DESIGN OF A NOVEL INHIBITOR FOR *HELICOBACTER PYLORI* UREASE

By

MONA SAMER MINKARA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

11

LIST OF SCHEMES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DESIGN OF A NOVEL INHIBITOR FOR *HELICOBACTER PYLORI* UREASE

By

Mona Samer Minkara

August 2015

Chair: Kenneth Malcolm Merz,Jr.
Cochair: Erik Deumens
Major: Chemistry

The bacterium *Helicobacter pylori* has been implicated in numerous gastrointestinal diseases and disorders. Approximately two thirds of the world's population is infected with *H. pylori,* and current treatment options function poorly against advanced infections. In order to survive in acidic conditions, the bacterium produces the urease enzyme (*H. pylori* urease, HP urease), which hydrolyzes urea into ammonium and carbamate ions. Ammonium is then released into the environment as ammonia, increasing the local pH and allowing the survival of *H. pylori* in the stomach lining. This thesis aims to identify or design a novel inhibitor of HP urease as a means to effectively combat infections. First, we sought to gain a better understanding of the system of HP urease through the use of molecular dynamics. Second, we docked the ligands found in the ZINC database in order to uncover prospects for inhibition. Regarding the molecular dynamics of HP urease, we conducted studies in three different environments. In an aqueous solution, we observed the wide open flap state discussed earlier by Roberts et al., in which a similar flap state observed in *Klebsiella aerogenes* (KA) urease was described. In a simulation with 10.5 M urease, we observed certain residues with high RMSF that acted as a net, shuttling urease molecules into the active site; this suggested another possible means of inhibiting the enzyme. In a 150 mM ammonium chloride simulation, we observed the possible mechanism by

which ammonium was expelled into the environment and deprotonated, which also suggested

possible sites for inhibition. Concerning our docking studies, we docked the 14 million structures

within the ZINC database and obtained possible options. We tested these options experimentally

on KA urease to determine inhibitory activity; however, high levels of inhibition were not

observed. After these various studies were performed to understand what might qualify as a good

inhibitor, a quantum mechanical study on urea and its interactions with all individual amino

acids was conducted. Though some trends were uncovered, nothing distinctive was confidently

observed. We hope that the work within this thesis brings us closer to the discovery of an

inhibitor for *H. pylori.*

CHAPTER 1
INTRODUCTION AND BACKGROUND OF *HELICOBACTER PYLORI* UREASE

The underlying causes of ulcers remained elusive for decades, with initial speculation suggesting that they were caused by stress. However, when the *Helicobacter pylori* bacterial strain was first discovered in 1982, Australian researcher Barry Marshall hypothesized that these bacteria may be linked to gastritis. In a 1984 experiment, Marshall consumed a petri dish of the bacteria and within days developed symptoms consistent with gastritis.[1] These bacteria commonly infect the lining of the stomach (Figure 1-1). Subsequently, the bacteria were linked to gastric and duodenal ulcers, hepatic coma (liver failure), non-ulcer dyspepsia and adenocarcinoma (stomach cancer), along with other potentially fatal diseases.[2]
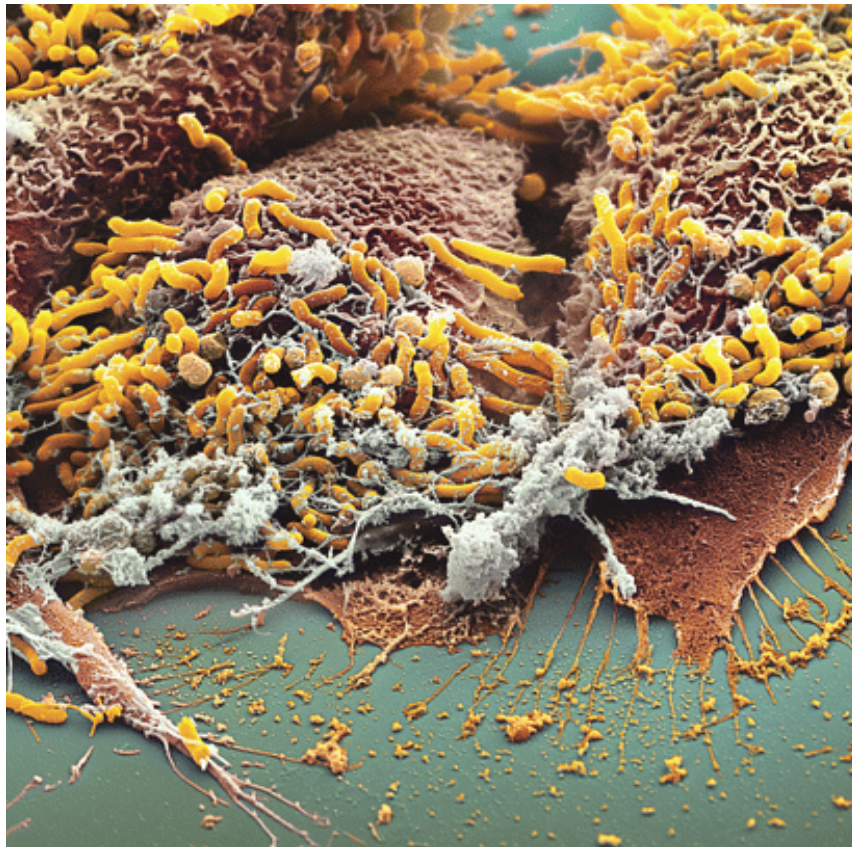
Figure 1-1. Stomach infected with *Helicobacter pylori*. Image courtesy of Martin Oeggerli (Micronaut and School of Applied Sciences Northwestern Switzerland) and featured as the cover image of the September 6, 2011 issue of *Proceedings of the National Academy of Sciences* (http://www.pnas.org/content/108/36.cover-expansion)

There exists an obvious healthcare and disease prevention gap between the richer, more developed nations and the poorer, underdeveloped nations and regions. In these underdeveloped areas, disease and infection run rampant throughout the population, whereas individuals living in more affluent areas rarely experience these diseases at the same frequency. It has been estimated by many outlets that approximately half of the world's population is infected with *Helicobacter pylori*, and the World Health Organization places its estimate at two-thirds. It is further estimated that around 4 percent of those infected die of diseases caused by these colonies. Though the percentage itself is low, 4 percent of half of the world's population still represents a vast number of potential fatalities. Figure 1-2 displays the percentage of the population infected with *H. pylori* based on geography.

The goal of preventing these fatalities creates a dire need for efficient treatments for these possibly fatal infections. One of the most difficult aspects of eradicating *H. pylori* from one's system involves the environment in which it resides; by colonizing within the stomach lining, *H. pylori* is surrounded by an acidic environment, one that confers a natural defense against available antibiotics. Specifically, the high acidity contributes to the degradation of antibiotics before they are able to effectively act upon a well-developed *H. pylori* colony. After the discovery that *H. pylori* was the main cause of ulcers, more work was done to understand the mechanism by which *H. pylori* survives in such acidic conditions. It was subsequently discovered that the enzymatic hydrolysis of urea was a critical pH-raising process for these bacterial colonies. Urease [urea amidohydrolase EC 3.5.1.5] is an enzyme that hydrolyses urea into ammonium and carbamate ions (Scheme 1-1), thereby increasing the pH of the surrounding system. In doing so, it counteracts the acidity of the stomach, creating a neutral environment that allows *H. pylori* to survive.
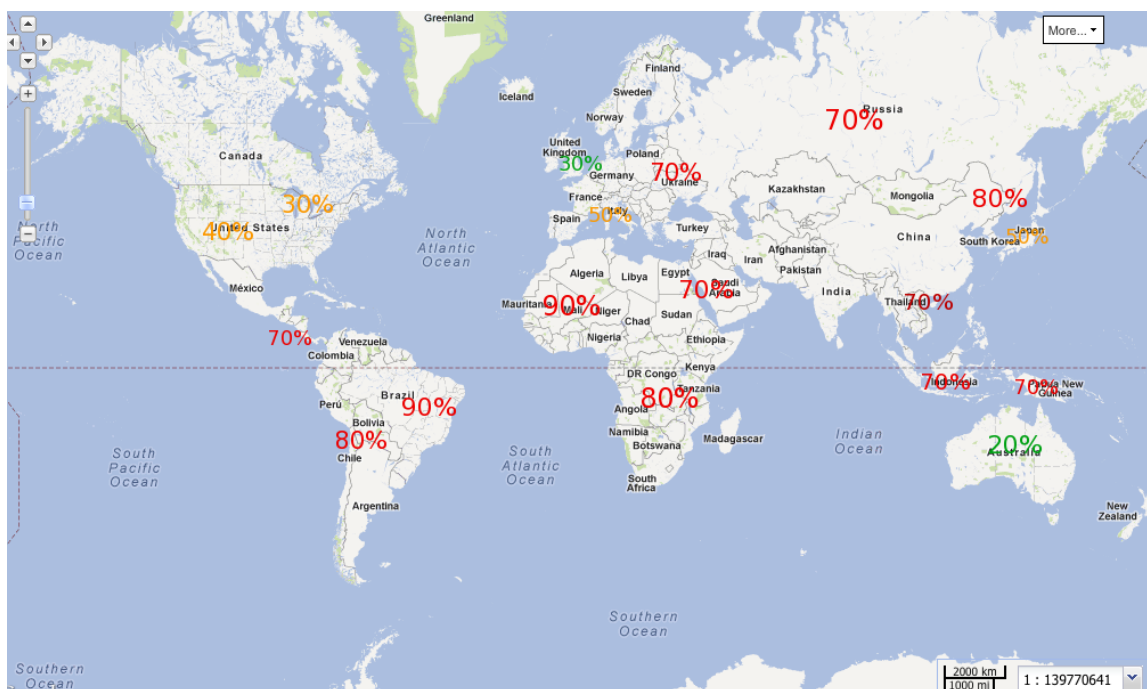
Figure 1-2. Percentage of individuals in each region infected with *H. pylori* Bioinformotics
    Consortium Taiwan, http://cbs.ym.edu.tw/cbs-
    01/index.php?option=com_content&view
    =article&id=306&Itemid=332
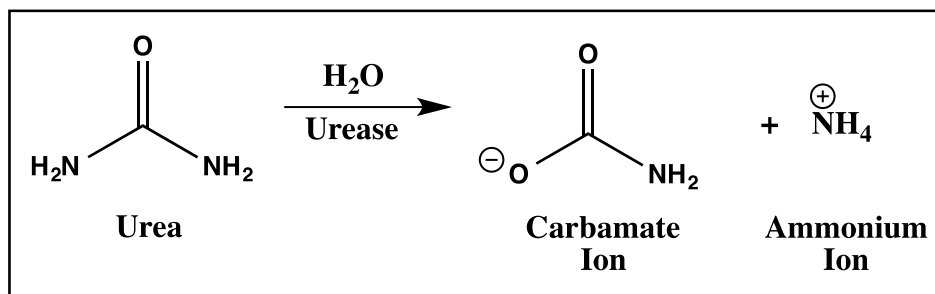
Scheme 1-1. Products of enzymatic urea hydrolysis.

Urea is a molecule that is exceptionally stable within the pH range of 2-12 and is present

in the stomach in 1.7-3.4 mM concentrations.[3] The stability of urea arises from the 30-40

kcal/mol of resonance stabilization as a result of nitrogen lone pair delocalization into the

carbonyl π^* orbital. While the rate of decomposition of urea is quite slow, with a half-life of 3.6

years at 38 °C, urease is able to enhance the rate of decomposition by approximately 10^{14}-fold.

Despite the small concentrations of urea present in the stomach, urease is able to effectively

utilize these amounts due to its high efficiency, with a Michaelis constant (K_M) of 0.17 mM.[4]

The means by which the nonenzymatic decomposition of urea proceeds are not entirely

clear, though two plausible mechanisms do exist (Scheme 1-2). Both cases involve

decomposition into cyanate and ammonium ions through elimination pathways, though each

approach these products in slightly deviant fashions. In the first mechanism direct elimination of

the ammonium ion occurs through intramolecular rearrangement of the molecule; one amine

group becomes protonated through following deprotonation of the second amine group, and is

eliminated as ammonia. The free ammonia species is then further protonated by removal of a

proton from cyanic acid, producing cyanate and ammonium. The second mechanism differs in

that tautomerization occurs first, converting urea into a species featuring an imidol functional

group. From here, elimination echoing the first mechanism occurs, followed by a similar

protonation of the ammonia via cyanic acid.
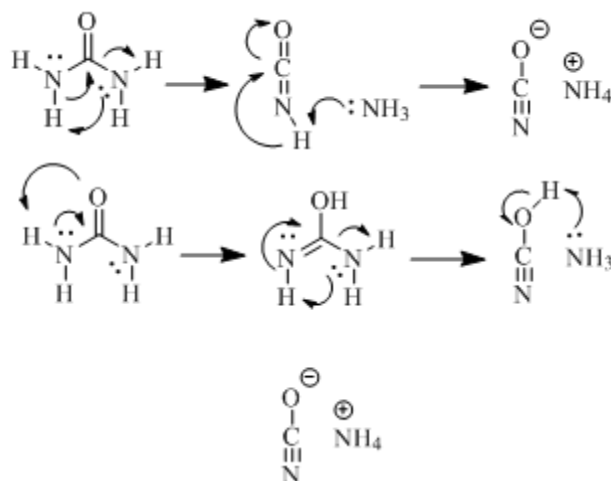
Scheme 1-2. Nonenzymatic urea decomposition by elimination (top) and rearrangement
        (bottom).

When catalyzed by urease, the decomposition of urea follows a different mechanism

altogether, proceeding via hydrolysis and acid-base catalysis rather than direct elimination

according to the reaction schematic presented by Karplus et al. (Scheme 1-3). This results in the

production of carbamate rather than cyanic acid. The reaction begins via coordination of both

urea and a hydroxyl group (derived from water) with the Ni^{2+} ion pair. Attack of the central

carbon by the coordinated hydroxyl group produces a tetrahedral carbonyl addition intermediate,
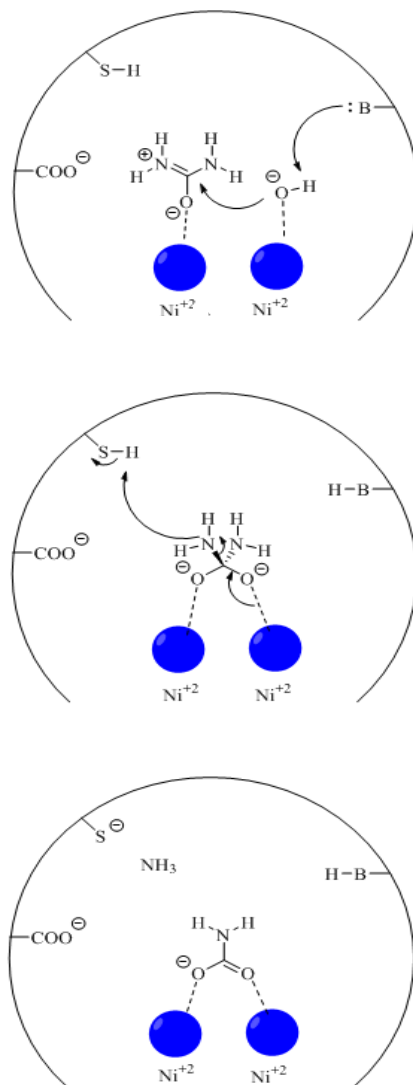
Scheme 1-3. Enzymatic decomposition of urea within the generalized active site.

which then eliminates an amide ion which is protonate by the sidechain thiol of a cysteine within the active site. The thiol and general base are subsequently regenerated prior to the next catalytic reaction.

It wasn't until the pioneering 1926 work of Sumner in which the first enzyme, urease, was crystallized.[5] It was assumed until this point that enzymes could not be crystallized. After the first crystallized enzyme – Jack Bean urease – was produced, other types of urease were crystallized, including *Klebsiella aerogenes* (KA) and others. It was not until recently that HP urease was crystallized. First reported by Ha et al. in 2001,[6] urease is essential for the survival of HP in the acidic environment of the human stomach. As discussed elsewhere in the life cycle, the presence of urease is crucial to the survival of the bacterial colonies' regulation of environmental pH. Urease is an enzyme that hydrolyzes urea into ammonium and carbamate, utilizing an active site that is quite unique among enzymes in that nickel ions serve as the catalytic centers – much like the aforementioned Jack Bean urease, which was identified as the first nickel-containing enzyme in 1977.[7] Nickel is not the sole metal to appear in the ureases; for example, in the *Helicobacter mustelae* urease, the active site contains iron ions.[8] HP urease specifically is a very large enzyme, featuring 12 active sites and a molecular mass of 1.06 MDa as determined by mass spectrometry.[9] The dodecamer enzyme is a homoteteramer of homotrimers of heterodimers, an $(\alpha\beta)_{12}$ structure. A single $\alpha\beta$ dimer is depicted in Figure 1-3 and the overall structure in Figure 1-4. The active site (Figure 1-5) of the enzyme possesses a barrel-like shape and features 2 nickels at the bottom of the barrel, which are respectively penta- and hexacoordinated.[6] A water molecule separately coordinates both Ni^{2+} ions. Both a bridging hydroxide ion and a bridging carbamylated lysine residue further coordinate these nickel ions. Each Ni^{2+} ion is coordinated by a histidine residue as well. One of the nickels is also coordinated with an aspartate. The twelve

active sites possess flaps that cover the entrance to the active site cavity. Key residues of the

active site and the residues of the active site-covering flap are listed in Table 1-1. It has a large

hollow in the center that has ten entrances; 4 are along the 3-fold symmetry axes, and 6 are along

