUT traj
248:        ! (at each time, arrayC & arrayE must exist)
249:        do r = 1,n_C
250:           do p = 1,n_E
251:              if ( arrayC(r,3) == arrayE(p,1) ) then
252:                 pathDUAL = pathDUAL + 1
253:              end if
254:           end do
255:        end do
256:
257:        ! Count path with BULK character
258:        ! (at each time, arrayA, arrayD but not arrayE must exist)
259:        do r = 1,n_A
260:           do p = 1,n_D
261:              if ( n_E /= 0) then
262:                 do q = 1,n_E
263:                    if (arrayA(r,3)==arrayD(p,1) .AND. &
264:                       & arrayA(r,3)/=arrayE(q,1))then
265:                       pathBULK = pathBULK + 1
266:                       write (*,*) "pathBULK    ", pathBULK
267:                       write (*,*) arrayA(r,:), arrayD(p,:), arrayE(q,:)
268:                    end if
269:                 end do
270:              else
271:                 if ( arrayA(r,3) == arrayD(p,1) ) then
272:                    pathBULK = pathBULK + 1
273:                    write (*,*) "pathBULK    ", pathBULK
274:                    write (*,*) arrayA(r,:), arrayD(p,:)
275:                 end if
276:              endif
277:           end do
278:        end do
279:
280:        do r = 1, n_F
281:           dirBULK = dirBULK + 1
282:           write(50,*) "dirBULK ", dirBULK
283:           write(50,*) "arrayF ", time_last, n_F, arrayF(r,:)
284:        end do
285:
286:                 ! FINISHED COUNTING PATHS
287:
288:
289:                 write (*,*) 'Resetting data array..........'
290:                 iresnum(1:bins) = 0
291:                 iresname(1:bins) = " "
292:                 iatom_index(1:bins) = 0
293:                 iatom_type(1:bins) = " "
294:                 icoordXYZ(1:bins,:) = 0.0d0
295:                 CALL hb_deallocate
296:                 bins = 1
297:                 HOH = 0
298:                 iresnum(bins) = tresnum
299:                 iresname(bins) = tresname
300:                 iatom_index(bins) = tatom_index
301:                 iatom_type(bins) = tatom_type
302:                 icoordXYZ(bins,:) = tcoordXYZ(:)
303:
304:           end if
```

```
305:
306:           time_last = time_this
307:           lresnum = tresnum
308:           lresname = tresname
309:           latom_index = tatom_index
310:           latom_type = tatom_type
311:           lcoordXYZ(:) = tcoordXYZ(:)
312:
313:           if ( iostat /= 0 ) then
314:              write(*,*)
315:              write(*,*) 'END OF FILE REACHED'
316:              write(20,*) 'TOTAL # OF RECORDS: ', records
317:              close(11)
318:              close(30)
319:              exit
320:           end if
321:
322:     enddo ReadData
323:
324:     perc_dirCOO = (real(dirCOO)/real(timesteps))*100
325:     perc_dirBULK = (real(dirBULK)/real(timesteps))*100
326:     perc_COO = (real(pathCOO)/real(timesteps))*100
327:     perc_DUAL = (real(pathDUAL)/real(timesteps))*100
328:     perc_BULK = (real(pathBULK)/real(timesteps))*100
329:
330:
331:     write(20,*)
332:     write(20,*) "::::::::::::::::::::::::::::::::::::::::::::::::::"
333:     write(20,*) " Number of time steps: ", timesteps
334:     write(20,*) " % Prob of PathCOO : ", perc_COO
335:     CALL standard_dev(pathCOO,timesteps)
336:     write(20,*)
337:     write(20,*) " % Prob of PathDUAL: ", perc_DUAL
338:     CALL standard_dev(pathDUAL,timesteps)
339:     write(20,*)
340:     write(20,*) " % Prob of PathBULK: ", perc_BULK
341:     CALL standard_dev(pathBULK,timesteps)
342:     write(20,*) "::::::::::::::::::::::::::::::::::::::::::::::::::"
343:     write(20,*)
344:              close(20)
345:     write(40,*) "Input filename: ", input_file
346:     write(40,*) "::::::::::::::::::::::::::::::::::::::::::::::::::"
347:     write(40,*) " Number of time steps: ", timesteps
348:     write(40,*) " HOH2 : ", HOH2
349:     write(40,*) " % HOH2 : ", (real(HOH2)/real(timesteps))*100
350:     write(40,*)
351:     write(40,*) " PathCOO : ", pathCOO
352:     write(40,*) " % Prob of PathCOO : ", perc_COO
353:     CALL standard_dev(pathCOO,timesteps)
354:     write(40,*)
355:     write(40,*) " dirCOO : ", dirCOO
356:     write(40,*) " % Prob of dirCOO : ", perc_dirCOO
357:     CALL standard_dev(dirCOO,timesteps)
358:     write(40,*)
359:     write(40,*) " PathDUAL: ", pathDUAL
360:     write(40,*) " % Prob of PathDUAL: ", perc_DUAL
361:     CALL standard_dev(pathDUAL,timesteps)
362:     write(40,*)
363:     write(40,*) " PathBULK: ", pathBULK
364:     write(40,*) " % Prob of PathBULK: ", perc_BULK
365:     CALL standard_dev(pathBULK,timesteps)
366:     write(40,*)
367:     write(40,*) " dirBULK : ", dirBULK
368:     write(40,*) " % Prob of dirBULK : ", perc_dirBULK
369:     CALL standard_dev(dirBULK,timesteps)
370:     write(40,*) "::::::::::::::::::::::::::::::::::::::::::::::::::"
371:     write(40,*)
372:              close(40)
373:              close(50)
374:
375: end program build_triads
376:
377:
378: !!c*************************************************************
379: !!c   ROUTINE: standard_dev
380: !!c*************************************************************
381:
382: subroutine standard_dev(m,time)
383:
384: implicit none
385:
386: integer, intent(in) :: m, time
387: real(8) :: p, stdev, stderr
388:
389:    p = (dble(m)/dble(time))
390:    stdev = p * (1.d0 - p)    !-- standard deviation
391:    stderr = SQRT(stdev/dble(time))        !-- standard error
392:    write (20,*) 'STD DEV (20) ', stdev
393:    write (20,*) 'STDERR (20) ', stderr
394:    write (40,*) 'STD DEV (40) ', stdev
395:    write (40,*) 'STDERR (40) ', stderr
396:
397: end subroutine standard_dev
398:
399: !!c*************************************************************
400: !!c   ROUTINE: path_arrays
401: !!c*************************************************************
402:
403: subroutine path_arrays(num_triads, n_input_triad, input_triad, A, arrayA, &
404:                       & B, arrayB, C, arrayC, D, arrayD, E, arrayE, F, &
405:                       & arrayF, H, arrayH, K, arrayK, J, arrayJ)
406:
```