the 2-fold axes of symmetry.[6] HP urease is therefore a highly symmetrical enzyme.
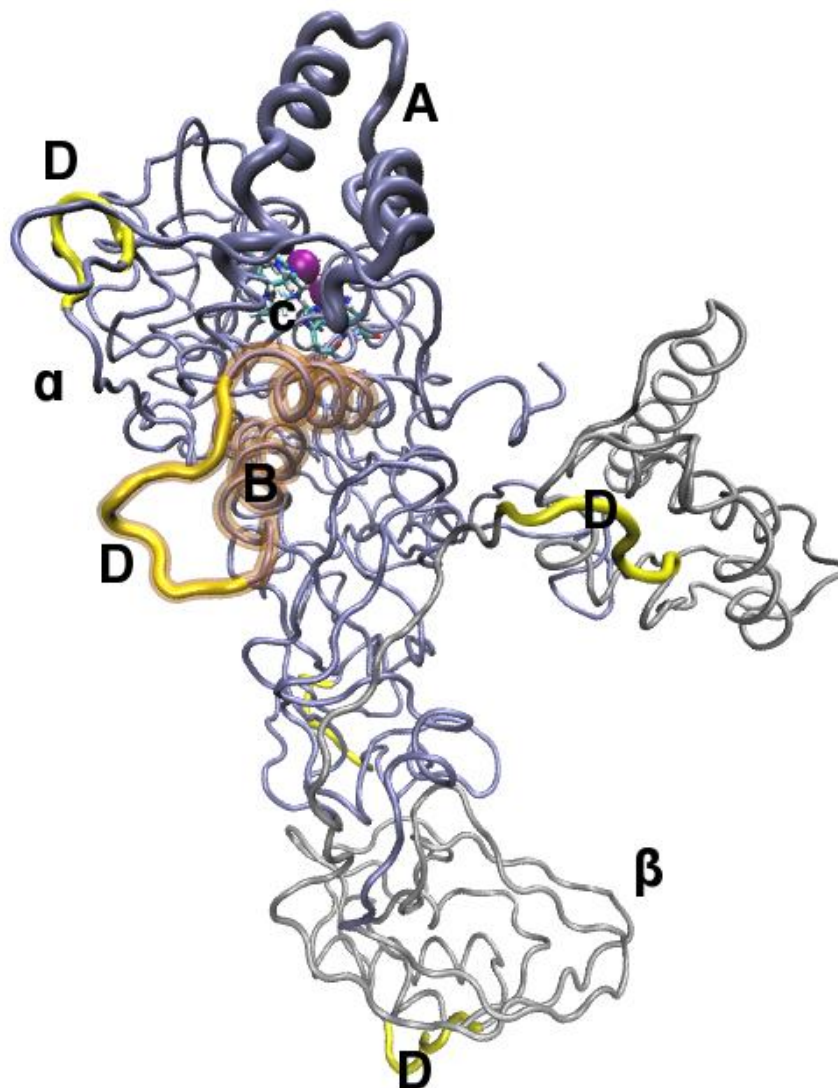
Figure 1-3. Single αβ dimeric subunit: chain β is colored silver (β); chain α is colored iceblue
(α); the active site covering flap is given in bold iceblue (A); Ni^{2+} ions are shown as
purple spheres; transparent brown highlights the α-helices of the second mobile flap
(B), (see results); active site residues in licorice (C); high RMSF regions are given in
bold yellow (D).
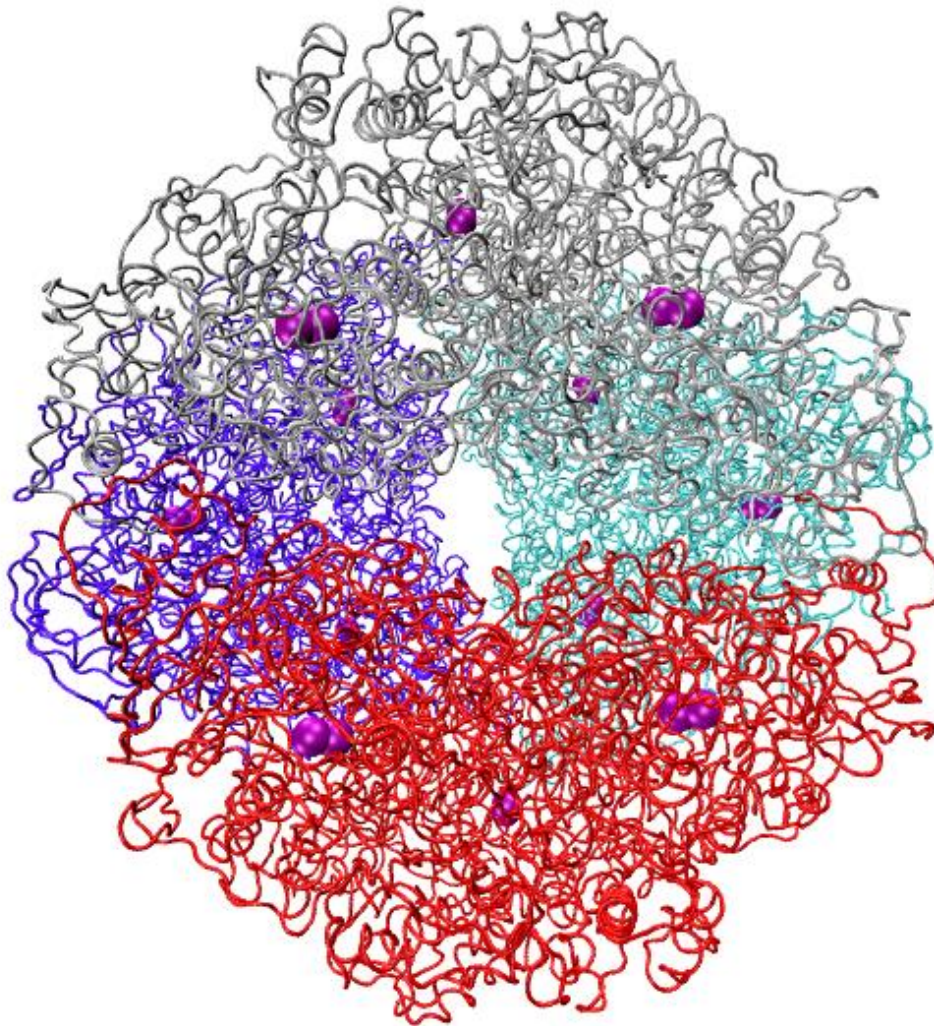
Figure 1-4. Entire structure of *H. pylori* urease. Each trimeric subunit of the tetramer is individually colored to depict the overall assembly. Ni^{2+} ions are depicted as purple spheres.
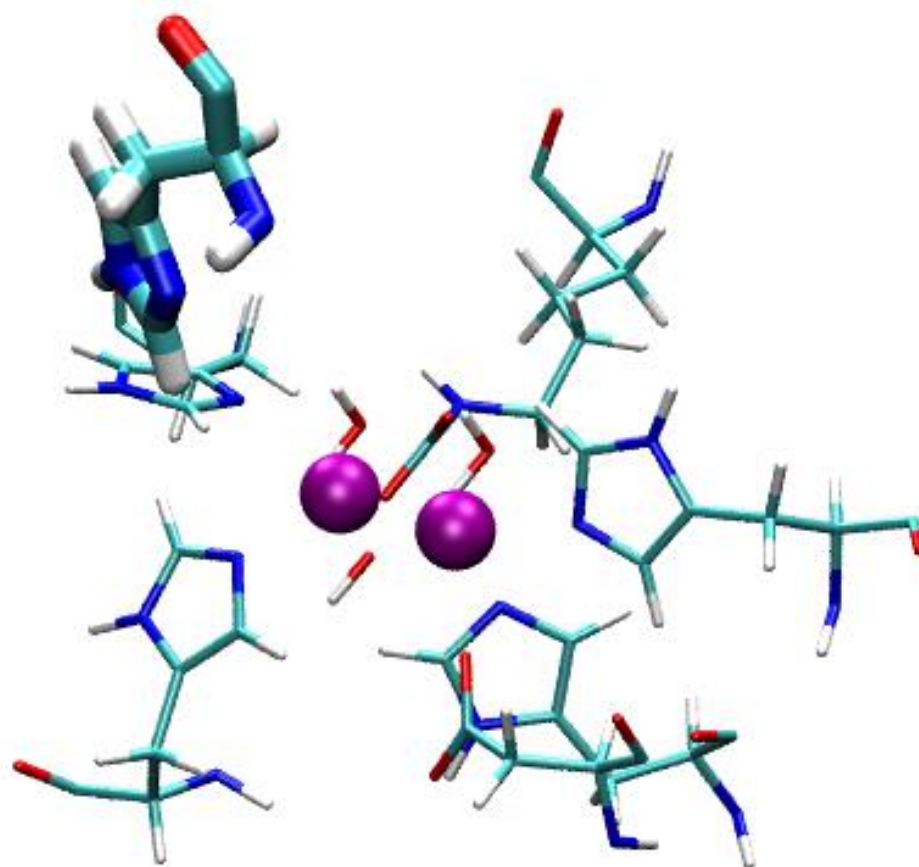
Figure 1-5. Urease active site. Ni^{2+} ions are depicted as purple spheres.

Table 1-1. Key regions and residues in HP urease. Flap residues refer to areas as outlined in the crystal structure. Numbering is according to Ha et al.[6]

Region	Residue Numbers and Types[a]
pentacoordinate Ni^{2+}	LYSα219(CO_2^-), HISα248(δN), HISα274(εN), bridging ^-OH, H_2O
hexacoordinate Ni^{2+}	LYSα219(CO_2^-), HISα136(εN), HISα138(εN), ASPα362(CO_2^-), bridging ^-OH, H_2O
Flap α-helix 1	α313-α322 GLU-HID-MET-ASP-MET-LEU-MET-VAL-CYS-HIE
Flap turn	α323-α329 HIP-LEU-ASP-LYS-SER-ILE-LYS
Flap α-helix 2	α330-α336 GLU-ASP-VAL-GLN-PHE-ALA-ASP

[a]HID, HIE and HIP refer to histidines with protonated δ, ε and both imidazole nitrogens, respectively.

The need to understand urease better is very important. Ever since it was shown that HP infections cause ulcers and other diseases, it has been indicated that inhibiting HP urease is a viable strategy that would help to kill off HP in the stomach lining of humans. Given that my research directly involves the inhibition of HP urease, it is useful to consider all forms of inhibition: competitive, noncompetitive, uncompetitive, product, and suicide inhibition. Competitive inhibition occurs when the inhibitor binds to the active site of the enzyme, preventing binding of the substrate. Most competitive inhibitors function by binding reversibly to the active site of the enzyme. During initial efforts to identify potential HP urease inhibitors, we focused largely on competitive inhibition. Noncompetitive inhibition involves a molecule that binds to an enzyme in a location other than the active site, thereby changing the enzymes 3-D structure and allowing the active site to bind substrate with the usual affinity but no longer in optimal arrangement. Uncompetitive inhibition involves an enzyme inhibitor that binds solely to the complex created between the enzyme and the substrate. Product inhibition occurs when the product of an enzymatic reaction binds to the enzyme and inhibits its activity. Product inhibition is a self-regulatory process. Suicide inhibition is a form of irreversible enzyme inhibition that occurs when an enzyme binds a substrate analogue and forms an irreversible complex through a covalent bond during the normal catalyzed reaction.

There have been multiple studies conducted on urease aimed at identification of inhibitors of urease that operate via any of the aforementioned mechanisms. In 1975, Dixon et al. identified phosphoramidate as a reversible inhibitor of urease.[10] It enters the active site and replaces a hydroxide and water located in the coordination site of HP urease. However, this inhibitor is not specific enough and can bind to many other targets, leading to potentially harmful side effects if administered to humans. Sometime later, a group in China began studying the effects of flavinoids on HP urease. Xiao et al. performed a study on analogs of quercetin and HP urease.[11] There was some level of inhibition observed for green tea extract and quercetin, but neither compound was able to compete with urea, which is the natural substrate of HP urease. Matsubara from Japan further studied green tea extract, which was found to contain a variety of catechins.[12] Subsequently, Benini and coworkers performed mechanistic studies on boric acid, discovering that it was a tight binder; however, the issues tied to AHA also surfaced with boric acid. Subsequently, Dominguez and coworkers studied phosphoramides, observing binding behavior both computationally and experimentally; they concluded that both compounds were reasonable inhibitors.[13] Abid and coworkers ran docking studies on small libraries to identify new scaffolds and identified a scaffold that demonstrated better potency then thiourea.[14] Finally, Benini studied citrate complexed with *Sporosarcina pasteurii* urease in order to propose new patterns for new inhibitors.[15] They suggested that the most ideal inhibitors should possess three elements: a nickel binding group, flexible carbon scaffolds, and properly aligned hydrogen bonding sites.

The research reported in this thesis describes efforts at ultimately identifying novel, potent inhibitors of the urease enzyme. Chapter 2 details the computational methods employed in these studies along with appropriate background. In Chapter 3, a 400 ns simulation of the HP

urease enzyme in aqueous solution is reported with particular focus given to motions associated with the flap covering the active site cavity. The effects of a 10.5 M solution of aqueous urea, the substrate of the enzyme, on the structure and dynamics HP urease are detailed in Chapter 4. Chapter 5 covers the effects of a 150 mM aqueous ammonium chloride solution on the enzyme, where ammonium is an immediate product of the enzymatic hydrolysis of urea. In Chapter 6, the results of docking studies where 14 million ligands were bound to the active site are reported. Finally, Chapter 7 details the unique experiences of a blind graduate student pursuing a doctoral degree in both the classroom and research environments, incorporating multiple perspectives to best illustrate the experience and provide insight toward the education of future blind scientists.

CHAPTER 2
THEORY AND METHODOLOGY

2.1 Molecular Dynamics

Molecular dynamics (MD) is a powerful theoretical tool utilized to elucidate motions

associated with proteins, as well as other chemical and biochemical systems, both macro- and

micromolecular. MD involves the simulation of physical movement of atoms and/or molecules

of interest within an N-body simulation using molecular mechanics. Following the significant

work of Metropolis et al. in 1953[16] using adapted Monte Carlo simulation techniques, the

fundamentals of MD were constructed in the late 1950s through the efforts of Alder and

Wainwright[17] and by Rahman[18] five years later. The development of accessible computational

resources subsequently allowed MD to become a feasible technique for the study of molecular

systems. Typical MD simulations begin with a single starting position for all relevant atoms or

molecules in the system of study. Then, using molecular mechanics, Newtonian motion

equations and force field calculations, the acceleration, and thus the movement, of the involved

atoms or molecules within a single time step are determined. These determinations are then

reiterated to construct a sufficient trajectory over a designated period of time. This trajectory

occurs within a set phase space, which is dependent upon the coordinates of position q (x, y, z)

and momentum p (p_x, p_y, p_z) of each particle of the system. Thus this phase space is 6N-

dimensional in nature.[19] Though MD is a versatile technique, as evident through its use in

biochemistry, biophysics, and materials science, it nonetheless features several major limitations.

Time-step determination, for example, demands a balance between long lengths, which can

result in deviation of the trajectory to low-probability regions of phase space, and short lengths,

which can become computationally demanding. The number of atoms in a simulation, reliance

on quality and power of computational equipment, and the upper limit of computational time all reflect additional challenges as well.

A wide assortment of algorithms and methods has been developed over time for the use of MD. One such method is the Verlet algorithm, initially created as a computationally stable alternative to Euler's approximation.[20] The Verlet algorithm is constructed by first obtaining Taylor expansions for the forward and reverse time-steps Δt, each truncated at third-order. These Taylor expansions are then combined, and rearrangement results in Equation 2.1.

$$q(t + \Delta t) = 2q(t) - q(t - \Delta t) + a(t)(\Delta t)^2 \tag{2.1}$$

The resulting equation dictates that the position of a particle of interest after a particular time step is dependent on its previous and current position vector, as well as its acceleration vector. Said position is therefore independent of both velocity and momentum. Though this may limit the application of Verlet's algorithm in certain contexts, it is useful for studying momentum-independent properties of a system using fewer computational resources.

To both reintroduce the velocity vector and provide a means to control temperature within a model system, the Verlet method can be altered further to produce what is known as the leapfrog algorithm. To produce the leapfrog algorithm, the Taylor expansions of the position vector about t + Δt/2 are constructed and truncated at second order, then combined to produce equation 2.2.

$$q(t + \Delta t) = q(t) + v\left(t + \frac{1}{2}\Delta t\right)\Delta t \tag{2.2}$$

Note that the position function is now dependent on a velocity function calculated one half time step out of phase. This velocity function can be defined by performing similar Taylor expansions on velocity to produce equation 2.3.

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + a(t)\Delta t \qquad (2.3)$$

The result is a coupling of the position and velocity vectors, where the velocity can be scaled to control temperature within the system. Note that in this method, force field calculations are only performed at integral time steps; half-time step velocities are determined from the integral time-step force field calculations, not directly from the applied force fields.

2.2 Force Fields

The AMBER (Assisted Model Building with Energy Refinement) software package was utilized to conduct all molecular dynamics simulations and molecular mechanics based structure minimizations.[21] To perform these studies an AMBER force field specifically designed for protein analysis, FF99SB, was employed. The basic functional form of the AMBER potential is presented in equation 2.4.[22]

$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2$$
$$+ \sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\varphi - \gamma)]^2 + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}}\right] \qquad (2.4)$$

This basic form consists of four unique summations. The first summation represents the bond term, which describes the molecular mechanics between two bonded atoms using a simple diagonal harmonic expression. The second summation describes the angular mechanics of atoms in the system using a similar expression. The third summation represents the dihedral energies as expressed by a minimal parameter set involving two central atoms. The final summation includes the electrostatic terms, which are represented by a Lennard-Jones expression and a Coulombic term.

Table 2-1. Atom types that pertain to AMBER force fields.[22]

atom	type	description
carbon	CT	any sp³ carbon
	C	any carbonyl sp² carbon
	CA	any aromatic sp² carbon and (Cε of Arg)
	CM	any sp² carbon, double bonded
	CC	sp² aromatic in 5-membered ring with one substituent + next to nitrogen (Cγ in His)
	CV	sp² aromatic in 5-membered ring next to carbon and lone pair nitrogen (e.g. Cδ in His (δ))
	CW	sp² aromatic in 5-membered ring next to carbon and NH (e.g. Cδ in His (ε) and in Trp)
	CR	sp² aromatic in 5-membered ring next to two nitrogens (Cγ and Cε in His)
	CB	sp² aromatic at junction of 5- and 6-membered rings (Cδ in Trp) and both junction atoms in Ade and Gua
	C*	sp² aromatic in 5-membered ring next to two carbons (e.g. Cγ in Trp)
	CN	sp² junction between 5- and 6-membered rings and bonded to CH and NH (Cε in Trp)
	CK	sp² carbon in 5-membered aromatic between N and N-R (C8 in purines)
	CQ	sp² carbon in 6-membered ring between lone pair nitrogens (e.g. C2 in purines)
nitrogen	N	sp² nitrogen in amides
	NA	sp² nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp)
	NB	sp² nitrogen in 5-membered ring with lone pair (e.g. N7 in purines)
	NC	sp² nitrogen in 6-membered ring with lone pair (e.g. N3 in purines)
	N*	sp² nitrogen in 5-membered ring with carbon substituent (in purine nucleosides)
	N2	sp² nitrogen of aromatic amines and guanidinium ions
	N3	sp³ nitrogen
oxygen	OW	sp³ oxygen in TIP3P water
	OH	sp³ oxygen in alcohols, tyrosine, and protonated carboxylic acids
	OS	sp³ oxygen in ethers
	O	sp² oxygen in amides
	O2	sp² oxygen in anionic acids
sulfur	S	sulfur in methionine and cysteine
	SH	sulfur in cysteine
phosphorus	P	phosphorus in phosphates
hydrogen	H	H attached to N
	HW	H in TIP3P water
	HO	H in alcohols and acids
	HS	H attached to sulfur
	HA	H attached to aromatic carbon
	HC	H attached to aliphatic carbon with no electron-withdrawing substituents
	H1	H attached to aliphatic carbon with one electron-withdrawing substituent
	H2	H attached to aliphatic carbon with two electron-withdrawing substituents
	H3	H attached to aliphatic carbon with three electron-withdrawing substituents
	HP	H attached to carbon directly bonded to formally positive atoms (e.g. C next to NH₃⁺ of lysine)
	H4	H attached to aromatic carbon with one electronegative neighbor (e.g. hydrogen on C5 of Trp, C6 of Thy)
	H5	H attached to aromatic carbon with two electronegative neighbors (e.g. H8 of Ade and Gua and H2 of Ade)

ᵃ See refs 5 and 6.

2.2.1 Atom Types

Before construction of a force field of interest, definitions and categorization of atom types is crucial. Atoms are categorized based on their hybridization and the other atoms to which they are bonded. Table 1 of Cornell et al. includes a list of atom types relevant to this discussion of AMBER force fields (Table (2-1)).

2.2.2 Bond and Angle Parameters

Once we have defined these atom types, we can begin to discuss how said atoms interact with each other through descriptions of the resulting bonds and angles. The final goal involved determining a set of parameters with which to characterize the nature of these bonds and angles. To do so, we begin by determining starting values for the equilibrium bond length r and bond angle θ through X-ray structural data, as well stretching constants, K_r, through linear interpolation and bending constants, K_θ, from vibrational analysis. From here, a simple functional fragment was constructed and energy calculations were conducted to obtain the optimal set of parameters. Said calculations were performed using the 6-31G* basis set, which was more computationally intensive and complex than the STO-3G basis set employed by Weiner et al. a decade prior to the Cornell et al. force field.[22-23]

2.2.3 Dihedral Parameters

The dihedral parameters used in these studies were constructed similarly to those developed by Weiner et al.[23] In both cases, parameters were first optimized using simple molecules, and then applied to molecules of increased size and complexity. This minimalist design was maintained by largely avoiding the addition of additional Fourier components to dihedral energy calculations unless necessary, such as in the case of clear "gauche" effects, phosphorus-ester bonds, and various dipeptide and nucleoside dihedrals.

2.2.4 VDW Parameters

Following a methodology similar to that used in OPLS, Van der Waals (VDW) parameters for sp^3 and sp^2 hybridized carbons – as well as both aromatic and aliphatic hydrogens – were constructed through Monte Carlo simulations on small-molecule liquids using the 6-31G* basis set. Remaining VDW parameters were taken directly from the OPLS model. As in previous parameter sets, the VDW set was built with a minimalistic approach; all atoms of similar type were given a specific parameter, with select exceptions as dictated by the OPLS model. The uniqueness of this model stems from the way that hydrogen parameters are assigned and treated. Namely, the sensitivity of hydrogen atoms to the bonding environment was focused upon.

2.2.5 Electrostatic Parameters

In addition to relying on electrostatic potentials taken from the 6-31G* basis set, the parameters employed in this model involve the use of multiple molecules and conformations, as well as RESP[24] fitting. The latter two address the underestimation of charge with regards to buried atoms. Comparison with experimental values for liquid enthalpies and solvation energies of methanol and N-methylacetamide supported the quality of calculations involving these parameters.

2.2.6 Force Field Types

The force fields used to compare this parameter set can be grouped into four categories: force fields with rigid geometries, force fields with dampened/no electrostatic terms, simple diagonal force fields, and complex force fields. Force fields of rigid geometry, such as the ECEPP[25] and FLEX (JUMNA)[26] force fields, can allow for more effective study of conformational space, but at the cost of increased conformational barrier energies. Force fields with omitted or reduced electrostatic components, such as the TRIPOS[27] (SYBYL) and YETI[28] force fields, are useful in the study of models where the increase in accuracy for taking into

account complex electrostatic terms do not justify the increased computational expense. However, these force fields are not well suited for studying phenomena reliant on electrostatic interactions, such as condensed-phase properties. Simple diagonal force fields use simple harmonic diagonal representations for bond and angle terms. The OPLS[29] force field is an example of a simple diagonal force field. Charge derivations and VDW parameterization may differ between force fields in this category (which include Weiner et al., CHARMM[30], and GROMOS[31]), though dihedral energy terms are typically represented with simple Fourier expansions across all force fields (excusing differences in energy assignment to bonds). Complex force fields, meanwhile, extend beyond simple diagonal functions, utilizing extra computational resources to model necessary subtleties. Examples include the MM2[32]/MM3[33] force fields for small molecules, the "Class II" force field[34], and the Merck Molecular Force Field[35] (MMFF94).

2.2.7 Molecular Dynamics: Methodology

The AMBER suite of programs was utilized in order to perform all molecular dynamics simulations and all molecular mechanics-based energy minimizations.[21] These calculations entailed multiple steps of preparation. Described in this section are the various computational steps involved in the simulations, with sample input files provided for each job type. To run a molecular dynamics simulation, first it was necessary to prepare the protein, which was accomplished using the tLeap function found in the AmberTools[21b] software suite. After this step, the proper restart file is produced. Missing hydrogen atoms are placed onto the protein, the proper parameters are established for the Ni-bonded sites and the protein is properly solvated and placed in orthorhombic periodic box. The system was subsequently neutralized.

After this step, the geometry is relaxed to an energy minimum using a two-step procedure. After reaching a satisfactory minimum, the system is allowed to reach equilibrium.

During the equilibration process the temperature is raised from 0K to 300K and following this process the MD production run can be obtained.