```
407: implicit none
408:
409: integer, intent(in) :: num_triads, n_input_triad
410: integer, intent(out) :: A, B, C, D, E, F, H, J, K
411: integer, DIMENSION(40,3), intent(out) :: arrayA, arrayB, arrayC
412: integer, DIMENSION(40,3), intent(out) :: arrayD, arrayE, arrayF
413: integer, DIMENSION(40,3), intent(out) :: arrayH, arrayK, arrayJ
414: integer, intent(in) :: input_triad(n_input_triad,3)
415:
416: integer :: i
417:
418: arrayA = 0.d0
419: arrayB = 0.d0
420: arrayC = 0.d0
421: arrayD = 0.d0
422: arrayE = 0.d0
423: arrayF = 0.d0
424: arrayH = 0.d0
425: arrayK = 0.d0
426: arrayJ = 0.d0
427:
428: A = 0
429: B = 0
430: C = 0
431: D = 0
432: E = 0
433: F = 0
434: H = 0
435: K = 0
436: J = 0
437:
438: Build_hbond_vec:
439: do i = 1, num_triads
440:
441:    ! Build array for triad with A character (1 2 Xo)
442:    if ( input_triad(i,1) == 1838 .AND. input_triad(i,2) == 1839) then
443:       A = A + 1
444:       arrayA(A,:) = input_triad(i,:)
445:       write (*,*) 'arrayA ', A, arrayA(A,:)
446:    end if
447:
448:    ! Build array for triad with B character (Xo Xh 1)
449:    if ( input_triad(i,3) == 1838 ) then
450:       B = B + 1
451:       arrayB(B,:) = input_triad(i,:)
452:       write (*,*) 'arrayB ', arrayB(B,:)
453:    end if
454:
455:    ! Build array for triad with C character (3 4 Xo)
456:    if ( input_triad(i,1) == 1864 .AND. input_triad(i,2) == 1865) then
457:       C = C + 1
458:       arrayC(C,:) = input_triad(i,:)
459:       write (*,*) 'arrayC ', arrayC(C,:)
460:    end if
461:
462:    ! Build array for triad with D character (Xo Xh 3)
463:    if ( input_triad(i,3) == 1864 ) then
464:       D = D + 1
465:       arrayD(D,:) = input_triad(i,:)
466:       write (*,*) 'arrayD ', arrayD(D,:)
467:    end if
468:
469:    ! Build array for triad with E character (Xo Xh 5 or Xo Xh 6)
470:    if ( input_triad(i,3) == 3593 .OR. input_triad(i,3) == 3595 ) then
471:       E = E + 1
472:       arrayE(E,:) = input_triad(i,:)
473:       write (*,*) 'arrayE ', arrayE(E,:)
474:    end if
475:
476:    ! Build array for triad with F character (1 2 3)
477:    if ( input_triad(i,1) == 1838 .AND. input_triad(i,2) == 1839 .AND. &
478:       & input_triad(i,3) == 1864 ) then
479:       F = F + 1
480:       arrayF(F,:) = input_triad(i,:)
481:       write (*,*) 'arrayF ', arrayF(F,:)
482:    end if
483:
484:    ! Build array for triad with H character (1 2 5)
485:    if ( input_triad(i,1) == 1838 .AND. input_triad(i,2) == 1839 .AND. &
486:       &( input_triad(i,3) == 3593 .OR. input_triad(i,3) == 3595) ) then
487:       H = H + 1
488:       arrayH(H,:) = input_triad(i,:)
489:       write (*,*) 'arrayH ', arrayH(H,:)
490:    end if
491:
492:    ! Build array for triad with J character (3 4 5)
493:    if ( input_triad(i,1) == 1864 .AND. input_triad(i,2) == 1865 .AND. &
494:       &( input_triad(i,3) == 3593 .OR. input_triad(i,3) == 3595) ) then
495:       J = J + 1
496:       arrayJ(J,:) = input_triad(i,:)
497:       write (*,*) 'arrayJ ', arrayJ(J,:)
498:    end if
499:
500:    ! Build array for triad with K character (3 4 1)
501:    if ( input_triad(i,1) == 1864 .AND. input_triad(i,2) == 1865 .AND. &
502:       & input_triad(i,3) == 1838 ) then
503:       K = K + 1
504:       arrayK(K,:) = input_triad(i,:)
505:       write (*,*) 'arrayK ', arrayK(K,:)
506:    end if
507: enddo Build_hbond_vec
508: end subroutine path_arrays
```

```
!************************************************************************
! This program was used to extract energy and temperature output
! from GROMACS log files to be plotted with XMGrace  (GROMACS forcefield).
! The if statements strip out extraneous text between lines in the log file
! The program also converts the character 'Step' into a real value
!************************************************************************

program get_doc
  implicit none
  character*15::Step,Temp,TotE,KE, PE
  character*100::bbb,bbb_out
  real(8) :: counter,StepReal, StepConv

  open(unit=11,status='old',file='test_1-2PS.log')
  open(unit=20,status='new',file='test_SvsTemp.txt')
  open(unit=30,status='new',file='test_SvsTotE.txt')
  open(unit=40,status='new',file='test_SvsKE.txt')
  open(unit=50,status='new',file='test_SvsPE.txt')

counter=-5.0
  inner: do
     read(11,'(a100)',end=99) bbb
     if(bbb(1:15)=='      Potential')then
        read(11,'(a100)',end=99)bbb
        counter=counter+5
        StepConv=counter*0.001
                if (counter == 20005) then
                   read(11,'(a100)',end=99)bbb
                end if
                if (counter == 20010) then
                   read(11,'(a100)',end=99)bbb
                   counter=-5.0
                end if
        PE = (bbb(1:15))
        KE   = (bbb(19:30))
        TotE = (bbb(34:46))
        Temp = (bbb(49:61))
           write(20,'(a15)') Temp
           write(20,'(F17.1,5X,a15)') counter,Temp
           write(30,'(F17.1,5X,a15)') counter,TotE
           write(40,'(F17.1,5X,a15)') counter,KE
           write(50,'(F17.1,5X,a15)') counter,PE
     endif
  enddo inner
  close(11)
  close(20)
  close(30)
  close(40)
  close(50)
99 continue
end program get_doc
```