2.2.7.1 tLeap

The input required for tLeap is provided below.

source leaprc.ff99SB

loadAmberParams metals.frcmod

loadoff ions08.lib

loadamberparams frcmod.ionsjc_tip3p

addAtomTypes {

  { "HX" "H" "sp3" }

  { "HQ" "H" "sp3" }

  { "N5" "Ni" "sp3" }

…

  { "OO" "O" "sp3" }

  { "OQ" "O" "sp3" }

  { "OU" "O" "sp2" }

}

WAT=TIP3P

frcmod = loadamberparams 1EJX_high.frcmod

loadamberprep 1EJX.prep

model = loadpdb "hptotal.pdb"

bond model.808.NI model.457.OQ1

bond model.808.NI model.486.ND1

bond model.808.NI model.512.NE2

…

bond model.9741.NI model.9532.OD1

bond model.9741.NI model.9742.O

bond model.9741.NI model.9744.O

solvateoct model TIP3PBOX 10.0

addions model Na+ 0

savepdb model testhp.pdb

saveamberparm model testhp.prmtop testhp.inpcrd

quit

In this tLeap input file, the first line is the command to load the force field necessary to run the MD simulation. The force field is FF99SB[36], which is a force field specialized for proteins. The chemical uniqueness of this system (due in large part to the nickel cations) required creation of a set of specialized parameters, which were subsequently loaded these parameters using the "loadamberparams" command. A set of parameters regarding ions was required as well, as well as the proper water model parameters, using the line "loadamberparams frcmod.ionsjc_tip3p." After loading these parameters to run the simulation, the particular atom types within the force field required definition, using the "addatomtypes" command. The water model was specified in the line "WAT = TIP3P." Several additional parameters were required regarding the coordination of the Ni^{2+} ions with the bonded ligands* (1EJX_high.frcmod vs 1EJX.prep). The variable was subsequently defined and the initial PDB of the system loaded, defined "model." The coordination sphere of each of the 24 Ni^{2+} cations (12 coordinated to 5 ligands, 12 coordinated to 6 ligands) was defined using a bonded model ("bond, name of protein.residue number.atom, model.resid.coordinated atom"). Then we solvated the system,

deciding to use an octahedral box, hence the "solvateoct" command. The model was solvated

using the "TIP3PBox" water model, with waters a minimum 10 Å from the surface of the

protein. Neutralization was accomplished by adding ions to the system. We determined that there

was a total charge of -44; to neutralize the protein, we used the command "addions" to the

model, using Na^+ cations for neutralization. After finishing, we saved the PDB file containing

the changes, as well as the topology parameter file and inputCRD file.

2.2.7.2 Minimization

The minimization was conducted over two steps requiring separate input files. The first

of these input files, shown below, relaxes the geometry of only the binding pocket.

min0: everything else but the binding pocket kept fixed

&cntrl

imin   = 1,

maxcyc = 100000,

ntmin = 2,

ntf = 1,

ntb   = 1,

ntr   = 1,

restraintmask='!(:H11,H12,H21,H22,A21,NI1,NI2,ONB,ON1,ON2,KCB)',

restraint_wt=10.0,

cut   = 8.0

/

Following the title line is the "&cntrl" line, which precedes the general set of parameters,

which signifies the beginning of command input. A minimization calculation is selected using

the flag command of "imin=1". The number of steps to be used is established with "maxcyc =

100000". The "ntmin" flag determines the minimization method; setting it equal to 2 indicates sole use of the steepest descent method. All interaction in the binding pocket is calculated with the flag "ntf = 1". To carry out the minimization at constant volume, we use "ntb = 1". The next commands restrict the geometry of all portions of the enzyme and solvent system that are not a part of the active site. Setting ntr=1 enabled restrictions, then the restraintmask flag was used to specify the atoms that need to be restrained. For this minimization, the "restraintmask" was set to "!(:H11,H12,H21,H22,A21,NI1,NI2,ONB,ON1,ON2,KCB)',", indicating that all but the listed active site residues were restrained. The restraint weight was defined as 10 kcal/mol-$Å^2$. This restraint is of the form $k(dx)^2$. Finally, a cutoff length was set for nonbonded interactions, using the suggested value of 8 Å.

The second input file, relaxes the entire structure, allowing complete minimization, and is shown below.

min1: min whole structure

&cntrl

imin   = 1,

maxcyc = 10000,

ntmin = 2,

dx0   = .000001,

ntf   = 1,

ntb   = 1,

cut   = 8.0

/

As in the previous input file, the &cntrl flag signifies the beginning of the parameter list. A minimization calculation is selected with the "imin=1", with 10,000 minimization cycles. Like the previous input, the steepest descent method was employed throughout with the "ntmin=2" flag. Exclusive to this input file is the initial step length of .000001 ps, indicated by "dx0 = 0.000001"; this was to address an issue with the initial step of the steepest descent minimization.

2.2.7.3 Equilibration

In the first of 2 steps of equilibration, we use the following input file:

hp equilibration

&cntrl

nstlim=1000000, dt=0.002, ntpr=5000, ntwr=5000, ntwx= 5000,

tempi=0, temp0=300.0, ntt=3, imin=0,

ntb=1, cut=8, ig=-1, ntc=2, ntf=2, gamma_ln=2.0,

ioutfm=1, ntwv=-1, ntr=1, restraintmask=':1-9744', restraint_wt=10.0

/

&wt TYPE='TEMP0', restraint_wt=10.0, istep1=0, istep2=500000,

value1=0, value2=300.0, /

&wt TYPE='TEMP0', restraint_wt=1.0, istep1=500001, istep2=1000000,

value1=300.0, value2=300.0, /

&wt TYPE='END' /

The number of molecular dynamics steps to be performed was set to 1000000 using the "nstlim" parameter. The time step length was set using the "dt" parameter, which is .002 psec in this simulation. A series of flags to describe the frequency at which data of the equilibration is printed out is then established with three keywords ("ntpr, ntwr, ntwx"). The "temp1" indicates an initial temperature of 0 K, which will then change over the course of the equilibration to the

temp0 parameter of 300.0. The flag "ntt=3" indicates the use of Langevin dynamics. In using this flag, we will need to define the collision frequency later in the input file. "Imin=0" indicates that no minimization was performed in this step. A constant-volume condition was imposed via the flag "ntb=1," and a non-bonded interaction cutoff of 8 Å via the cut parameter. It is recommended to use a random number generator to produce the value for the MD starting velocity, which can be chosen to be based upon the current date and time with the flag "ig=-1". The following "ntc" flag is used to allow the SHAKE[37] algorithm to constrain all bonds involving hydrogen. As such, forces do not need to be calculated for these H-bonds; said calculations are omitted from SHAKE using the "ntf=2" flag. The "gamma_ln" parameter provides the aforementioned collision frequency to be used in Langevin dynamics calculations. The following "ioutfm=1" flag sets the format out the output trajectory file to binary NetCDF. "Ntwv=-1" represents another output flag, included to write velocities to a mdcrd coordinate/velocity trajectory file. As demonstrated in previous input files, ntr=1 is used to restrain atoms; in this case, the entire protein is restrained with a weight of 10 ("restraintmask=':1-9744'"), to reduce protein movement.

Beyond this initial setup, are two "&wt" namelist blocks utilized to construct a system of varying temperature. The first block operates on the "temp0" parameter ("TYPE="TEMP0"). Between the 0th and $500,000^{th}$ step ("istep=0, istep2=500000"), the parameter was raised from 0K to 300K ("value1=0.0, value2=300.0"). Throughout, "restraint_wt" was held at 10.0. The second block maintains the "temp0" parameter ("value1=300.0, value2=300.0") throughout the rest of the process ("istep1=500001, itstep2=1000000"). Note that the "restraint_wt" parameter is set to 1.0 in this step. Finally, an additional "&wt" block is added to end the process ("&wt=END").

Subsequently, the second equilibration input file is executed, with the syntax displayed

below:

hp equilibration NPT

 &cntrl

 nstlim=5000, dt=0.002,ntx=5,irest=1,ntpr=1000,ntwr=1000,ntwx=1000,

 tempi=300.0, temp0=300.0, ntt=3, imin=0,

 ntb=2, cut=8,

 pres0 = 1.0, ntp = 1,

 taup = 2.0, iwrap=1,

 ntc=2, ntf=2, gamma_ln = 1.0,

 ioutfm=1, ntwv=-1, ig=-1,

 /

In this input file, the number of molecular dynamics steps was set to 5,000 ("nstlim =

5000"). The time-step lengths were also set to 0.002 ps via the "dt" parameter. To read the initial

coordinates, velocities, and box size from the previous MD restart file, the flag "nxt=5" was

included along with an additional "irest=1" flag, used to restart the simulation using coordinates

and velocities from the previous restart file. Again, the "ntpr," "ntwr," and "ntwx" parameters

indicate the frequency at which particular data is exported. In this input, initial ("tempi") and

final ("temp0") are both set to 300.0K. Langevin dynamics are employed through the "ntt=3"

flag, with the collision frequency specified further in the input. Minimization is avoided in this

step using the "imin=0" flag. Constant pressure is established by setting "ntb=2," and a

nonbonding interaction cutoff at 8 Å was employed ("cut=8"). Reference pressure ("pres0") is

set to 1 bar, and constant pressure dynamics are controlled through isotropic position scaling via

the "ntp=1" flag. Pressure relaxation time was then set to 2.0 ps. Momentarily referring back to output structure, "iwrap=1" ensured that the coordinates included in the trajectory and restart files are kept in a primary box. The bonds containing hydrogen were constrained using the SHAKE algorithm through the "ntc=2" flag, subsequently omitting force calculation of these constrained bonds by including the "ntf=2" flag. Then, after including the "gamma_ln" collision frequency of 1.0, additional output was requested, formatting the trajectory file to a binary NetCDF format ("ioutfm=1") as well as including velocities to the mdcrd file ("ntwv=-1"). Finally, the random seed generation was set identically to the previous input file.

2.2.7.4 Production

The production phase was initiated using the input file shown below.

hp equilibration NPT

&cntrl

nstlim=50000000, dt=0.002,ntx=5,irest=1,ntpr=5000,ntwr=5000,ntwx=5000,

tempi=300.0, temp0=300.0, ntt=3, imin=0,

ntb=2, cut=8,

iwrap=0,

pres0 = 1.0, ntp = 1,

taup = 2.0,

ntc=2, ntf=2, gamma_ln = 1.0,

ioutfm=1, ntwv=-1, ig=-1,

/

The molecular dynamics step count was set to 50,000,000 ("nstlim") and the time step length to 0.002 ps ("dt"), in order to acquire a 100ns simulation. The flags describing frequency of data output were specified ("ntpr", "ntwr", and "ntwx)" and each flag is used in a fashion

similar to its usage in previous inputs), setting each to 5,000. Temperature was set at a constant 300.0K via the "tempi=300.0" and "temp0=300.0" parameters. Langevin dynamics were employed via "ntt=3," and minimization was again prevented with the flag "imin=0." Constant pressure was set using ("ntb=2") and the nonbonding interaction cutoff established at 8 Å ("cut=8"). The function of the remainder of the input mirrors that of the previous equilibration file; coordinates were kept in a primary box ("iwrap=1"), pressure was held at 1 bar via isotropic position scaling ("pres0=1.0, ntp=1") and assigned a relaxation time of 2.0ps ("taup=2.0"), hydrogen bonds were constrained by and omitted from SHAKE calculations ("ntc=2, ntf=2"), the collision frequency was set to 1.0 ("gamma_ln=21"), and additional output flags were established ("ioutfm=1, ntwv=-1, ig=-1").

2.3 Quantum Mechanical Methods

The Hartree-Fock (HF) approximation is a powerful method for determining appropriate wave functions for many-electron systems.[3] It exists as an updated form of the self-consistent field (SCF) method, which was initially used by Hartree as a means of accounting for repulsive electron interactions in Hamiltonians. The requirement for the SCF method arose when the single-electron operators derived from the many-electron Hamiltonian (separable because it only features kinetic energy and nuclear attraction terms for single electrons) were modified to include such pairwise interactions. This modification is shown in Equation 2.5, where h_i is the single-electron Hamiltonian, and V is the repulsive term defined in Equation 2.6.

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{k=1}^{M}\frac{Z_k}{r_{ik}} + V_i\{j\} \tag{2.5}$$

$$V_i\{j\} = \sum_{j\neq i}\int \frac{\rho_j}{r_{ij}}d\boldsymbol{r} \tag{2.6}$$

In Equation 2.6, ρ_j represents, for an electron j, the charge probability density. ρ_j is defined as

the square of the magnitude of the associated eigenfunction that satisfies Equation 2.7, the one-

electron Schrodinger equation linked to the corresponding Hamiltonian.

$$h_i \psi_i = \varepsilon_i \psi_i \tag{2.7}$$

This introduced a problem: the eigenfunctions must be known in order to determine h_i. The

purpose of the calculation is to find those single-electron eigenfunctions so that one may

determine the proper multi-electron eigenfunction for the parent multi-electron Hamiltonian,

shown in Equation 2.8, where the target eigenfunction, ψ_{HP}, is known as the Hartree product

wavefunction, and satisfies Equation 2.9

$$\psi_{HP} = \psi_1 \psi_2 \dots \psi_N \tag{2.8}$$

$$H\psi_{HP} = \left(\sum_{i=1}^{N} \varepsilon_i \right) \psi_{HP} \tag{2.9}$$

Knowing the correct eigenfunctions beforehand is therefore an unreasonable requirement. The

SCF method addresses this by first requiring a guess for the proper single-electron wave

functions, from which the operators h_i are built. These operators are then used with Eq. 2.7 to

derive a new set of wave functions. This new set, if sufficiently different from the previous set,

is used to build a new set of h_i, and the process is repeated until a set of eigenfunctions is

produced which is close enough to the previous set to achieve convergence. As a final change,

note that Eq. 2.5 requires further adjustment to prevent double-counting of repulsive forces,

resulting in the eigenvalue of Eq. 2.9 becoming that shown in Equation 2.10.

$$E = \sum_{i} \varepsilon_i - \frac{1}{2} \sum_{i \neq j} \iint \frac{|\psi_i|^2 |\psi_j|^2}{r_{ij}} \, d\mathbf{r}_i \mathbf{r}_j \tag{2.10}$$

Unfortunately, the use of Hartree product wavefunctions fails to take critical principles regarding electron spin into account – namely, Pauli exclusion. Suppose, for example, that there exists a HP wavefunction for a system of two same-spin electrons shown in Equation 2.11.

$$^3\Psi_{HP} = \psi_a(1)\alpha(1)\psi_b(2)\alpha(2) \tag{2.11}$$

In this wavefunction α indicates one of the two spin states. This wavefunction violates the Pauli exclusion principle, as it is not anti-symmetric. An alternative, known as the Slater determinant (SD), can produce an anti-symmetric wavefunction that satisfies the Pauli principle. The generalized form of the SD can be written as

$$^3\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} = |\chi_1\chi_2\cdots\chi_N\rangle \tag{2.12}$$

Where N is the electron total and χ represents a spin-orbital. When applied to the two-electron model described previously, for example, the SD becomes

$$^3\Psi_{SD} = \frac{1}{\sqrt{2}} [\psi_a(1)\alpha(1)\psi_b(2)\alpha(2) - \psi_a(2)\alpha(2)\psi_b(1)\alpha(1)] \tag{2.13}$$

The SD is unique in that it accounts for several quantum mechanical properties. Not only is particle indistinguishability addressed, but the concept of Fermi holes and same-spin electron interactions are inherently accounted for. Given its advantages over HP wave functions, Fock extended Hartree's SCF method to include SD wave functions. The result was then adapted for use with an MO basis set using linear algebra (specifically, Roothaan equations[38]), resulting in computationally feasible "restricted Hartree-Fock" (RHF) calculations. The overall mathematical methodology is summarized below.

To begin, in place of h_i, Fock operators are used, with the single-electron operator f defined for an electron, i, in Equation 2.14.

$$f_i = -\frac{1}{2}\nabla_i^2 - \sum_k^{nuclei} \frac{Z_k}{r_{ik}} + V_i^{HF}\{j\} \tag{2.14}$$

Note that the third term in the operator is defined as $2J_i - K_i$, where J is the Coulombic repulsive

energy and K is the reduction of repulsive forces as a result of Fermi holes. With f defined, the

following secular equation must be solved for roots E_{ij}, as shown in Equation 2.15. The S

elements are from the overlap matrix and the elements F are defined in Equation 2.16.

$$\begin{vmatrix} F_{11} - ES_{11} & F_{12} - ES_{12} & ... & F_{1N} - ES_{1N} \\ F_{21} - ES_{21} & F_{22} - ES_{22} & ... & F_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ F_{N1} - ES_{N1} & F_{N2} - ES_{N2} & ... & F_{NN} - ES_{NN} \end{vmatrix} = 0 \tag{2.15}$$

$$F_{\mu\nu} = \langle \mu | -\frac{1}{2}\nabla^2 | \nu \rangle - \sum_k^{nuclei} Z_k \langle \mu \left| \frac{1}{r_k} \right| \nu \rangle + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma) \right] \tag{2.16}$$

Each of the components in F follow specific notation. In the first and second terms, we have the

notation given in Equation 2.17 where g is an operator and Φ is a basis function.

$$\langle \mu | g | \nu \rangle = \int \phi_\mu (g\phi_\nu) d\boldsymbol{r} \tag{2.17}$$

The final term is presented in Equation 2.18, where probability densities are represented as Φ_μ

and Φ_ν for the first electron and Φ_λ and Φ_σ for the second.

$$(\mu\nu|\lambda\sigma) = \iint \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2)d\boldsymbol{r}(1)d\boldsymbol{r}(2) \tag{2.18}$$

Furthermore, in the same term, are the elements $P_{\lambda\sigma}$ in Equation 2.19, which depict a density

matrix P indicating relative contributions of each basis function, where coefficienct $a_{\zeta i}$ depicts

the normalized contribution to molecular orbital i by basis ζ.

$$P_{\lambda\sigma} = 2 \sum_{i}^{occupied} a_{\lambda i} a_{\sigma i} \tag{2.19}$$

These coefficients are determined by the secular equation (Equation 2.15), which in turn requires said coefficients to be evaluated. Thus SCF is required to determine the proper coefficients.

Prior to docking individual ligands, the structures require preparation. Ligands are built by hand using GaussView to get a rough initial structure. If a particular ligand had multiple possible minima, those initial structures were all built. Then, using Gaussian, the Cartesian coordinates of the ligands were placed into an input file and an optimization calculation was set up at the B3LYP/6-31+G* level of theory. The specific parameter line in the file was:

B3LYP/6-31+g* scf=xqc opt=(calcall)

B3LYP refers to the hybrid density functional theory (DFT) method used in this optimization. Specifically, it refers to the exchange-correlation functional formed from the linear combination of the Becke 88 exchange functional E_x^{GGA}, the Lee-Yang-Parr correlation functional E_c^{GGA}, and the Vosko-Wilk-Nusair local-density approximation E_x^{LDA}.[39] The complete functional is expressed in Equation 2.20, where E_x^{HF} is the HF exchange functional, common to all DFT methods.

$$E_{xc}^{B3LYP} = E_x^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + E_c^{LDA}$$
$$+ a_c(E_c^{GGA} - E_c^{LDA}) \tag{2.20}$$

The 6-31+G* basis set is one of several sets included in Gaussian, which combines the 6-31G set with additional d- and f-type polarization functions on relevant atoms, as well as extra diffuse functions for heavy atoms to better analyze systems containing large anions. SCF refers to the self-consistent field calculation, which repeatedly calculates orbitals from the directly preceding orbital until convergence is achieved. XQC acts similarly to the quadratically convergent

procedure (QC) in that a forced convergence is achieved; however, XQC adds an additional QC step in the event that convergence does not occur in a first-order calculation. Finally, the opt keyword prompts geometry optimization, with calcall indicating that, for the method being used, force constants are computed for every point. Afterwards, the optimized structure was then used for docking procedures.

For each amino acid, three additional input files were prepared for geometry optimization, followed by charge determination. These files are discussed below in the context of methyl-capped alanine. The first of these files, CAPALA, focuses on structural optimization, and contains the parameter line:

Opt=calcall B3LYP/6-31g(d) geom=connectivity n scf=xqc

As in the previous Gaussian input file, "opt=calcall" refers to a calculation where first derivatives (forces) are computed at every step. Similarly, "scf=xqc" still refers to the usage of an additional QC step to increase the odds of convergence if not attained in a first-order SCF calculation. "B3LYP" still refers to the hybrid DFT method discussed previously, and the specified basis set, "6-31g(d)," is functionally similar to 6-31+G*, apart from the omission of the additional diffuse functionals. The "geom=connectivity" term modifies the input source for the optimization; in addition to the default input source, this term adds a second input source containing atom bonding data that specifies unique features in connectivity, including bond orders. This term requires additional information within the com file, structured in the following manner:

M N1 Order 1 N2 Order 2…

Where M is an atom number, Nx are atoms bonded to atom M, and Order indicates the bond order of that specified bond. In the case of CAPALA, said entry appears as:

1 2 2.0 3 1.0 11 1.0
2

3 4 1.0 5 1.0 7 1.0

The latter two input files handle charge distribution of the amino acid, incorporating the

optimized geometry included in the CAPALA output. The first of these input files, "mkcapala",

features the following parameter line:

Pop=mk B3LYP/6-31g(d) geom=check guess=read n scf=xqc

Again, the "B3LYP/6-31g(d)" and "scf=xqc" parameters retain their previous definitions. One of

the terms exclusive to this com file is "pop=mk," which is a parameter that addresses

determination of atomic charges and the handling of species populations. Specifically, this

parameter takes points as chosen by the Merz-Kollman scheme and creates charges that fit to the

electrostatic potential at those points. "geom=check" indicates the use of the checkpoint file

arising from the capala output to determine molecular geometry. Said checkpoint file is also

involved through the "guess=read" parameter, which takes the initial guess required for HF

calculations from the checkpoint file.