```
!************************************************************************
! This program was used to extract energy and temperature output from GROMACS
! log files to be plotted with XMGrace  (OPLS forcefield).
!************************************************************************

program get_doc2
  implicit none
  character*15::Step,Temp,TotE,KE, PE
  character*100::bbb,bbb_out
  real(8) :: counter,StepReal, StepConv

  open(unit=11,status='old',file='DHODop_consEq0.all.log')
  open(unit=20,status='new',file='DHODop_consEq0_TempTotEKePe.txt')

counter=-5.0

  inner: do
     read(11,'(a100)',end=99) bbb

     if(bbb(1:15)=='   Coulomb (LR)')then
        read(11,'(a100)',end=99)bbb
        counter=counter+5
        StepConv=counter*0.001
        if (counter == 10005) then
            read(11,'(a100)',end=99)bbb
        end if
        if (counter == 10010) then
            read(11,'(a100)',end=99)bbb
            counter=-5.0
        end if
        PE   = (bbb(19:30))
        KE = (bbb(34:46))
        TotE = (bbb(49:61))
        Temp = (bbb(64:75))

        write(20,'(a15,5X,a15,5X,a15,5X,a15)') Temp, TotE, KE, PE

     endif

  enddo inner

  close(11)
  close(20)

99 continue

end program get_doc2
```

```
!*************************************************************************
! This program was used to extract percentage of hydrogen bonds from GROMACS
! trajectory files to be plotted with XMGrace  (not forcefield specific).
!  This file reads the 5-column output of distances and
!  calculates the percentage of h-bonding to molecule
!*************************************************************************

program get_PercHBond

  implicit none
  character (LEN=30) :: input_file
  character (LEN=30) :: output_file
  integer :: i, ios, t_lines, ser, thr, o71, o72
  integer :: ser_count, thr_count, o71_count, o72_count
  integer :: path_ser_thr, path_ser_coo, path_dual
  real(8) :: time, PercHB_ser, PercHB_thr, PercHB_o71, PercHB_o72
  real(8) :: PercHB_pathST, PercHB_pathSC, PercHB_pathDual

!-------------------------------------------------------------
! Ask for name of input and output files
! Maximum of 3 attempts allowed
  DO i=1,3
     PRINT *, "Enter name of INPUT file then OUTPUT file"
     READ (*,*) input_file, output_file
     OPEN(unit=11,status='old',file=input_file,IOSTAT=ios)
     OPEN(unit=15,status='old',file=output_file,position='append')

     IF (ios==0) EXIT
     PRINT *, "Error opening file - try again"
  END DO
!-------------------------------------------------------------

  t_lines = 0
  ser_count = 0
  thr_count = 0
  o71_count = 0
  o72_count = 0
  path_ser_thr = 0
  path_ser_coo = 0
  path_dual = 0

  inner: do
   read(11,*,end=99) time, ser, thr, o71, o72

    t_lines = t_lines + 1

    if( ser /= 0) ser_count = ser_count + 1
    if( thr /= 0) thr_count = thr_count + 1
    if( o71 /= 0) o71_count = o71_count + 1
    if( o72 /= 0) o72_count = o72_count + 1

    if( ser /= 0 .AND. thr /= 0 .AND. o71 == 0) &
       & path_ser_thr = path_ser_thr + 1
    if( ser /= 0 .AND. thr == 0 .AND. o71 /= 0) &
       & path_ser_coo = path_ser_coo + 1
    if( ser /= 0 .AND. thr /= 0 .AND. o71 /= 0) &
       & path_dual = path_dual + 1

  enddo inner

  close(11)

99 continue

   time = (t_lines*0.025)/1000
   PercHB_ser = 100.d0*real(ser_count)/real(t_lines)
   PercHB_thr = 100.d0*real(thr_count)/real(t_lines)
   PercHB_o71 = 100.d0*real(o71_count)/real(t_lines)
   PercHB_o72 = 100.d0*real(o72_count)/real(t_lines)

   PercHB_pathST = 100.d0*real(path_ser_thr)/real(t_lines)
   PercHB_pathSC = 100.d0*real(path_ser_coo)/real(t_lines)
   PercHB_pathDual = 100.d0*real(path_dual)/real(t_lines)