The parameterization line of the second input file, nbocapala, reads:

Pop=nbo B3LYP/6-31g(d) geom=check guess=read n scf=xqc

and is nearly identical to that of the mkcapala input file. However, this input specifies a

population parameter that utilizes Natural Bond Orbital[40] analysis ("pop=nbo"), as opposed to

the Merz-Kollman scheme. This type of analysis places a greater emphasis on pairwise electron

distribution in the context of atomic bonding.
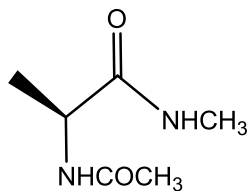
Figure 2-1. Structure of capala.

2.4 Density Functional Theory

To circumvent the inherent challenges and disadvantages associated with the use of wavefunctions to derive energies of a system, an alternative set of methods was built to rely upon physical, observable phenomena. Early attempts at such methods separated the total energy into potential and kinetic components, and involved the use of functionals – that is, functions whose arguments are themselves functions. In this case (Equations 2.21-2.23), the functionals (V and T) relied on a density function, hence giving rise to Density Functional Theory (DFT), a method designed to work independently from wavefunctions.

$$V_{ne}[\rho(\mathbf{r})] = \sum_k^{nuclei} \int \frac{Z_k}{|\mathbf{r} - r_k|} \rho(r) dr \qquad (2.21)$$

$$V_{ee}[\rho(\mathbf{r})] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r_1})\rho(\mathbf{r_2})}{|\mathbf{r_1} - \mathbf{r_2}|} d\mathbf{r_1} d\mathbf{r_2} \qquad (2.22)$$

$$T_{ueg}[\rho(\mathbf{r})] = \frac{3}{10} (3\pi^2)^{2/3} \int \rho^{5/3}(\mathbf{r}) d\mathbf{r} \qquad (2.23)$$

These early attempts were, however, flawed in that they did not account for various key factors, including the omission of the effects of correlation and exchange energies. This was first

corrected through the use of hole functions *h* to account for the shortcomings associated with classical terms as shown in Equation 2.24.

$$\langle\psi\left|\sum_{i<j}^{electrons}\frac{1}{r_{ij}}\right|\psi\rangle = \frac{1}{2}\int\int\frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 + \frac{1}{2}\int\int\frac{\rho(\mathbf{r}_1)h(\mathbf{r}_1;\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 \qquad (2.24)$$

Specifically, *h* corrects for self-interaction error and includes exchange and correlational energies, which were both accounted for in early models by Slater exchange (Equation 2.25), which took advantage of the fact that correlation corrections were insignificant compared to exchange corrections.

$$E_x[\rho(\mathbf{r})] = -\frac{9\alpha}{8}\left(\frac{3}{\pi}\right)^{1/3}\int\rho^{4/3}(\mathbf{r})d\mathbf{r} \qquad (2.25)$$

While these early models found widespread use in solid-state physics, they lacked practical applications in chemistry due to calculation errors and a basis in unproven theories. The transition to modern DFT models began with the proofs of two crucial theorems by Hohenberg and Kohn. The first, the H-K Existence Theorem, demonstrates that ground-state density can be used to determine external potential; the second, the H-K Variational Theorem, states that this aforementioned density adheres to a variational principle, and that the energy expectation value can be calculated as shown in Equation 2.26.

$$\langle\psi_{cand}|H_{cand}|\psi_{cand}\rangle = E_{cand} \geq E_0$$
$$(2.26)$$

This appears to be no different from MO theory, as reliance on a Hamiltonian (and thus a wave function) to calculate energy remains an issue. The difference occurs when the Hamiltonian is treated as an operator for non-interacting electron systems. By choosing a non-interacting version of the system of interest as a starting point – with the same ground-state density as the original system – one can write the resulting energy functional as

$$E[\rho(\mathbf{r})] = T_{ni}[\rho(\mathbf{r})] \; V_{ne}[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] + \Delta T[\rho(\mathbf{r})] + \Delta V_{ee}[\rho(\mathbf{r})]$$

(2.27)

Where T_{ni} is the kinetic energy of the noninteracting system, V_{nc} is the nuclear-electron interaction, V_{cc} is classical repulsion between electrons, and ΔT and ΔV are corrections to kinetic and repulsive contributions, respectively. These terms can be rewritten as

$$E[\rho(\mathbf{r})] = \sum_i^N \left(\langle \chi_i | -\frac{1}{2}\nabla_i^2 | \chi_i \rangle - \langle \chi_i \left| \sum_k^{nuclei} \int \frac{Z_k}{|r_1 - r_2|} \right| \chi_i \rangle \right)$$

(2.28)

$$+ \sum_i^N \langle \chi_i \left| \frac{1}{2} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' \right| \chi_i \rangle + E_{xc}[\rho(\mathbf{r})]$$

where E_{xc}, the term that combines ΔT and ΔV, is our exchange-correlation energy. By finding orbitals χ that minimize E (in a manner akin to HF theory), we satisfy

$$h_i^{KS}\chi_i = \varepsilon_i \chi_i$$

(2.29)

Where h_I^{KS} is defined as

$$h_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_k^{nuclei} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' + V_{xc}$$

(2.30)

$$V_{xc} = \frac{\delta E_{xc}}{\delta \rho}$$

(2.31)

DFT largely differs from HF and MO theory with regard the degree of approximation; whereas HF theory is an approximation constructed from exact components, DFT is an exact theory whose equations must then be solved approximately. Compared to other methods, DFT is, as a whole, computationally efficient, but handling of noncovalent interactions and – more importantly – inaccuracies based on the exchange interaction approximated confer limitations.

2.4.1 Jacob's Ladder

The ongoing pursuit of complete chemical accuracy with regard to DFT can best be depicted

through a Biblically-inspired concept known as Jacob's ladder, which depicts a hierarchy of

density approximations in which each upward rung builds upon the previous rungs, slowly

ascending towards a "heaven" of chemical accuracy (Figure 2-2). In the context of DFT,

continually climbing the ladder by adding more terms and corrections to the exchange-

correlation functional is a primary, if not THE primary, goal. The major limitation, of course, is

tied to computational efficiency and resources; computational chemists can only climb the ladder

as fast as developing theory and computational power might let them. Nonetheless, Jacob's

ladder succinctly demonstrates the overall strategy with regard to DFT; by incorporating

systematic, iterative improvements to current functionals, the methodology can continually

improve until the ideal goal of total accuracy can be reached. In the next section, the rungs of

Jacob's ladder – local density approximation (LDA), generalized gradient approximation (GGA),

meta-GGA, and exact exchange (EXX) – will be discussed in detail.

Chemical Accuracy
(1 kcal/mol)

EXX – Partial Exact Correlation (Generalized RPA)
EXX – Correlation (hyperGGA)
metaGGA
GGA
LDA
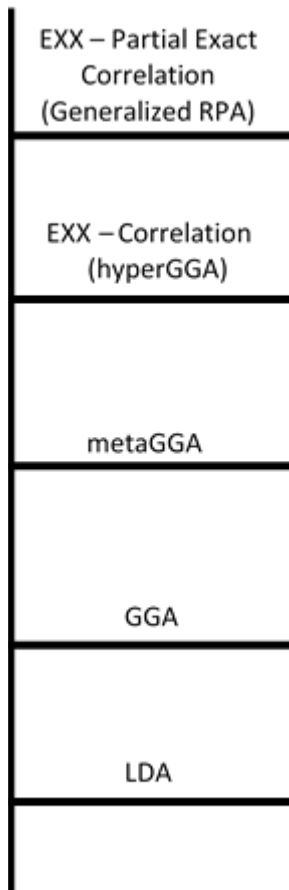
Hartree World

Figure 2-2. Jacob's Ladder; a recurring goal of DFT centers around ascending the ladder

2.4.2 Exchange-Correlation Functionals

Improvement of the accuracy of DFT – and ascension of Jacob's ladder - involves the determination of the exchange-correlation functional E_{xc}. Recall that exchange energy refers to the interactions between identical particles, and that correlation energy stems from interactions between electrons in a quantum system, usually quantified by the difference between the HF limit of a system and its exact energy. This functional is dependent on electron density such that

$$E_{xc}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{xc}[\rho(\mathbf{r})]d\mathbf{r} \qquad (2.32)$$

where the energy density ε_{xc} is defined as

$$\varepsilon_{xc}[\rho(\mathbf{r})] = -\frac{9\alpha}{8}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}\rho^{1/3}(\mathbf{r}) \qquad (2.33)$$

and electron spin, a concept not mentioned above but crucial for later calculations, is

$$\xi(\mathbf{r}) = \frac{\rho^{\alpha}(\mathbf{r}) - \rho^{\beta}(\mathbf{r})}{\rho(\mathbf{r})} \qquad (2.34)$$

Several general strategies have been employed to approximate the exchange-correlation functional, each represented by a successive rung on Jacob's ladder. The first, the local density approximation (LDA), involved the determination of ε_{xc} at a position r solely from the local density value. While this could technically be used in any system where a position results in one specific density, functionals determined in this way are usually derived from uniform electron gas systems. In doing so, the exchange energy can be calculated as

$$\varepsilon_x[\rho(\mathbf{r}),\xi] = \varepsilon_x^0[\rho(\mathbf{r})] + \{\varepsilon_x^1[\rho(\mathbf{r})] - \varepsilon_x^0[\rho(\mathbf{r})]\}\left[\frac{(1+\xi)^{4/3} + (1-\xi)^{4/3} - 2}{2(2^{\frac{1}{3}} - 1)}\right] \qquad (2.35)$$

Where spin is set to 0 for an unpolarized system, and quantified elsewhere (in the latter case, the approximation is referred to as local spin density approximation [LDSA]). Correlation energy is noticeably more difficult to procure. Ceperly and Alder employed MC techniques to calculate correlation energy to high degrees of accuracy, and Vosko et al. proposed local functionals based on those calculations:

$$\varepsilon_c^i(r_S) = \frac{A}{2}\left\{ ln\frac{r_S}{r_s + b\sqrt{r_s + c}} + \frac{2b}{\sqrt{4c - b^2}}\tan^{-1}\left(\frac{\sqrt{4c - b^2}}{\sqrt{r_s} + b}\right)\right. \tag{2.36}$$

$$-\frac{bx_0}{x_0^2 + bx_0 + c}\left\{ln\left[\frac{\left(\sqrt{r_s} - x_0\right)^2}{r_s + b\sqrt{r_s} + c}\right]\right.$$

$$\left.\left.+ \frac{2(b + 2x_0)}{\sqrt{4c - b^2}}\tan^{-1}\left(\frac{\sqrt{4c - b^2}}{2\sqrt{r_s} + b}\right)\right\}\right\}$$

Where A, x_0, b, and c are a set of empirical constants, and r_s is defined as

$$r_S(\text{r}) = \left(\frac{3}{4\pi\rho(\text{r})}\right) \tag{2.37}$$

The result is unintuitive and mathematically difficult, but provides a good treatment of systems with spatially uniform electron densities.

In most molecular systems, however, non-uniform densities are to be expected. As such, improving on LDA by incorporating local change of the density may lead to increased accuracy. By including such a correction, the generalized gradient approximation (GGA) is developed:

$$\varepsilon_{x/c}^{GGA}[\rho(\text{r})] = \varepsilon_{x/c}^{LSD}[\rho(\text{r})] + \Delta\varepsilon_{X/C}\left[\frac{|\nabla\rho(\text{r})|}{\rho^{\frac{4}{3}}(r)}\right] \tag{2.38}$$

The Becke (B) exchange functional was among the first to incorporate a gradient correction, featuring accurate asymptotic behavior for energy densities at a significant range as well as an optimized empirical parameter. Other exchange functionals followed suit, though some instead forwent empirical parameters and focused on rational function expansions (B86, LG, P, PBE, mPBE). Popular correlation functionals, meanwhile, directly relied on either 8.29 (such as B88, P86, or PW91), or on computed correlation energies entirely rather than included correction

terms (such as LYP). Combining functionals from either set (such as B and LYP to form BLYP) produced complete exchange-correlation functionals for use in the literature.

Current functional development extends beyond LDA and GGA. One example involves meta-GGA, which incorporates the second derivative of the local density. Both exchange (BR) and correlation (Lap) functionals were developed to go beyond the gradient corrections of GGA, though any improvement in performance of metaGGA over GGA can sometimes be negated by calculation instability (such as in the case of computing the Laplacian of relevant densities). Functionals can also be constructed to better control the extent of interaction, such as in the case of the adiabatic connection method (ACM), which relies on the fact that, based on the Hellman-Feynman theorem,

$$E_{xc} = \int_0^1 \langle \Psi(\lambda)|\mathbf{V}_{xc}(\lambda)|\Psi(\lambda)\rangle d\lambda \tag{2.39}$$

Where λ is the extent of interelectronic interaction (from 0 to 1), producing an integral of the expectation value curve such that

$$E_{xc} = E_x^{HF} + z(E_{XC}^{DFT} - E_X^{HF}) \tag{2.40}$$

Or, if we choose a = 1 - z,

$$E_{xc} = (1-a)E_{XC}^{DFT} + aE_x^{HF} \tag{2.41}$$

The ACM is one subtype of the exact-exchange (EXX) method. Several popular examples include the B3PW91 functional, which adopts a 3-parameter scheme while combining B and PW91 functionals:

$$E_{xc}^{B3PW91} = (1-a)E_{XC}^{LSDA} + aE_x^{HF} + b\Delta E_x^B + E_c^{LSDA} + c\Delta E_C^{PW91} \tag{2.42}$$

Where a, b, and c are calculated optimized values (0.20, 0.72, 0.81). This was eventually

modified to the B3LYP functional expressed in Equation 2.43, which has since become one of

the more prominent functionals in existence due to its accuracy and performance, and is the

functional most commonly used in this work.

$$E_{xc}^{B3LYP} = (1-a)E_{XC}^{LSDA} + aE_x^{HF} + b\Delta E_x^B + (1-c)E_c^{LSDA} + c\Delta E_C^{LYP}$$

$$(2.43)$$

.

2.5 Docking

Docking involves the placement of ligands onto a particular active site, analyzing and

subsequently scoring its favorability to binding in its current orientation. This method plays a

large role in structure-based drug design, given that evaluation of ligands may reveal novel

inhibitors or other medicinally useful molecules. Multiple scoring functions have recently been

developed to evaluate large numbers of ligands. Several examples include Bohm's function[41],

which provided a faster, more flexible means of binding affinity evaluation, and VALIDATE[42],

which sacrifices much of the speed gained in the Bohm approach to increased accuracy (in

training set evaluations, VALIDATE obtained a cross-validated estimate of error of 6.5 kJ/mol,

opposed to the 9.3 kJ/mol of Bohm's method). Newer docking methodologies are based on

modifications of the ChemScore[43] function, a function that closely follows the philosophy behind

Bohm's design while improving upon it. Such improvements come largely from the use of a

large protein-ligand complex database, leading to increased robustness and reliability without

compromising speed (8.68 kJ/mol vs 9.3 kJ/mol).

2.5.1 ChemScore

The overall ChemScore scoring function is shown in Equation 2.44.

$$\Delta G_{bindng} = \Delta G_0 + \Delta G_{hbond} \sum_{iI} g_1(\Delta r) g_2(\Delta \alpha)$$

$$+ \Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{1L} f(r_{1L}) + \Delta G_{rot} H_{rot}$$

(2.44)

Beyond ΔG_0, the equation contains four terms, each describing an atom type to which ligand and receptor atoms are assigned. In each term, ΔG was determined via multiple linear regression; we will instead discuss the remaining components of each term here. In the first term, we define the hydrogen bond term as including all possible hydrogen bonds that occur between any ligand and receptor atom. These functions are defined similarly to the Bohm piecewise hydrogen bonding term shown in Equations 2.45 and 2.46.

$$g_1(\Delta r) \begin{cases} 1 & \text{if } \Delta r \leq 0.25 \text{ Å} \\ 1 - \dfrac{(\Delta r - 0.25)}{0.4} & if\ 0.25\text{Å} < \Delta r \leq 0.65 \text{ Å} \\ 0 & if\ \Delta r > 0.65 \text{ Å} \end{cases}$$

(2.45)

$$g_2(\Delta \alpha) \begin{cases} 1 & if\ \Delta \alpha \leq 30° \\ 1 - \dfrac{(\Delta \alpha - 30)}{50} & if\ 30° < \Delta \alpha \leq 80° \\ 0 & if\ \Delta \alpha > 80° \end{cases}$$

(2.46)

The terms Δr and $\Delta \alpha$ involve deviations of bond angles from ideal values. The second term, the metal term, includes all acceptor and acceptor/donor atoms as well as all metal atoms. The simple contact function, f(r), is represented in Figure (2-3(a)). The third term, the lipophilic term, includes a similar function as the metal term shown in Figure (2-3(b)).
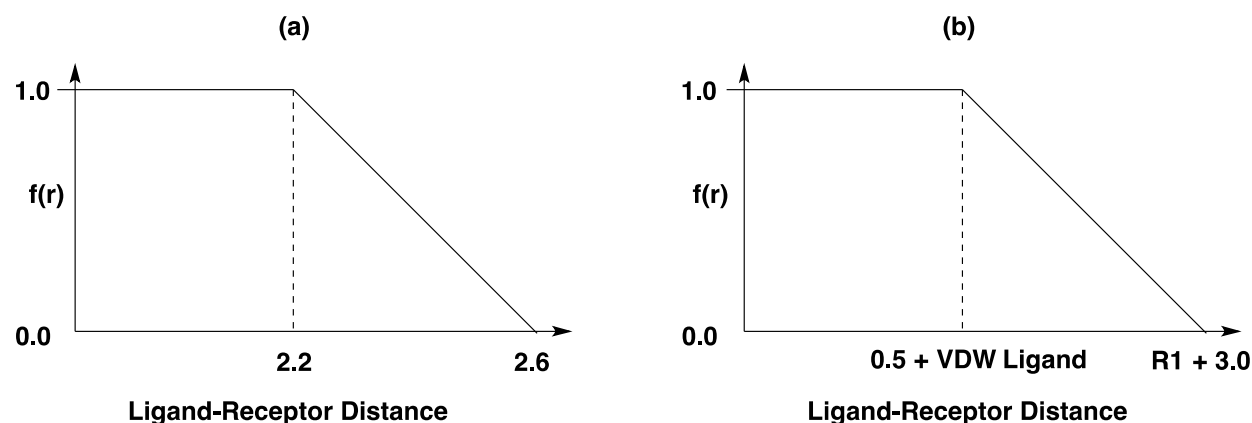
Figure 2-3. Contact function f(r) included in both the (a) metal and (b) lipophilic terms, where "VDW ligand" represents the van der Waals distance of the ligand atom. Distances on the x-axis are in Å.[43]

Finally, the fourth term includes frozen rotational bonds, which involve sp^3-sp^3 and sp^3-sp^2 bonds (involving non-terminal groups) where all involved atoms are impaired by the receptor, introducing significant entropic terms. Given the differing contributions by lipophilic and non-lipophilic atoms, H_{rot} is defined in Equation 2.47 below, where N_{rot} counts frozen rotatable bonds and the P and P' functions count percentages of non-lipophilic heavy atoms on either side of said bonds.

$$H_{rot} = 1 + (1 - 1/N_{rot}) \sum_{r} (P_{nl}(r) + P'_{nl}(r))/2 \tag{2.47}$$

2.5.2 SP Glide

ChemScore provides a valuable basis upon which to build improved scoring functions to better study binding affinity in different cases. The grid-based ligand docking with energetics (Glide) scoring function[44], for example, incorporates modifications to the ChemScore function that perform a more comprehensive ligand analysis without sacrificing significant speed. The standard precision (SP) Glide function accomplishes this through the use of a "funnel" strategy, using hierarchical filters to identify ligand orientations within a particular active site. After filtering potential ligands using the physical properties of the receptor, an exhaustive set of

ligand conformations are used to conduct prescreening of the phase space relevant to each

ligand, reducing the amount of intensive computational work that will be required for each

ligand. Ligand minimization using an OPLS-AA force field followed by Monte Carlo analysis of

torsional minima results in a select few hits, which are then subjected to scoring by the

GlideScore function, an adjustment of the ChemScore function unique to SP Glide shown in

Equation 2.48.

$$
\begin{aligned}
\Delta G_{bind} = \ & C_{lipo-lipo} \sum f(r_{ir}) \\
& + C_{hbond-neut-neut} \sum g(\Delta r)h(\Delta \alpha) \\
& + C_{hbond-neut-charged} \sum g(\Delta r)h(\Delta \alpha) \\
& + C_{hbond-charged-charged} \sum g(\Delta r)h(\Delta \alpha) \\
& + C_{max-metal-ion} \sum f(r_{lm}) + C_{rotb}H_{rotb} \\
& + C_{polar-phob}V_{polar-phob} + C_{coul}H_{coul} + C_{vdW}E_{vdW} \\
& + \text{solvation terms}
\end{aligned}
\tag{2.48}
$$

All four terms from the ChemScore function can be found here, with some modifications.

The first lipophilic term is kept identical to that of ChemScore. The hydrogen bonding term is

split into three summations, taking into account the charges of the acceptor and donor atoms. The

metal-ligand term, while maintaining the same functional form as its ChemScore counterpart,

only considers interactions where anionic acceptors are present, noting that anionic ligand

coordination is typically favored. Furthermore, SP GlideScore only observes the best metal

ligation in cases where multiples are discovered, and factors in increased favorability of an

anionic ligand should the metal ion have a positive charge in the apoprotein. Locked rotational bonds are treated as they were in the ChemScore function.

The first of the three unique Glide SP terms, derived from Schrodinger's active-site mapping facility, tracks the presence of polar, non-hydrogen bonding atoms within hydrophobic areas, and includes their influence appropriately. The next two terms include energies arising from Columbic and VDW interactions between ligand and receptor. These energies were reduced for version 2.5 of GlideScore, as original energies (except in some cases) led to significant inaccuracies. The final new term addresses the solvation requirements of the ligand and protein of interest, as well as their impact on ligand positioning. These solvation effects were obtained by including individual, explicit water molecules in docking simulations.

2.5.3 XP Glide

Despite the noticeable advantages gained by using SP Glide, Friesner et al. determined that additional modification could be done to further enhance the scoring function. The result, named Extra Precision (XP) Glide,[45] improves upon the SP Glide model through the introduction of additional terms that take hydrophobic enclosure of ligand atoms, as well as unique hydrogen bonding motifs, into account. Though more precise, XP Glide sacrifices computational efficiency and as such is an alternative to, rather than a substitute for, SP Glide. The XP Glide scoring function is given in Equation 2.49 with the binding and penalty terms shown in Equations 2.50 and 2.51, respectively.

$$\text{XP Glidescore} = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \tag{2.49}$$

$$E_{bind} = E_{hyd\_enclosure} + E_{hb\_nn\_motif} + E_{hb\_cc\_motif} + E_{coul} + E_{PI} + E_{hb\_pair}$$
$$+ E_{phobic\_pair} \tag{2.50}$$

$$E_{penalty} = E_{desolv} + E_{ligand\_strain} \qquad (2.51)$$

The overall functional form is similar to that of SP GlideScore, with Coulombic and van der Waals energy terms treated similarly to SP Glide, and E_{hb_pair} and E_{phobic_pair} defined as they are in ChemScore. Between the bind and penalty term groups, however, six new terms unique to XP Glide have been added. The first of these new terms, $E_{hydenclosure}$, enhances the treatment of interactions between lipophilic ligand atoms and lipophilic protein atoms by quantifying the impact of local geometry, specifically in the instances where ligand atoms are enclosed by such protein atoms (enclosure requires a 180° angle between protein atoms, as defined by Friesner et al.). The second and third terms, $E_{hb\_nn\_motif}$ and $E_{hb\_cc\_motif}$, include particular neutral-neutral and charged-charged hydrogen bonds that fit a special set of conditions as listed by Friesner et al. (Table (2-2)). As opposed to SP Glide, which substitutes the ChemScore hydrogen-bond term for three terms dividing the pairs by charge relationship, XP Glide maintains the ChemScore H-bond term while accounting for the additional impact of these "special" H-bond motifs, as defined by the set of conditions in Table (2-2). The fourth term, E_{PI}, addresses any instances of π-stacking and similar interactions; though non-trivial, this term typically contributes far less to the binding score than other terms. The next term, E_{desolv}, imposes desolvation penalties through the use of an explicit water model to evaluate the solvation of charged or polar groups on either the ligand or protein. This constitutes the most novel term of XP Glide relative to other scoring functions and as such required modifications to the sampling algorithm to avoid issues when applying such a penalty to random ligands. Finally, the E_{lig_strain} term imposes contact penalties, which account for the change in torsional angles from ideal values by ligands in certain cross-docking models.

The following ligand docking study was conducted with Glide through the Maestro modelling environment. Relevant options were first adjusted under the Settings tab of the Ligand Docking window. Within the "Receptor Grid" subfield, the appropriate grid file was selected. Next, within the "Docking" subfield, the Standard Precision version of Glide was selected, as opposed to HTVS, SP-Peptide, and XP Glide, given that it was appropriate for the quantity of ligands to be screened. Under Options, flexible docking was enabled, as opposed to rigid docking, refinement, or scoring in place. This flexible docking was further adjusted by enabling the sampling of pyramidal nitrogen inversions and ring conformations, as well as setting bias sampling of torsions to "none." Given the use of Epik as part of ligand preparation, appropriate penalties were involved by selecting "add Epik state penalties to docking score." Finally, "enhance planarity of conjugated π groups" was deselected to remove the associated constraint.

Table 2-2. Limitations on the pair types that can be included in the Ehb_nn_motif and Ehb-_cc_motif terms, as listed by Friesner et al.[19]

conditions for not applying the special pair hydrogen-bond scores
ignore poor quality hydrogen bonds ($<0.05 \, E_{hb\_pair}$)
ignore pairs involving the same neutral protein atom
ignore pairs involved in a salt bridge if the electrostic potential at either ligand atom is above the cutoff
ignore salt bridge pairs if either protein atom is involved in a protein−protein salt bridge
ignore ligand donor/donor pairs that come from NH_x, where $x \geq 2$ groups, the nitrogen atom is not in a ring and has no formal charge
ignore formally charged protein atoms with more than eight second-shell waters in the unligated state
ignore charged−neutral hydrogen bonds unless the protein atom is in a salt bridge
ignore pairs of different neutral acceptor atoms on the ligand
neutral hydrogen-bond pairs must satisfy ring atom and hydrophobicity environment criteria as outlined in section 2
ignore ligand hydroxyl to protein hydrogen bonds if the protein atom has zero formal charge

Next, various adjustments were made under the "Ligands" category. Ligands were obtained from a file, using the appropriate .sdf files. The corresponding range was set to cover the entire set of ligands. "Use input partial charges" was unchecked, as they were obtained by default from the associated force field. The defaults for ligand type restrictions – 300 atoms, 50 rotatable bonds – were maintained. Van der Waals radii scaling values were also kept at default, with a scaling factor of 0.8 Å and partial charge cutoff of 0.15 C.

The next several available categories required relatively fewer modifications. Underneath "Core," core pattern comparison was set to "do not use," as no core definition was employed. Within "Constraints," "show markers" was enabled. Furthermore, some constraint flexibility was added by selecting "Must match at least 1" and "test constraint satisfaction only after docking" was unchecked, allowing the constraints to play a role in the actual docking process. Under "Torsional Constraints," "apply torsional constraints" was unchecked. Underneath the "Similarity" tab, "do not use similarity" was selected.

The final "Output" tab, which controlled the successful ligand pose output, also required additional changes. "Write pose viewer file(includes receptor)" was selected to produce a file viewable with Maestro Project Table. The file itself was written in Maestro format. A cap on written poses per docking run was set at 10,000 and a limit on poses per ligand was set to 5. Multiple options enabled by default were unchecked, including "Perform post-docking minimization," "Write per-residue interaction scores," "Write report file," and "Compute RMSD to input ligand geometries." After completing these and the preceding steps, the ligand docking job was started. The job name was set as "[database name]_[sdfid]_SP," and the host was set to "submit_8x1x1_150hrs." On the monitor window, job submission was then confirmed.

CHAPTER 3
MOLECULAR DYNAMICS STUDIES OF HP UREASE IN AQUEOUS SOLUTION

The following chapter has been adapted from *Molecular Dynamics Study of Helicobacter pylori Urease,* published in the Journal of Chemical Theory and Computation in 2014 (http://pubs.acs.org/doi/abs/10.1021/ct5000023)[46], with permission from the American Chemical Society.

3.1 Introduction

Urease has received computational interest in recent years, via both Quantum Mechanical (QM) and Molecular Dynamics (MD) approaches. Musiani et al. used density functional and docking methods in a 2001 study to further investigate the mechanism of *B. pasteurii* urease hydrolysis and used steric and electrostatic criteria to distinguish the two most likely mechanistic pathways.[47] Both MD and QM studies have been utilized in our laboratory in order to probe the pathway of urea decomposition in an attempt to further elucidate the mechanism of catalysis, with the MD study focusing on *B. pasteurii* urease.[48] Carlsson and Nordlander used density functional theory to probe the nature of urea coordination to the nickel active site while studying the hydrolytic mechanism.[49] Valdez and Alexandrova utilized their QM/DMD method in order to probe the natural selection of two Ni^{2+} ions in the urease enzyme active site, whereas nature selected two Zn^{2+} ions for β-lactamases, enzymes that share similar hydrolytic function.[50] Barros et al. studied the Jaburetox-2Ec protein, derived from urease, which has potent insecticidal properties in another MD study.[51] An older study by Manunza et al. probed only the active site of *K. aerogenes* urease in their investigation of urea and the urease inhibitors N-(*N-butyl*)-phosphoric triamide and acetohydroxamic acid binding to the dinickel site.[52]

These urease enzymes share a conserved set of residues comprising a mobile flap positioned over the active site. Previous MD simulations on *K. aerogenes* urease, a much smaller

enzyme than *H. pylori* urease, revealed a third flap state we refer to as the wide-open flap state which is notably distinct from the previously known open and closed flap states.[53] Inspired by our simulations on the urease of *K. aerogenes*, we conducted MD simulations on *H. pylori* urease in order to ascertain whether a wide-open flap state is observed in this enzyme as well. Furthermore, we analyzed other regions of the enzyme and finally investigated the behavior of the sodium. Herein we present the results of the first MD simulation on *H. pylori* urease, which spans a 400 ns duration on this large protein (148,648 atoms in the protein and 307,839 in total). Importantly, this simulation provides us with up to 4.8 μs (400ns * 12 flaps) of flap motion information if the motions are not correlated (weak, if any, correlation has been noticed between the flaps based on this simulation).

3.2 Methods

A 400 ns molecular dynamics simulation was run on the urease of *Helicobacter pylori*. This was conducted using the PMEMD version of Amber 12[21a, 54] on an M2090 GPU using the FF99SB force field. We used the PDB structure 1E9Z as the starting point.[6] The PDB contains one twelfth of the total structure, therefore PyMOL[55] was used in order to reflect the structure eleven times, thereby generating the entire dodecameric HP urease structure. The online protonation server H++ was used to obtain the correct protonation states of the amino acids throughout the structure at a pH of 7.4.[27, 56] We took extra care with the nickel coordination site, which is located at the bottom of each of the twelve active sites. A bonded model was created with the aid of the tLeap facility of AmberTools version 1.5[21b] where bonds involving Ni^{2+} ions were defined; subsequently, hydrogen atoms and water molecules were added to the entire structure. The prepared model was neutralized with the addition of 44 Na^+ ions and solvated with a periodically replicated octahedral water box using tLeap using the TIP3P triangulated water model. A total of 53,048 water molecules were used to hydrate the structure. The metal

parameters developed by Roberts et al. with the MTK++/MCPB utility of AmberTools for *K.*

aerogenes urease[57] were modified by utilizing previously developed Lennard-Jones parameters[58]

($R*$ and ε) for Zn^{2+} as parameters for the Ni^{2+} radii.

The energy minimization of the structure was performed using a two-stage procedure: we

first minimized the nickel coordination sites by imposing weak harmonic positional restraints of

10 kcal/(mol•$Å^2$) on all atoms outside the Ni coordination sphere. The steepest descent method

was first used for $1x10^5$ steps to minimize the active site. Subsequently, the entire structure was

relaxed and minimized using steepest descent for $1x10^4$ steps. The initial step length was

decreased to $1x10^{-6}$ Å. After minimization we equilibrated the structure employing a two-stage

process. In the first stage, the temperature was raised from 0 to 300 K over $1x10^6$ steps of MD

with a step size of .002 ps (2 fs) in the canonical (NVT) ensemble. This applied a weak harmonic

positional restraint on the whole protein. After the system was brought to 300 K, the simulation

was run for 10 ps in the isobaric, isothermal (NPT) ensemble after removal of all the harmonic

restraints. The temperature control was performed using Langevin dynamics with a collision

frequency γ of 2.0 ps^{-1} in the first and 1.0 ps^{-1} in the second equilibration stage. SHAKE[37] was

used to constrain all hydrogen-containing bonds during both equilibration steps.

The production MD run was procured over 400 ns in the isobaric, isothermal (NPT)

ensemble, which was acquired using $2.0x10^8$ steps, with a 0.002 ps time step. The temperature

was kept constant at 300 K using Langevin dynamics with a collision frequency of 1.0 ps^{-1} while

the pressure was maintained at 1 bar with a pressure relaxation time of 2.0 ps. SHAKE was again

used to constrain all hydrogen-containing bonds. For the calculation of the non-bonded

interactions, we made use of the particle mesh Ewald method and a cutoff distance of 8 Å was

employed while computing the long-range electrostatic interactions. Frames were saved every

5×10^3 steps (10 ps), providing 4×10^4 frames. The frames were analyzed for flap state-defining residue separations, root mean square deviation (RMSD), atomic fluctuations (RMSF) and correlation matrices using the ptraj utility of AmberTools version 1.5. RMSD was obtained for each atom in the entire protein with the first frame of the trajectory as the reference and gives an average deviation of the protein geometry from the reference structure at each point over the simulation time. We also separated out the residues involved in the motion of the active site-covering flaps and obtained the RMSD for each of these. The RMSF was computed on a per residue basis, using the alpha carbon of each residue as the reference point and providing a time-averaged value.

Throughout our discussion of the residues we will identify both the residue numbering according to Ha et al. (α and β chains), as well as the residue numbering from the PDB (1-807). For the purposes of this analysis the flap covering the active site was considered to span residues α304-347 (542-585) with the first α-helix composed of residues α304-322 (542-560), the turn composed of α323-329 (561-567) and the second α-helix spans α330-347 (568-585). The α-helices are extended from the helical section described by Ha et al.[6] in order to account for all residues in the flap region that adopt α-helical character at some point during the simulation.

3.3 Results

3.3.1 Root Mean Square Deviation

The RMSD (Figures 3-1 through 3-5) was obtained for the entire structure and of each flap individually using the first frame after equilibration as the reference point. The RMSD of the entire structure levels off approximately 250 ns into the simulation at a value of approximately 2.5 Å. For the flap that remains closed (flap 1), the maximum RMSD found over time was 1.87 Å and overall this flap undergoes only modest fluctuation. For the semi-open flap state displayed in Figure (3-1) (flap 5), the maximum observed RMSD is approximately 3 Å; Flap 5 was

observed to have an RMSD that did not vary appreciably from 2.6 Å after 150 ns. In the wide-open flap (flap 11) the RMSD reaches 5.25 Å and after two-thirds of the simulation remains around 3.8 Å before dipping slightly over the final 50ns, although at a value well above the total simulation RMSD. Further RMSD plots for the remaining flaps are provided in Figures (3-1)-(3-4).
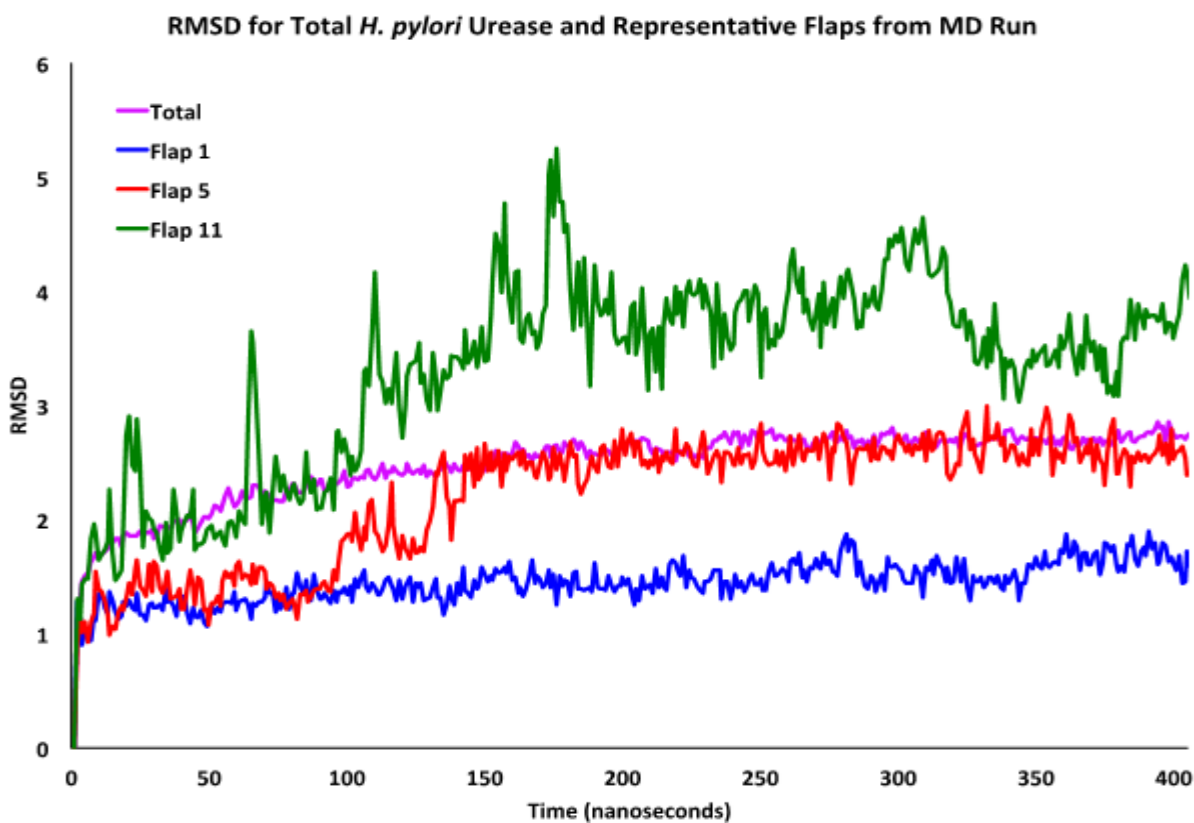
Figure 3-1. RMSD of the entire *H. pylori* urease structure (purple), and the closed (blue), semi-open (red) and wide-open (green) flap states.
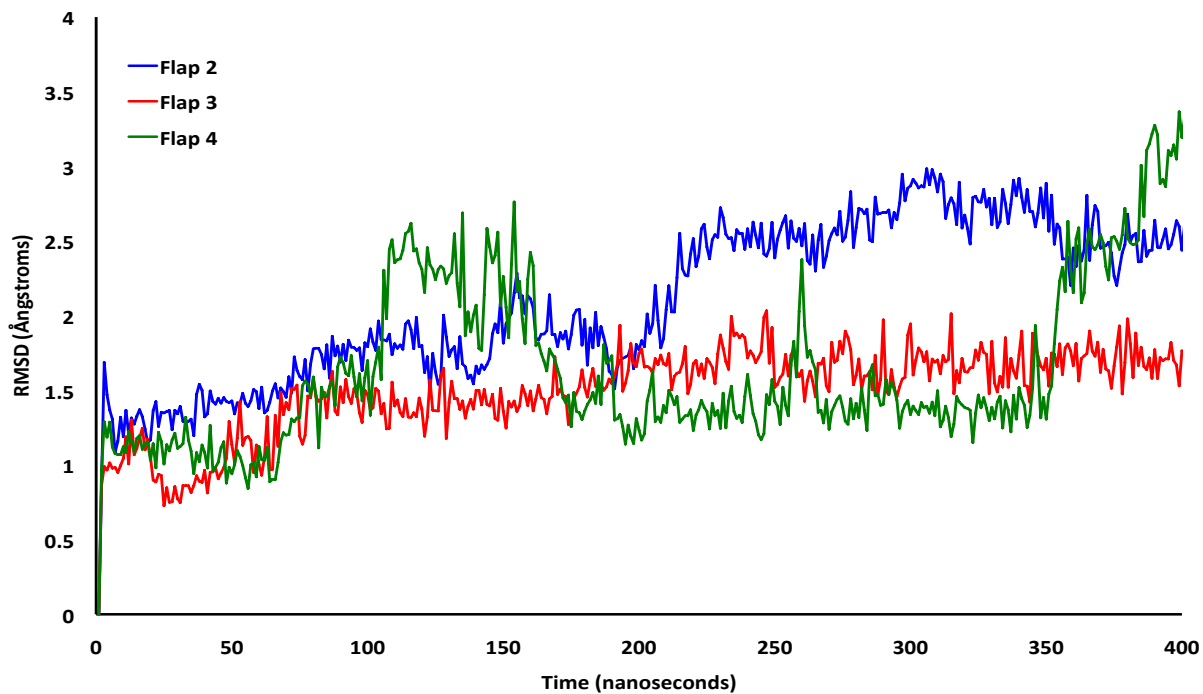
Figure 3-2. RMSD for *H. pylori* urease flaps 2 (blue), 3 (red) and 4 (green).
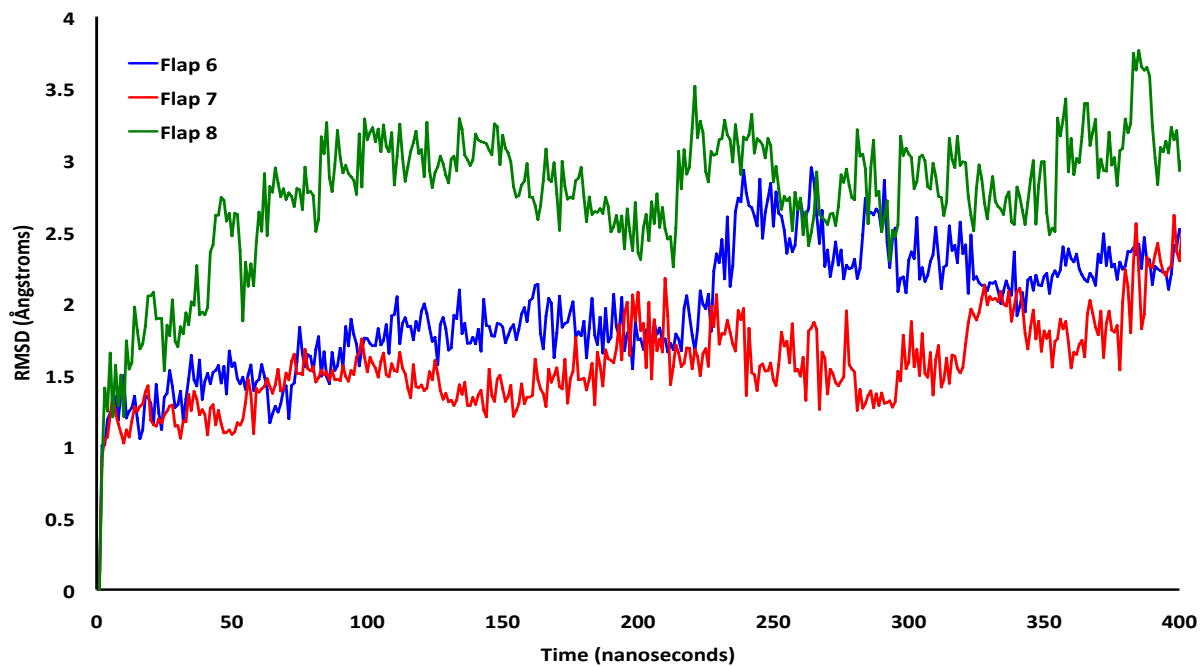
Figure 3-3. RMSD for *H. pylori* urease flaps 6 (blue), 7 (red) and 8 (green).
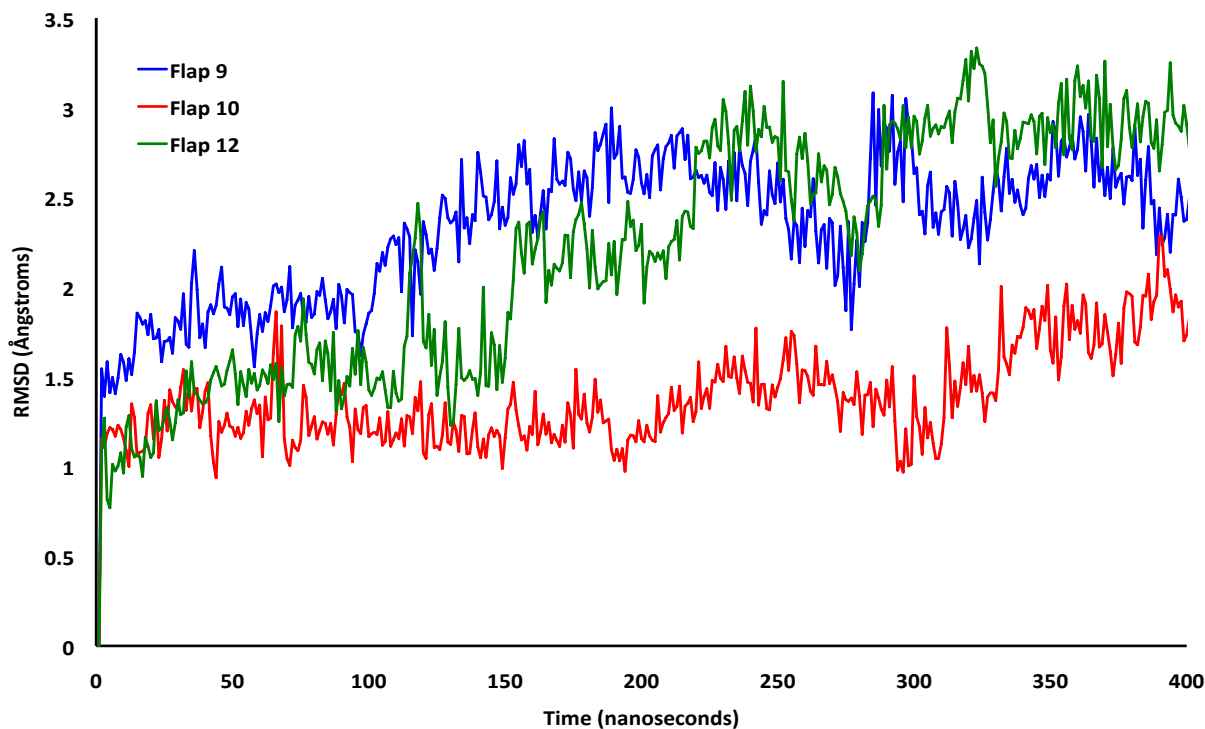
Figure 3-4. RMSD for H. pylori urease flaps 9 (blue), 10 (red) and 12 (green).