!----------Print to screen-------------------------------
 write(*,*) input_file
   write(*,'(a15,5X,a10,5X,a10,5X,a10,5X,a10)') "Total Time (ns)", &
        & "PercHB_ser", "PercHB_thr", "PercHB_o71", "PercHB_o72"
   write(*,'(F10.5,7X,F10.5,7X,F10.5,5X,F10.5,5X,F10.5)') time, &
        & PercHB_ser, PercHB_thr, PercHB_o71, PercHB_o72

   write(*,'(a15,5X,a10,5X,a10,5X,a10)') "Total time (ns)", &
        &  "pathST", "pathSC", "pathDual"
   write(*,'(F10.5,7X,F10.5,7X,F10.5,5X,F10.5,5X,F10.5)') time, &
        & PercHB_pathST, PercHB_pathSC, PercHB_pathDual

!-----------Print to file--------------------------------

   write(15,'(a15,5X,a30)') "Input Filename: ", input_file
   write(15,'(a15,5X,a10,5X,a10,5X,a10,5X,a10,a15,5X,a10,5X,a10,5X,a10)') &
         & "Total Time (ns)", "PercHB_ser", "PercHB_thr", "PercHB_o71",   &
         & "PercHB_o72", "pathST", "pathSC", "pathDual"
   write(15,'(F10.5,7X,F10.5,7X,F10.5,5X,F10.5,5X,F10.5,F10.5,7X,F10.5,7X,&
         & F10.5,5X,F10.5,5X,F10.5)') time, PercHB_ser, PercHB_thr,       &
         & PercHB_o71, PercHB_o72, PercHB_pathST, PercHB_pathSC,          &
         & PercHB_pathDual
   write(15,*)

   close(15)

end program get_PercHBond
```

```
205: !*************************************************************************
206: ! Program to calculate the distance from the center of mass of an atom.
207: ! This file reads the 7-column output of center of mass in xyz coordinates
208: !*************************************************************************
209:
210: program com_dist
211:
212:   implicit none
213:   character*100::ReadLine
214:   real(8) :: time, comX, comY, comZ, serX, serY, serZ, thrX, thrY, thrZ
215:   real(8) :: cooX, cooY, cooZ, serA, serB, serC, thrA, thrB, thrC, cooA
216:   real(8) :: cooB, cooC, timeNS, serDist, serDistance
217:   real(8) :: thrDist, thrDistance, cooDist, cooDistance, ang1, ang2, ang3
218:
219:    open(unit=11,status='old',file='coord.xvg')
220:    open(unit=20,status='new',file='CoordDistAll.txt')
221:
222:   inner: do
223:     read(11,*,end=99) time, comX, comY, comZ, serX, serY, &
224:         &  serZ, thrX, thrY, thrZ, cooX, cooY, cooZ
225:
226:     serA = (comX - serX)*(comX - serX)
227:     serB = (comY - serY)*(comY - serY)
228:     serC = (comZ - serZ)*(comZ - serZ)
229:          serDist = sqrt(serA + serB + serC)
230:          serDistance = abs(serDist)
231:          timeNS = (time/1000)
232:     thrA = (comX - thrX)*(comX - thrX)
233:     thrB = (comY - thrY)*(comY - thrY)
234:     thrC = (comZ - thrZ)*(comZ - thrZ)
235:          thrDist = sqrt(thrA + thrB + thrC)
236:          thrDistance = abs(thrDist)
237:     cooA = (comX - cooX)*(comX - cooX)
238:     cooB = (comY - cooY)*(comY - cooY)
239:     cooC = (comZ - cooZ)*(comZ - cooZ)
240:          cooDist = sqrt(cooA + cooB + cooC)
241:          cooDistance = abs(cooDist)
242:    write(*,*)
243:    write(20,'(F17.5,5X,F17.5,5X,F17.5,5X,F17.5)') timeNS, &
244:          & serDistance, thrDistance, cooDistance
245:
246:   enddo inner
247:   close(11)
248:   close(20)
249: 99 continue
250:
251: end program com_dist
252:
253:
254:
255:
```