3.3.2 Root Mean Square Fluctuation

The RMSF of each dimer was obtained and representative results are shown in Figure (3-5) for the three flap states (remaining RMSF plots are provided in Figures (3-6)-(3-8)). RMSF was determined based on the α-carbon of each residue. We define "average RMSF plus one standard deviation" as the threshold for a residue considered as "high RMSF." Recall that HP urease consists of twelve identical αβ dimers, for a total of 24 chains. Each dimer contains one binding pocket with two Ni^{2+} ions and one flap, which will be referred to as flaps 1-12 to differentiate between them. After plotting the RMSF of each dimer it was apparent that there were repeating regions of high flexibility.

It was expected that the residues of the flap region covering the active site, α304-347 (542-585) would have high RMSF values and this was indeed true for all twelve dimers, with the exception of the first, where the flap was observed to remain essentially closed. The other

regions that have high RMSF values, almost in every dimer, were residues β100-106 (100-106), β226-232 (226-232), α60-65 (298-303), α388-395 (626-633), and α538-545 (776-783) (Table (3-1)). The β100-106 (100-106) region lies within a bridge in a small group of amino acids that link two approximately 100 amino acid regions. Near the end of the β-chain are residues β226-232 (226-232) that make up one of the 8 vertices of the dodecameric structure. These amino acids are part of a C-terminus extension of the beta chain as compared to the ureases of *K. aerogenes* and *B. pasteurii*.[6] Residues α60-65 (298-303) are located on the interior of the dodecamer near a vertex where ion density is observed in the MD simulation (*vide infra*). Finally, the residues α538-545 (776-783) comprise the loop portion of an ancillary flap close in proximity to the flap covering the active site. Of note is the observation that in the dimer where the flap remains closed, these residues have the highest observed RMSF of any of the remaining 11 dimers. In fact, in this dimer each residue from α533-553 (771-791) is observed to have
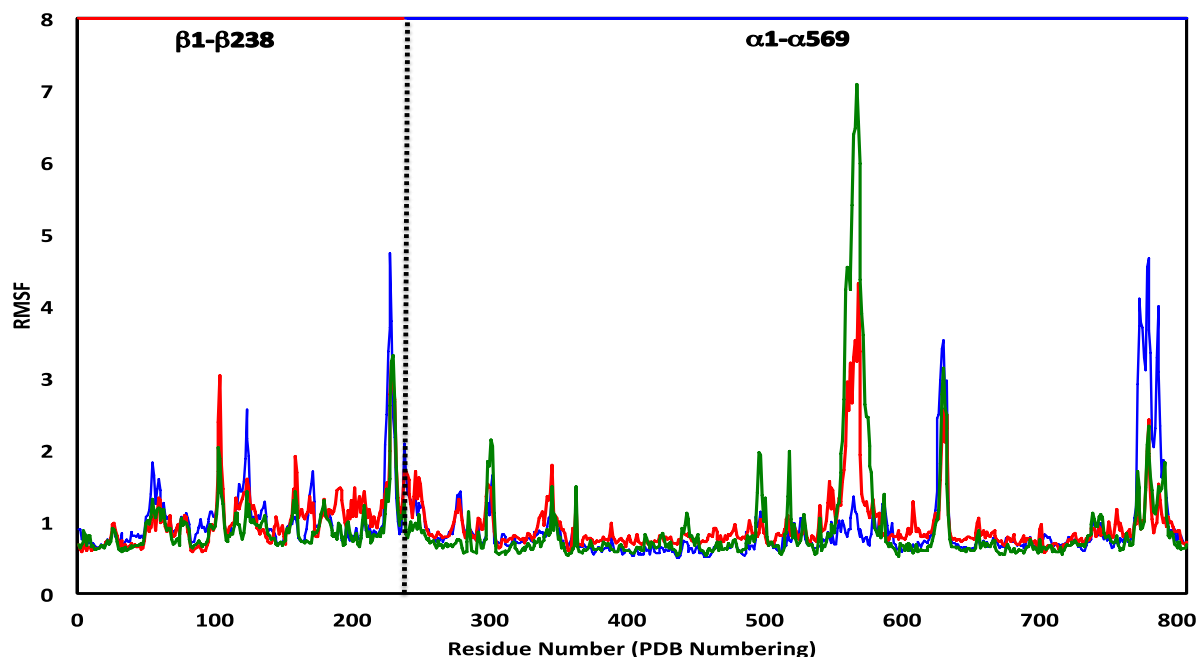
Figure 3-5. RMSF for αβ dimers representing the wide-open (green), semi-open (red) and closed (blue) flap states with sequential numbering from the PDB (1-807). Residues 1-238 correspond to the β-chain and residues 239-807 form the α-chain.

significantly higher RMSF values. Based on an analysis of the trajectory, this final high RMSF region is most mobile when the active site covering flap is closed, and exhibits the least amount of motion when the flap is in the wide-open state.
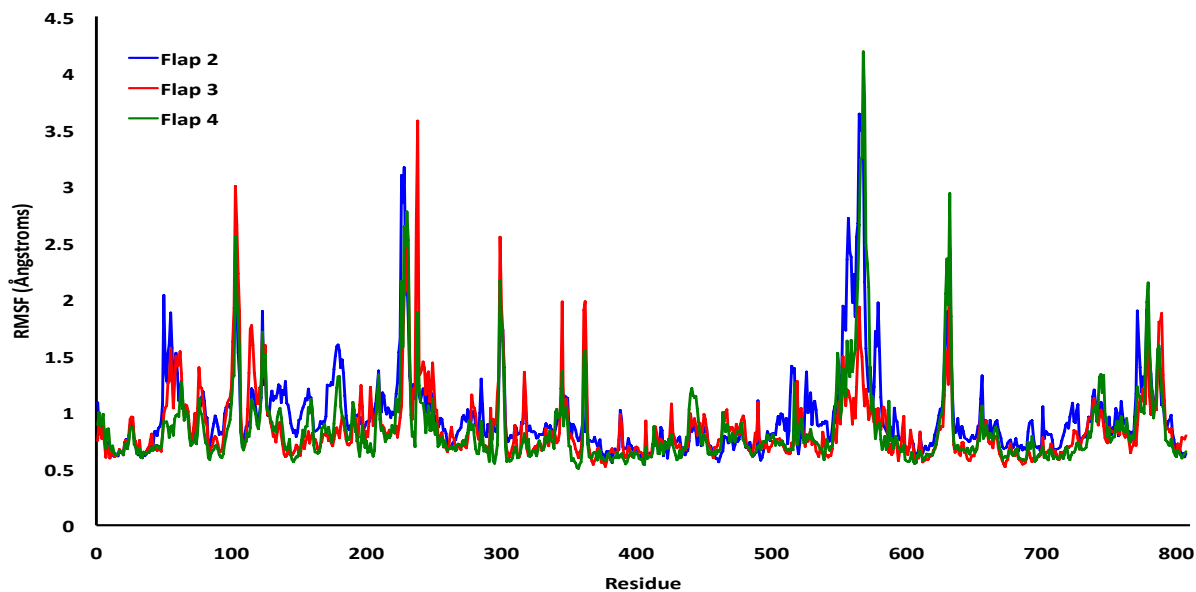
Figure 3-6. RMSF for *H. pylori* urease dimers 2 (blue), 3 (red) and 4 (green).
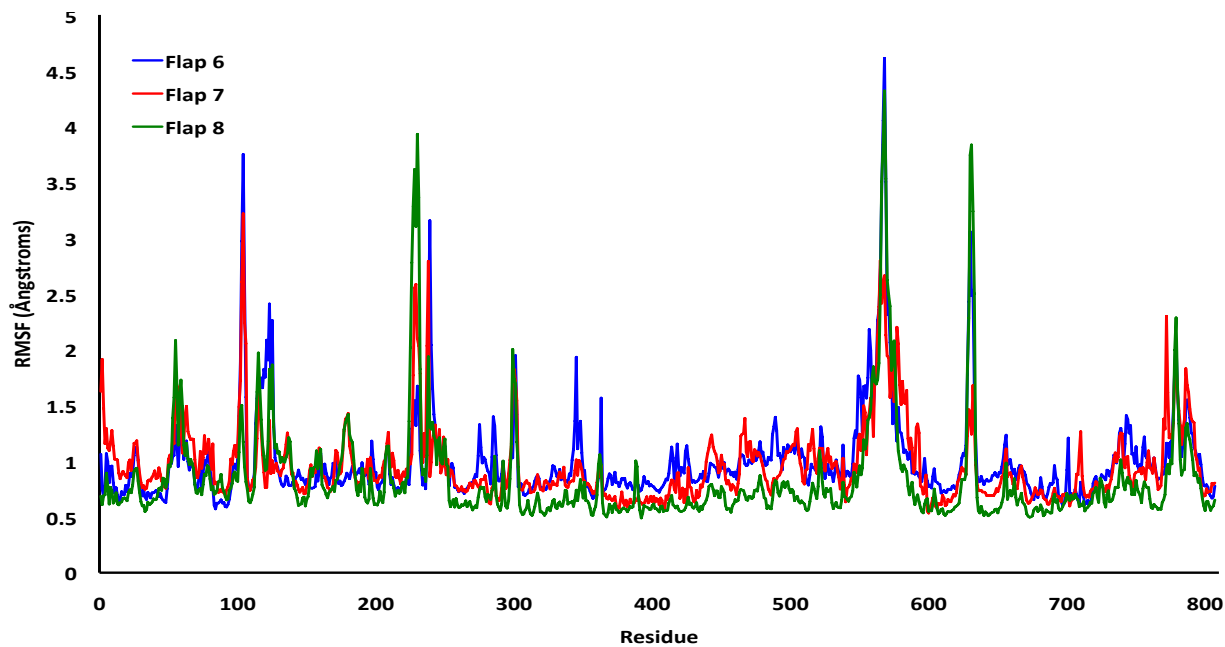
Figure 3-7. RMSF for *H. pylori* urease dimers 6 (blue), 7 (red) and 8 (green).
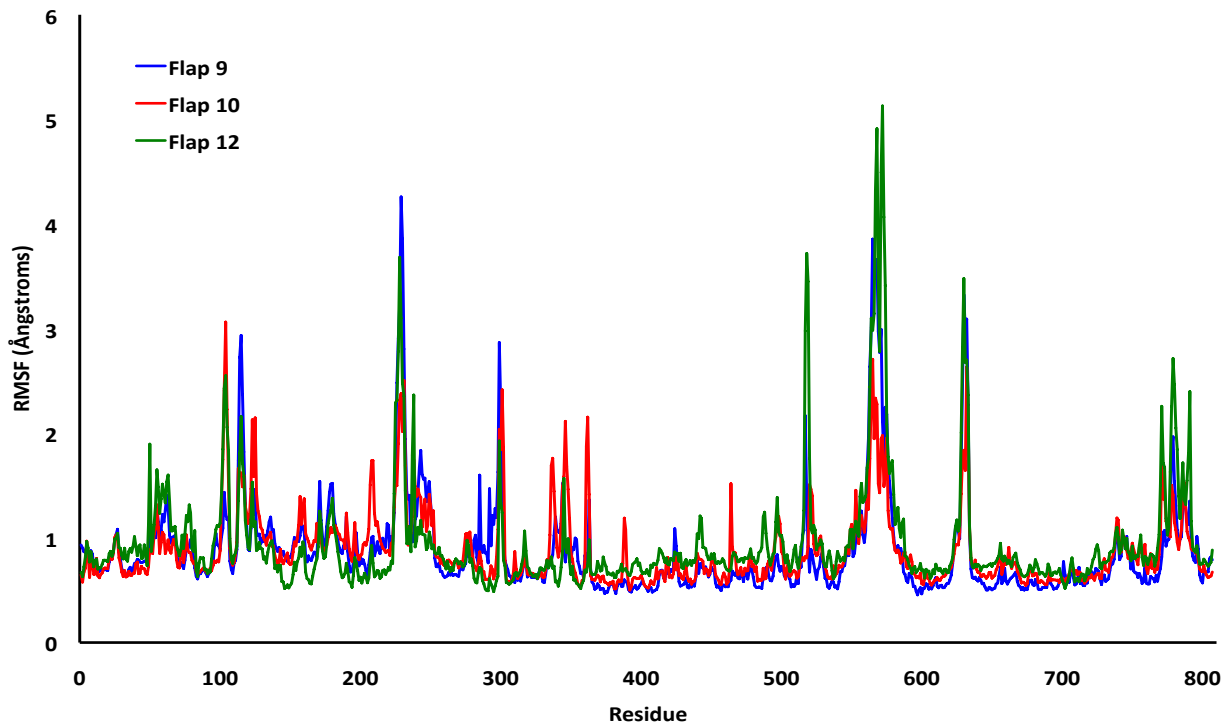
Figure 3-8. RMSF for *H. pylori* urease dimers 9 (blue), 10 (red) and 12 (green).

Figure (3-9) depicts the B factors from the X-ray structure of Ha et al.[6] plotted against scaled RMSF values for the dimer that adopts the wide-open active site flap conformation. The RMSF profile matches up quite well with the experimental B factors, which supports the reliability of our molecular dynamics simulation. One of the particularly interesting high RMSF regions spanned residues α388-α395 (626-633) and this group of amino acids constitutes the loop portion of what has been identified as a second mobile flap in the active site region (Figure (3-10)). This flap has a profile similar to the experimental data for inhibitor complexes of *B. pasteurii*[38] and *S. pasteurii*[39] reported by Benini et al. This flap consists of two α-helices spanning residues α372-α387 (610-625) and α398-α410 (636-648) and the loop portion, just as the flap that covers the active site. One of the key differences observed in the two flaps during the MD simulation was that while both α-helices were observed to lose α-helical character in the active site covering flap, the helices are quite rigid in this second flap. This loop is observed to

be highly flexible in each of the twelve αβ dimers over the course of the simulation. This flap is positioned near one of the vertices that allow direct access of molecules into the hollow, suggesting it may affect entrance into and egress from the internal cavity. While a number of hypotheses are possible we speculate that this flap may serve as an entrance/exhaust conduit that allows for the exit of hydrolysis products from the active site. Analysis of the MD trajectory reveals one sodium ion clearly passing through the first flap (residues 771-791; α533-α553) and other ions briefly interacting. Due to the large accumulation of sodium ions in the hollow, the low occurrence of Na^+ ion interaction with this flap may be a concentration issue and studies underway with higher ion concentration (Na^+ and NH_4^+) would be better able to support or reject the proposed conduit hypothesis.
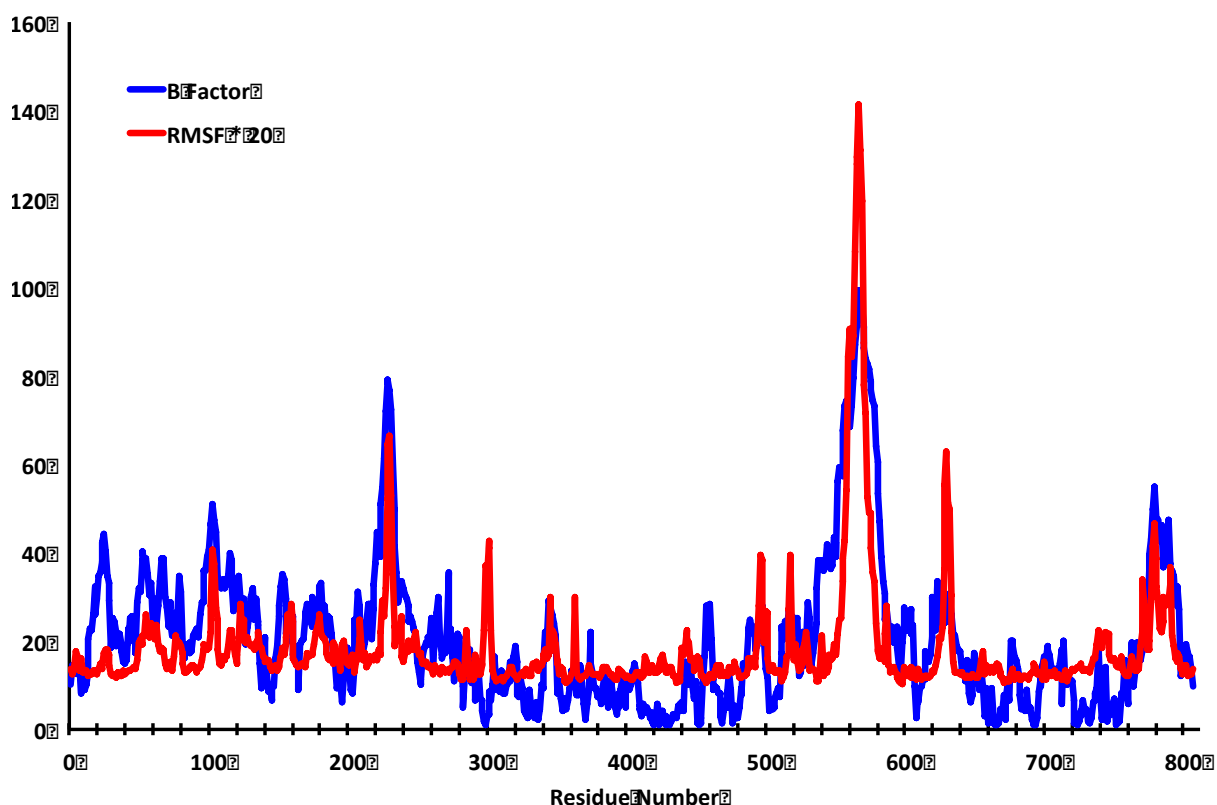
Figure 3-9. *H. pylori* urease X-ray B factors[9] versus scaled RMSF for the dimer reaching wide-open flap state (flap 11). The RMSF has been multiplied by a factor of 20 for easy visual comparison with the urease B factors.

78
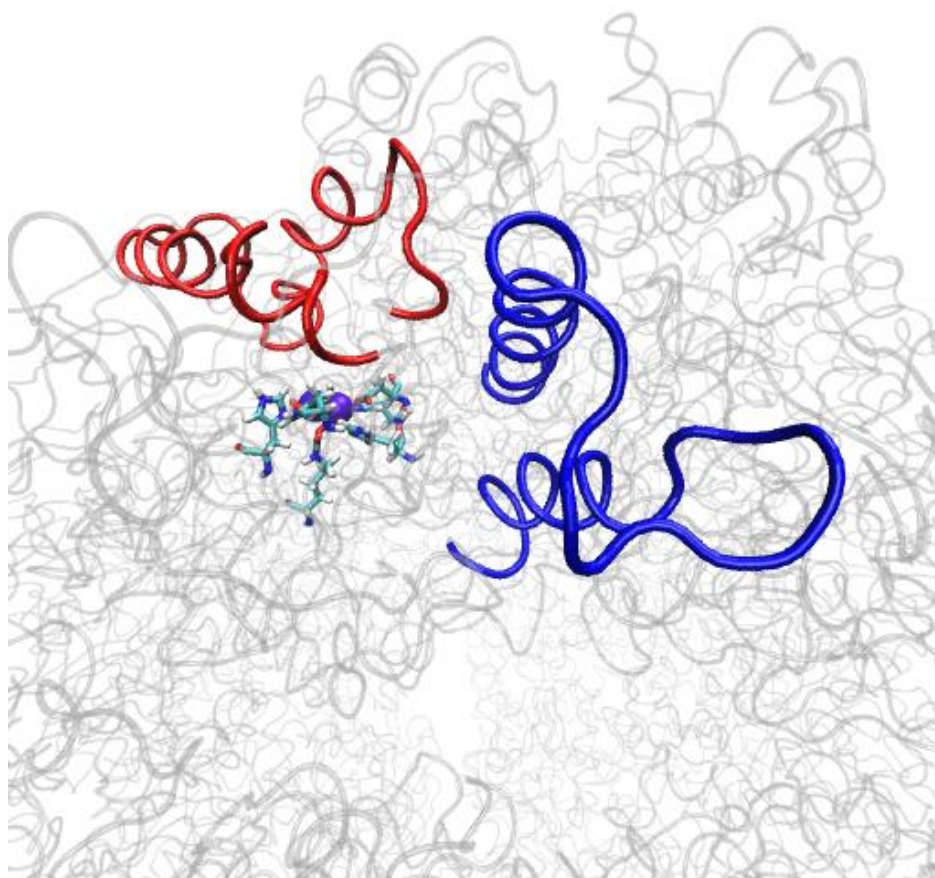
Figure 3-10. Positions of the second active site flap (blue), the flap covering the active site (red) and the active site.

Table 3-1. High RMSF regions and residues with α/β numbering according to Ha et al. and the corresponding sequential numbering from the PDB (1-807).

Residue Range	Amino Acids
β100-β106 (100-106)	ILE-GLU-ALA-ASN-GLY-LYS-LEU
β226-β232 (226-232)	ALA-LYS-SER-ASP-ASP-ASN-TYR
α60-α65 (298-305)	ASN-PRO-SER-LYS-GLU-GLU
α388-α395 (626-633)	GLY-ARG-LEU-LYS-GLU-GLU-LYS-GLY
α538-α545 (776-783)	VAL-ASN-PRO-GLU-THR-TYR-HIS-VAL

Analyzing the RMSF values of the flaps covering the active site (defined as residues α304-α347 (542-585) for our analysis) we observe that in eleven of the twelve flaps, one per αβ dimer, residues α324-α328 (562-566) have the highest RMSF values as expected based on the known flexibility of this region. Comparison of the two alpha helices [residues α304-α322 (542-560) that make up the first helix and residues α330-α347 (568-585) form the second as determined from the MD simulation] reveals that the residues comprising the second alpha helix have much higher RMSF values than those of the first. Additionally, we observe that the loop region residues have the highest RMSF values, particularly residues α326-α329 (564-567). In eleven of the twelve αβ dimers, these loop residues have RMSF values greater than any other residues of the α304-α347 (542-585) domain. As expected, the highest RMSF values in the loop region are observed in the flap (flap 11) that reaches the wide-open state. No flap residues were found to have high RMSF in the αβ dimer that were observed to remain in the closed flap state throughout the MD simulation.

3.3.3 Active Site Access Flap Motion Analysis

Having been previously implicated in allowing urea access to the active site, we focus our attention in the next several sections on an enhanced analysis of the behavior of the active site covering flap over the course of the MD simulation. Distances between points on the flap and opposite points at the access point are used to probe the extent of flap opening, and the distances between the critical HISα322 residue and the catalytically active Ni^{2+} centers of the active site are studied. Finally, free energy maps constructed using these distances are used to elucidate the flap states that can be readily identified in the *H. pylori* urease structure over the duration of the MD simulation.

3.3.4 Critical Distances

The flap motion was assessed by measuring the distances between three points on the flap and three facing points at the distal end of the flap on the protein. The first distance compares the separation between ILEα328 and ALAβ170. The ILEα328 residue resides at the tip of the flap and the ALAβ170 residue is a nearest non-bonding neighboring residue when the flap is closed. The remaining two distances involve the tips of the α-helices and the corresponding nearest neighbors upon flap closure. These distances are between HISα322 and GLYα47, and GLUα330 and ALAβ173, where the HIS and GLU residues are those that reside at the α-helix tips. These are illustrated in Figures (3-11)a-i.

We characterized the states associated with flap motion by recording and analyzing measurements of these three distances for each of the twelve flaps. Figure (3-12) displays representative ILEα328/ALAβ170 separations for the closed, semi-open and wide open flap states. In the closed state, the separation of these residues varies only slightly, oscillating within the 13.5-16 Å range, with slightly more variation observed in the semi-open state where the separation is observed to fluctuate between 12 and 22 Å. In the wide-open state, the distance between these residues increases dramatically to over 38 Å. The separation remains high until the last 50 ns, where it rapidly declines to about 20 Å. This is in accord with the visual inspection of the protein. Flap 11 is the only flap that adopts a clear wide-open state conformation. It is already known in the literature that there is an open state for the flaps that involves the loop region and a partial uncoiling of one of the α-helices.[6] It should be noted that generation of the crystal from the unit cell reveals the existence of protein-protein contacts that preclude the observation of the wide-open flap state in the X-ray structure.
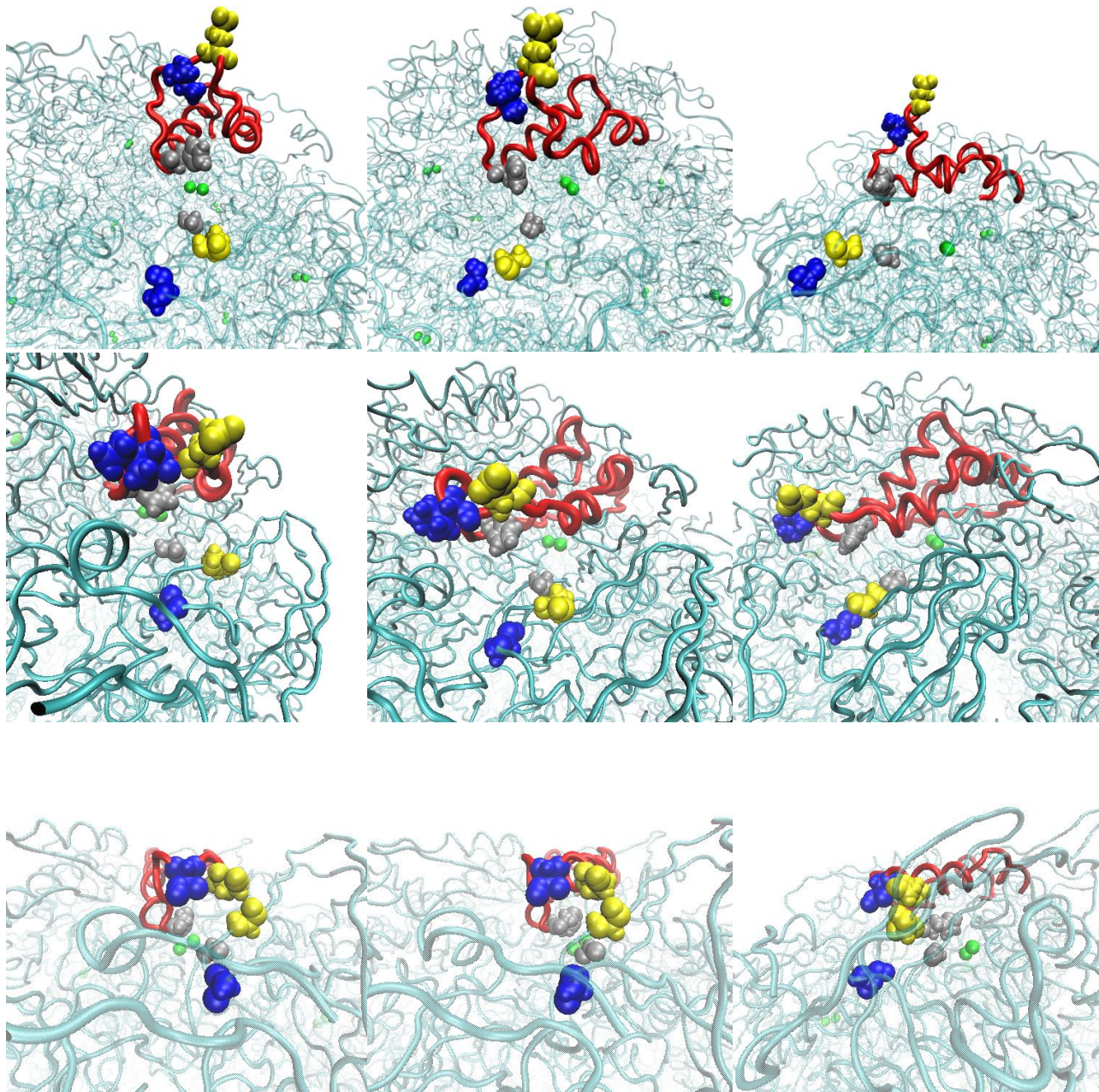
Figure 3-11. Representative conformations illustrating residue separation in the closed (row 3), semi-open (row 2) and wide-open states (row 1). The first column perspective is head-on, the second angled and the third perpendicular with respect to the flap. Blue residues depict the ILEα328/ALAβ170 separation, yellow shows HISα322/GLYα47 and silver shows GLUα330/ALAβ173.
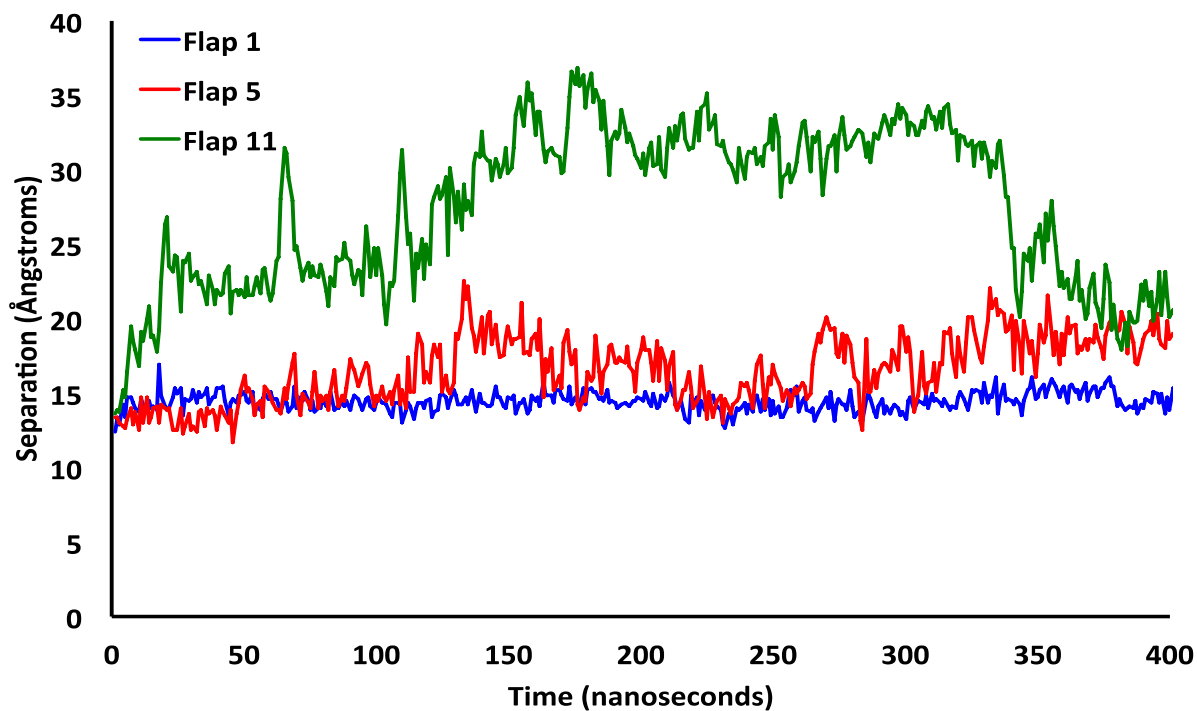
Figure 3-12. Separation between residues ILEα328 and ALAβ170 over time for flaps representing the closed (blue), semi-open (red) and wide-open (green) flap states.

HISα320 is a critical residue proposed to play an integral role in the urea hydrolysis mechanism for *K. aerogenes* and *B. pasteurii* ureases[3, 48b] and displacement of this residue is thought to be essential in reaching the wide-open flap state, allowing urea access to the active site. The corresponding residue in *H. pylori* urease is HISα322. Figure (3-13) shows that in the flap that achieves the wide-open state, the HISα322/GLYα47 separation starts around 8 Å, close to the value observed in the closed flap state, but as the simulation proceeds, the separation of these residues reaches a maximum of 20 Å, with distinct separation from the semi-open state occurring approximately 100ns into the simulation. The trajectory clearly reveals a marked displacement of the histidine in achieving the wide-open state both in comparison to the closed and semi-open states. In the closed state this separation is nearly constant at approximately 7 Å while in the semi-open state the separation is roughly 10 Å.
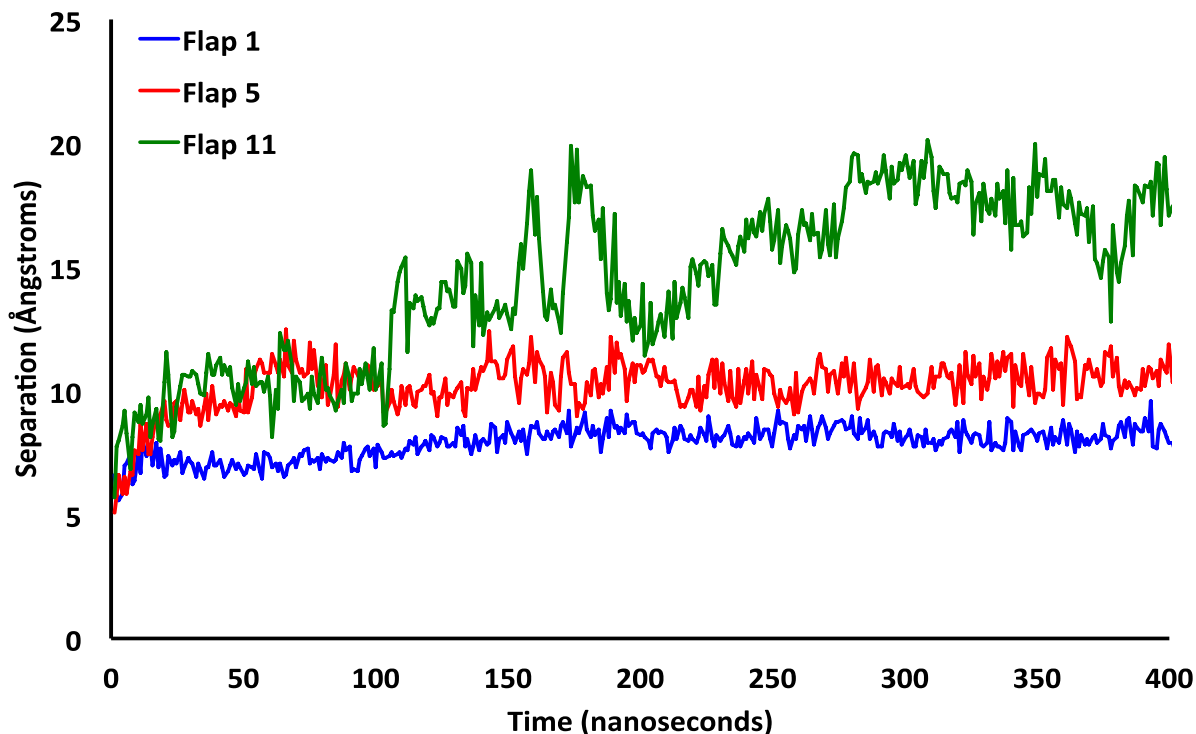
Figure 3-13. Separation between residues HISα332 and GLYα47 over time for flaps representing the closed (blue), semi-open (red) and wide-open (green) flap states.

The separation between GLUα330/ALAβ173 is observed to fluctuate in concert with the ILEα328/ALAβ170 distances, with the former separation reaching a maximum of ~30 Å in the wide-open state (Figure (3-14)). This separation is also observed to narrow over the last quarter of the MD run and begins its decline slightly earlier than the ILEα328/ALAβ170 separation. In the wide-open flap state all the identified residue pairs have separations that vary in phase with one another, which is in contrast to the semi-open flap state that has been reported previously in the literature.[6] The semi-open flap state is observed in the other 10 flaps and in these cases the ILEα328/ALAβ170 and GLUα330/ALAβ173 separations fluctuate in a synchronized manner, while the HISα322/GLYα47 separation fluctuates out of phase with respect to the former two separations. This behavior is expected, since ILEα328 and GLUα330 are only 2 amino acids apart and are more likely to move in a concerted fashion. It is further pointed out that each of the

distances discussed is shorter in the X-ray structure, 12.87 for ILEα328/ALAβ170, 4.91 for

GLUα330/ALAβ173 and 3.85 for HISα322/GLYα47.
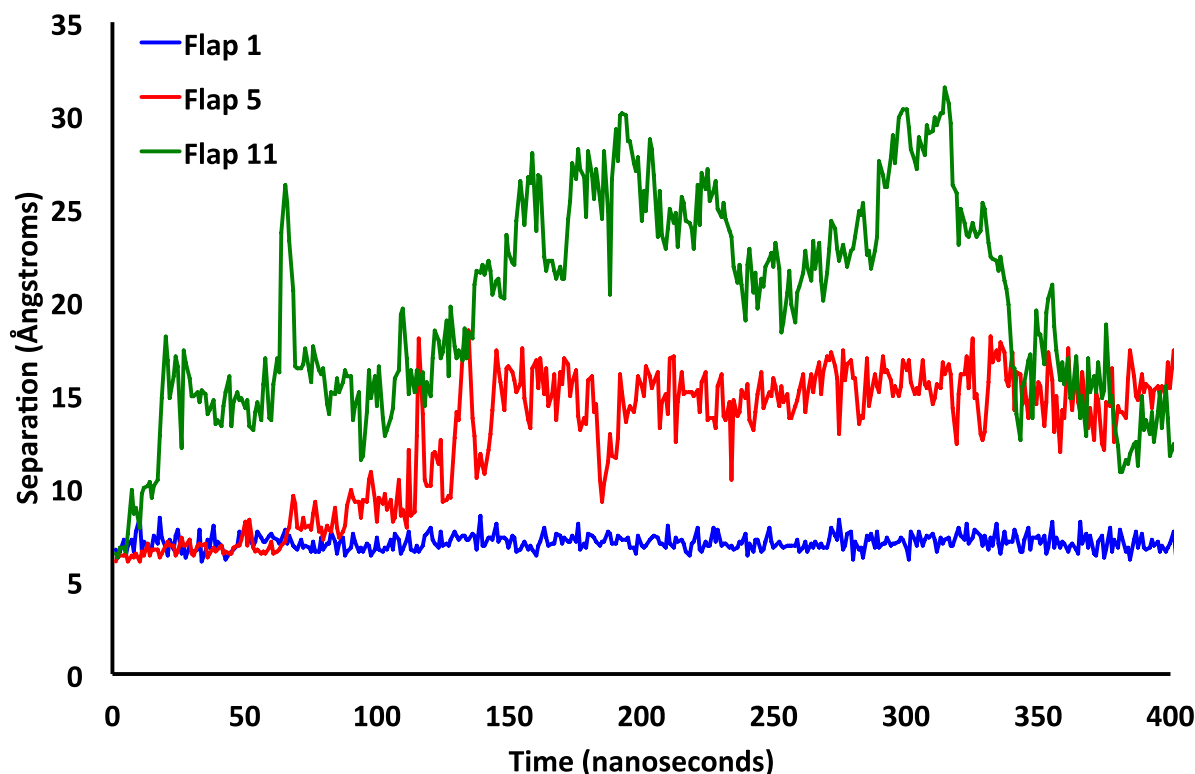
Figure 3-14. Separation between residues GLUα330 and ALAβ173 over time for flaps
representing the closed (blue), semi-open (red) and wide-open (green) flap states.