```
256: !*************************************************************************
257: ! This program calculates hydrogen bonding lifetimes
258: ! The hbond list file is read and the average lifetime per hbond
259: ! are calculated
260: !*************************************************************************
261:
262: program lifetime
263:
264:   implicit none
265:   character*100::bbb
266:   integer :: n_ones, n_zeros, n_cut, id, id_prev
267:   real(8) :: n_blocks, life_time
268:
269:   open(unit=11,status='old',file='get_lifetime_ab.txt')
270:   open(unit=20,status='new',file='get_lifetime_ab.log')
271:
272: n_ones = 0
273: n_zeros = 0
274: n_cut = 0
275:
276:   do
277:
278:      read(11,*,end=99) id
279:
280:      if( id_prev /= id) then
281:         n_cut =  n_cut + 1
282:         id_prev = id
283:      end if
284:
285:      if( id == 1) then
286:         n_ones = n_ones + 1
287:         id_prev = id
288:
289:         n_blocks = (n_cut + 1)/2
290:         life_time = (n_ones * 0.025)/n_blocks
291:
292:         write(20,'(I7.1,5X,F7.3)') n_ones, life_time
293:         write(*,*) n_ones, life_time
294:
295:       end if
296:   enddo
297:
298:   close(11)
299:   close(20)
300:
301: 99 continue
302:
303: end program lifetime
304:
305:
306:
```

```
!*****************************************************************************
! Sample IMPACT input script to automate a simulated annealing run
! Repeat these blocks in the file for the desired number of steps.
! NOTE: the files are saved in 20 ps chunks.
! NOTE: the protein and solvent are coupled to the temperature bath separately
.
!*****************************************************************************

The general order of commands (heat, equilibrate, cool):
! 200 steps of steepest decent minimization
! 300 steps of conjugate gradient minimization
! 2 ps of MD at zero K
! 2 ps of MD at 50 K
! 2 ps of MD at 100 K
! 2 ps of MD at 150 K
! 2 ps of MD at 200 K
! 2 ps of MD at 250 K
! 40 ps of MD at 300 K
! 2 ps of MD at 250 K
! 2 ps of MD at 200 K
! 2 ps of MD at 150 K
! 2 ps of MD at 100 K
! 2 ps of MD at 50 K
! 2 ps of MD at zero K

MINIMIZE
  steepest dx0 5.000000e-02 dxm 1.000000e+00
  input cntl mxcyc 200 rmscut 1.000000e-02 deltae 1.000000e-01
  run
  write maestro file -
"filename_steepmin_out.mae"
! saves coord, veloc and box size
   write restart coordinates and velocities box -
     formatted real8 file "filename_steepmin.rst"
QUIT

MINIMIZE
  conjugate dx0 5.000000e-02 dxm 1.000000e+00
  input cntl mxcyc 300 rmscut 1.000000e-02 deltae 1.000000e-01
  read restart coordinates and velocities box formatted -
     file "filename_steepmin.rst"
  run
  write maestro file -
"filename_congjmin_out.mae"
   write restart coordinates and velocities box -
     formatted real8 file "filename_congjmin.rst"
QUIT

DYNAMICS
  input cntl nstep 2000 delt 0.001
  input cntl constant temperature byspecies relax 0.01
  input target temperature 0 name species1 temperature 0 name species2
  read restart coordinates and velocities box formatted -
     file "filename_congjmin.rst"
  input cntl initialize temperature forspecies name species1 at 0 seed 1704616
  input cntl initialize temperature forspecies name species2 at 0 seed 1504518
  input cntl nprnt 5
  input cntl tol 1e-07
  input cntl stop rotations
  input cntl statistics on
  run verlet
  write maestro file -
"filename_heatzero_out.mae"
  write restart coordinates and velocities box -
     formatted real8 file "filename_heatzero.rst"
QUIT
```

```
!*****************************************************************************
! Scripts written in Tcl language for doing hydrogen bonding analyses'
!*****************************************************************************


set nf [molinfo top get numframes]
set sel1 [atomselect top "index 2884 or index 6370"]
puts "Processing $nf frames..."
set cutoff 3.0
set angle 30.0
set abond1 "2884 6370 2889"
set abond2 "{5845 6229} {6229 2923} {5847 6231}"
set nonbond "{} {} {}"
set chan1 [open test_output.log a]
set chan2 [open 2884_Tcl_xmgrHB.txt  a]
for {set frame 0} {$frame < $nf} {incr frame 1} {
          $sel1 frame $frame
          $sel1 update
          set sel2 [measure hbonds $cutoff $angle $sel1]
          puts $chan1 "h-bonds for frame $frame is $sel2"
          if {$sel2 == $abond1} {puts $chan2 "$frame      1"}
          if {$sel2 == $nonbond} {puts $chan2 "$frame     0"}
          if {$sel2 == $abond2} {puts $chan2 "$frame      34"}
}
close $chan1
close $chan2




!*****************************************************************************
set nf [molinfo top get numframes]
set sel1 [atomselect top "resid 189 or resid 555"]
puts "Processing $nf frames..."
set cutoff 3.0
set angle 30.0
set chan [open test_output.log a]
for {set frame 0} {$frame < $nf} {incr frame 1} {
          $sel1 frame $frame
          $sel1 update
          set sel2 [measure hbonds $cutoff $angle $sel1]
          puts $chan "h-bonds for frame $frame is $sel2"
}
close $chan




!*****************************************************************************
! Scripts written in Tcl language to calculate radius of gyration
!*****************************************************************************

et sel_rgyr [atomselect top "protein"]
set nf [molinfo top get numframes]
puts "Processing $nf frames..."
set chan1 [open test_rgyr.log a]
set chan2 [open Tcl_RGYR_Gmx_test_orig.log a]
for {set frame 0} {$frame < $nf} {incr frame 1} {
          $sel_rgyr frame $frame
          $sel_rgyr update
          set calc_rgyr [measure rgyr $sel_rgyr weight mass]
          puts $chan1 "RGYR of frame $frame is $calc_rgyr"
          puts $chan2 "$frame    $calc_rgyr"
}
close $chan1
close $chan2







!*****************************************************************************
! Scripts written in Tcl language to calculate Root Mean Square displacement
!*****************************************************************************


set reference [atomselect top "backbone"]
set compare [atomselect top "backbone"]
set nf [molinfo top get numframes]
puts "Processing $nf frames..."
set chan1 [open test_Gmx_test_orig.log a]
set chan2 [open Tcl_RMSD_Gmx_test_orig.log a]
for {set frame 0} {$frame < $nf} {incr frame 1} {
          $compare frame $frame
          $compare update
          set sel1 [measure fit $compare $reference weight mass]
          $compare move $sel1
          set rmsd [measure rmsd $compare $reference weight mass]
          puts $chan1 "RMSD of frame $frame is $rmsd"
          puts $chan2 "$frame    $rmsd"
}
close $chan1
close $chan2
```

VITA

Yolanda A. Small

Education

May 1998: B.S. in Applied Mathematics and Chemistry, University of Houston, Houston, Texas 77005

Publications

Y. A. Small, V. G. Guallar, A. V. Soudackov and S. Hammes-Schiffer. "Hydrogen Bonding Pathways in Human Dihydroorotate Dehydrogenase." J. Phys. Chem. B 110(39), 19704-19710.

Contributed Talks

Y. A. Small, "Mechanistic Investigation of Enzymes using QM/MM Methods", American Physical Society Annual Meeting, Montreal, Canada March 2004

Y. A. Small, "QM/MM study of an enzyme comparing the wild type to the mutant", CCP5 and Marie Curie Actions: Methods in Molecular Simulation, Cardiff, Wales UK, July 2004

Invited Talks

Y. A. Small, "Hydrogen Bonding Pathways and Hydrogen Transfer in Dihydroorotate Dehydrogenase", Gordon Research Conference on Computational Chemistry, Les Diablerets, Switzerland October 2006

Poster Presentations

Y. A. Small, S. Hammes-Schiffer, "Theoretical Study of the Proton Transfer Mechanism in the Dihydroorotate Dehydrogenase (DHOD) Enzyme", Graduate Research Exhibition, State College, April 2003

Y. A. Small, S. Hammes-Schiffer, "Quantum mechanical/molecular mechanical (QM/MM) study of the mechanism of dihydroorotate dehydrogenase", ACS National Meeting, Philadelphia, PA, August 2004

Y. A. Small, V. Guallar, S. Hammes-Schiffer, "Exploring a possible proton relay pathway and the influence of mutations in dihydroorotate dehydrogenase using classical molecular dynamics", ACS National Meeting, Washington, D.C, August 2004

Y. A. Small, V. Guallar, S. Hammes-Schiffer, "Quantum mechanical/molecular mechanical (QM/MM) study of a hydride transfer reaction in the enzyme dihydroorotate dehydrogenase", Gordon Research Conference on Computational Chemistry, Les Diablerets, Switzerland, October 2006