3.3.5 Ni-HIS Distances

One of the mechanisms proposed in the literature for the hydrolysis of urea by *K.*

aerogenes urease involves HISα320 on the flap interacting with the bridging hydroxide.[3] This

hydrogen bonding interaction with the histidine is thought to be critical for urea hydrolysis and

this residue corresponds to HISα322 in *H. pylori*. The hypothesis is that the wide-open flap state

will have this histidine removed far enough from the nickel centers to permit a molecule of urea

to enter the active site and interact with the pentacoordinate Ni^{2+} ion. In order to analyze this

separation over the course of the MD simulation, the distance between the HISα322 Cα and the

two Ni^{2+} ions for each frame were extracted and plotted (Figure (3-15)-(3-18)). As expected, the

closed flap (flap1) had very small ranges in which the separations between both Ni^{2+} ions and the

histidine were observed to vary between 9 and 11 Å. For the wide-open flap (flap11) the

separations range from 9 to 27 Å, a 9-fold increase in the ranges spanned by the flap1 distances.

In the other 10 cases we observed flaps that are beginning to show wide-open flap state

character. This is typically manifested as the unraveling of one, but not both, of the α-helices that

make up the flap. For example flaps 2, 5 and 8 have ranges from 9 to 17 Å, which clearly reveal

that the HIS residue is far away enough for a urea molecule to enter the active site. As for the

other flaps, they have ranges of separation spanning from 9 to 14 Å.
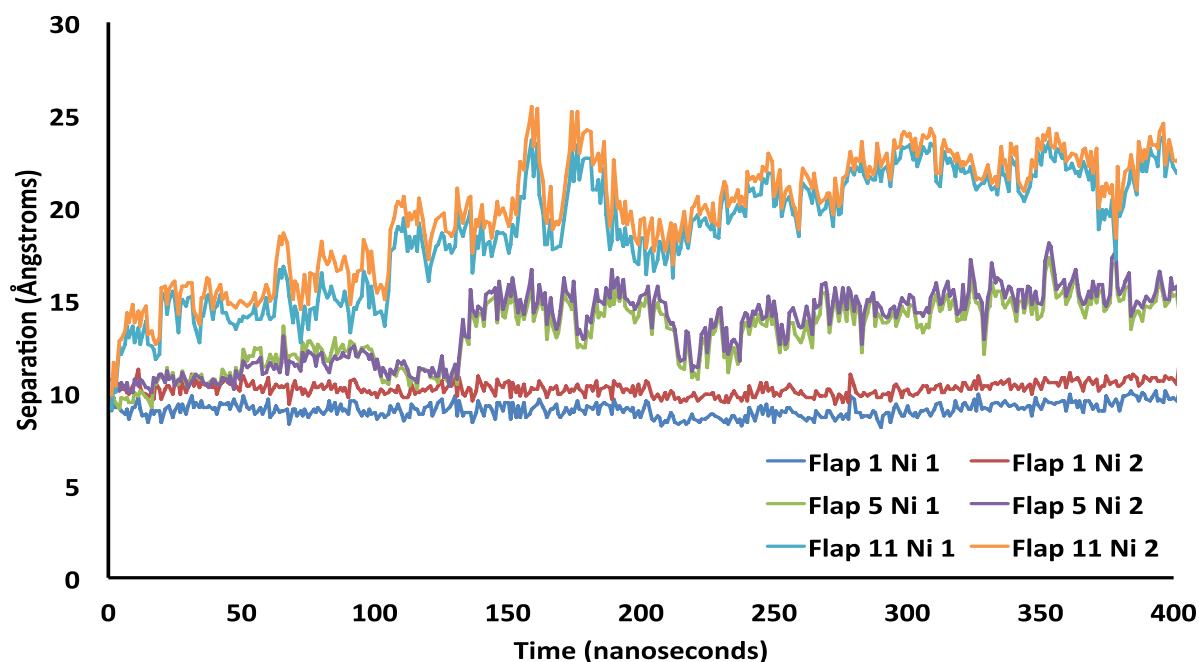
Figure 3-15. Ni^{2+}/HISα322-C_α distance fluctuations for the various flaps.
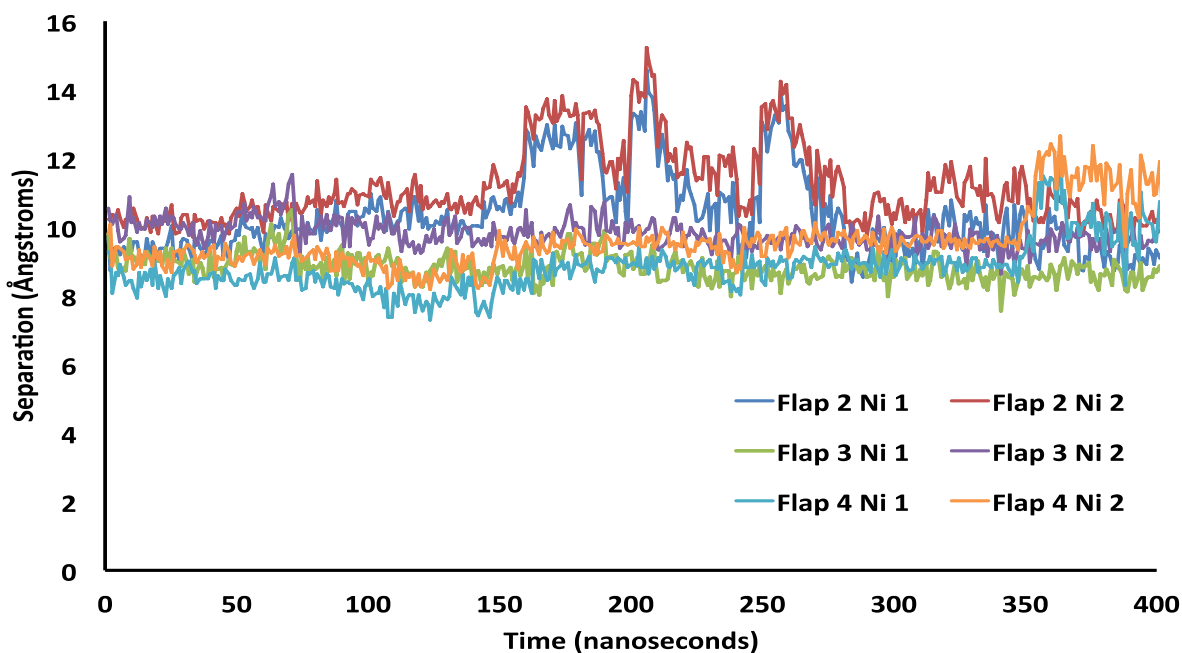
Figure 3-16. Flap two, three and four separations between the HISα322-C$_\alpha$ and each Ni^{2+} ion.
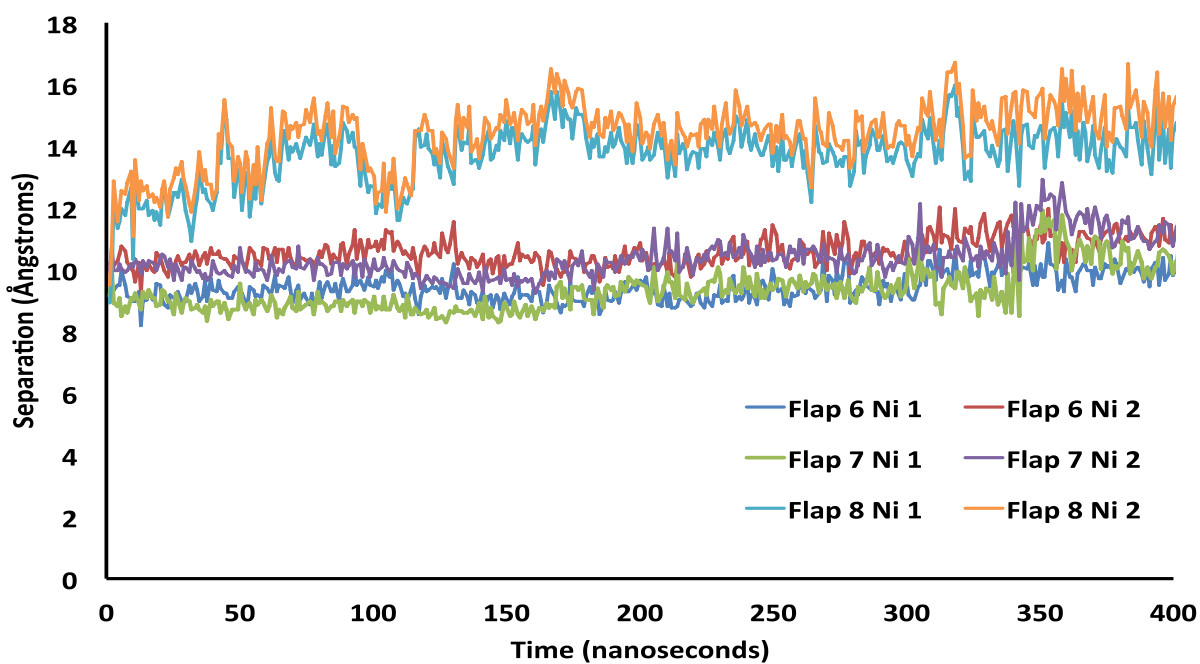
Figure 3-17. Flap six, seven and eight separations between the HISα322-C$_\alpha$ and each Ni^{2+} ion.
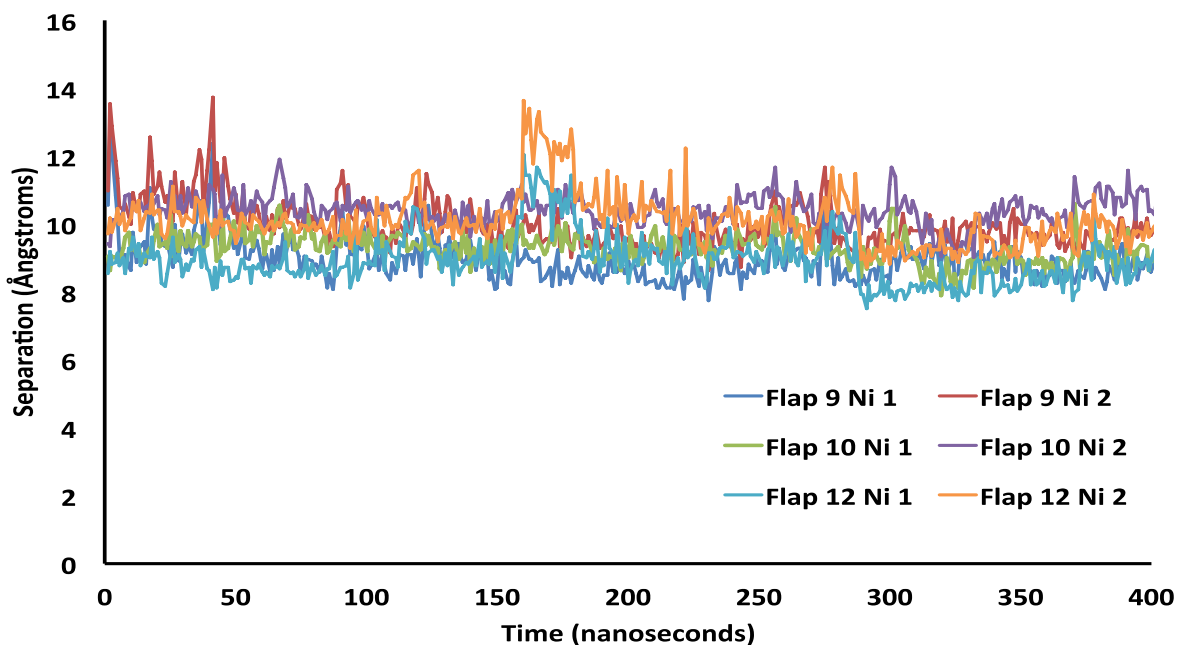
Figure 3-18. Flap nine, ten and twelve separations between the HISα322-C_α and each Ni^{2+} ion.

In order to further probe the importance of this key HIS residue, we extracted and plotted the distances between the HISα322-εN to both Ni^{2+} ions for each flap. All but 2 flaps exhibit regions where the εN/Ni^{2+} distance dramatically increases. Representative separations over the course of the MD simulation are presented in Figure (3-19) for the closed, semi-open and wide-open flap states, with remaining plots presented in Figures (3-20)-(3-24). In the flap that remains closed throughout the entire simulation, the minimum distance between the HIS322-εN and the pentacoordinate Ni^{2+} is 4.5 Å. This pentacoordinate nickel is the closest of the nickel ions to the HISα322 εN in nearly all cases. For the first flap, which was observed to remain closed, the maximum εN/Ni^{2+} separation is 9 Å. As for the other 10 flaps that exhibit a semi-open flap state there are sharp peaks indicative of large increases in the εN/Ni^{2+} separation and the maximum found in the 10 flaps that exhibit a semi-open flap state is 20 Å. As for the wide-open flap state the distance gradually increases reaching a maximum of 31 Å without a clear sharp peak as exhibited by the other flaps. In the open flap sate the minimum distance is generally around 5 Å.

The rapid increases in the $\varepsilon N/Ni^{2+}$ separation indicate that the HISα322 imidazole ring undergoes a 180° rotation at a particular point in time resulting in a significant increase in the observed distance between the atoms. Also of note is a region from 50-100 ns in flap 5 where the $\varepsilon N/Ni^{2+}$ separations invert, and Ni2 is closer to the εN than Ni1. This is the only flap where this inversion is observed and corresponds to an instance where the imidazole ring rotates 90°.
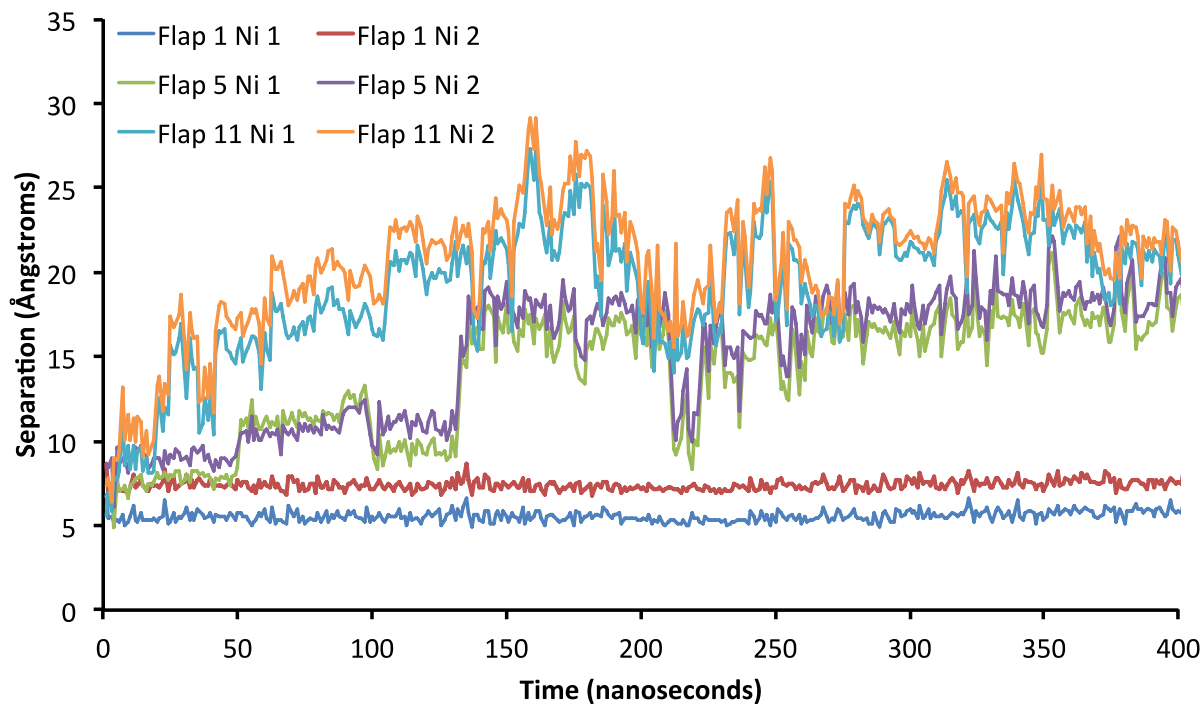
Figure 3-19. Separation between both active site Ni^{2+} ions and the HISα322-εN for the wide-open, closed, and a representative semi-open state.
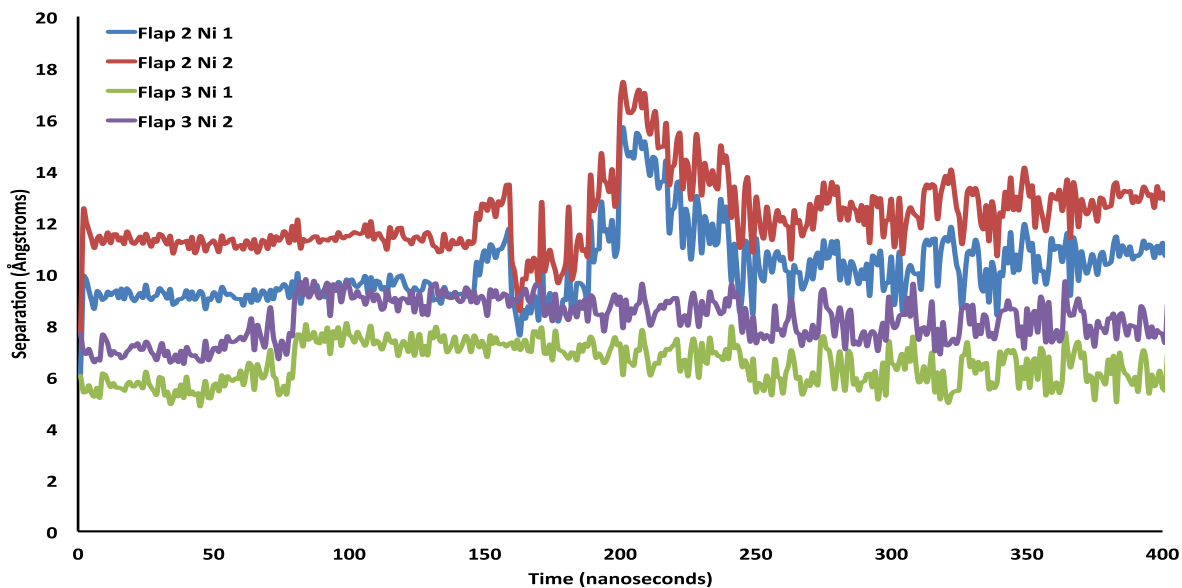
Figure 3-20. Flap two and three separations between the HISα322-N$_\varepsilon$ and each Ni^{2+} ion.

Figure 3-21. Flap four and six separations between the HISα322-N$_\varepsilon$ and each Ni^{2+} ion.

Figure 3-22. Flap seven and eight separations between the HISα322-N$_\varepsilon$ and each Ni^{2+} ion.

Figure 3-23. Flap nine and ten separations between the HISα322-N$_\varepsilon$ and each Ni^{2+} ion.

Figure 3-24. Flap twelve separations between the $HIS\alpha322\text{-}N_\varepsilon$ and each Ni^{2+} ion.

3.3.6 Free Energy Maps

In order to further elucidate the various flap states, relative free energy maps were

constructed based on the methodology described in Toba et al.[59] These maps were produced by

first ascertaining the probability of any two of the residue separations described above coexisting

in any given frame of the 40,000 collected during the simulation. These probabilities were then

utilized to determine the relative free energies of the states via Equation 3-1, where R is the ideal

gas constant, T the temperature and P_i the probability of any two distinct residue distances

coinciding in a randomly selected MD frame. The data was processed using an in-house script.

$$\Delta G_{relative} = -RT \ln P_i \qquad\qquad (3.1)$$

The first three relative free energy maps were produced based on the separations between

the residue pairs $ILE\alpha328/ALA\beta170$, $HIS\alpha322/GLY\alpha47$, and $GLU\alpha330/ALA\beta173$. The

variation observed in the separations between the $ILE\alpha328/ALA\beta170$ residues represents the

92

largest range of distances among the three residue pairs analyzed, ranging from 10 to 37 Å. The

GLUα330/ALAβ173 separation ranges from approximately 6 to 31 Å and the HISα322/GLYα47

distances were found to take values from 5 to 20 Å. From a general observation of the three free

energy maps produced it seems that the loop regions move more freely followed by the α-helix

ending in the Glutamate, then the alpha helix ending with HISα322. If we analyze the free

energy maps produced using the distances described above we observe a wide-open flap state in

each plot. In Figure (3-25) we have the HISα322 to GLYα47 distance versus the ILEα328 to

ALAβ170 distance. In the residue separation ranges of ~34 Å between ILEα328 and ALAβ170,

and ~19 Å between HISα322 and GLYα47 we observe local minima that lie 2 to 2.4 kcal/mol

above the minimum. The closed state occurs at distances of 13 to 15 Å on ILEα328/ALAβ170

and 7 to 8 Å on HISα322/GLYα47. We also observe many interim states, with one intermediate

state that is very evident located at distances of 16 to 20 Å separating ILEα328/ALAβ170 and a

HISα322/GLYα47 separation of roughly 9 to 10 Å.

The second free energy map (Figure (3-26)) was derived from the GLUα330/ALAβ173

and HISα322/GLYα47 separations. The HISα322/GLYα47 separations span a range that is much

smaller than those of GLUα330/ALAβ173. We can observe the wide-open flap state at

separations of 19 Å between HIS and GLY, and 26 to 27 Å between GLUα330 and ALAβ173.

The energy well associated with the wide-open flap state resides approximately 2.4 to 2.8

kcal/mol above the minimum. The minimum observed in this plot resides at distance ranges of 6

Å for GLUα330/ALAβ173 and 5.5 to 8 Å for HISα322/GLYα47. In this plot we observe large

ranges of stable energies while keeping the HISα322/GLYα47 distance constant and the

GLUα330/ALAβ173 distance varies implying more mobility possessed by the GLUα330 residue

and the alpha helix it is located on.
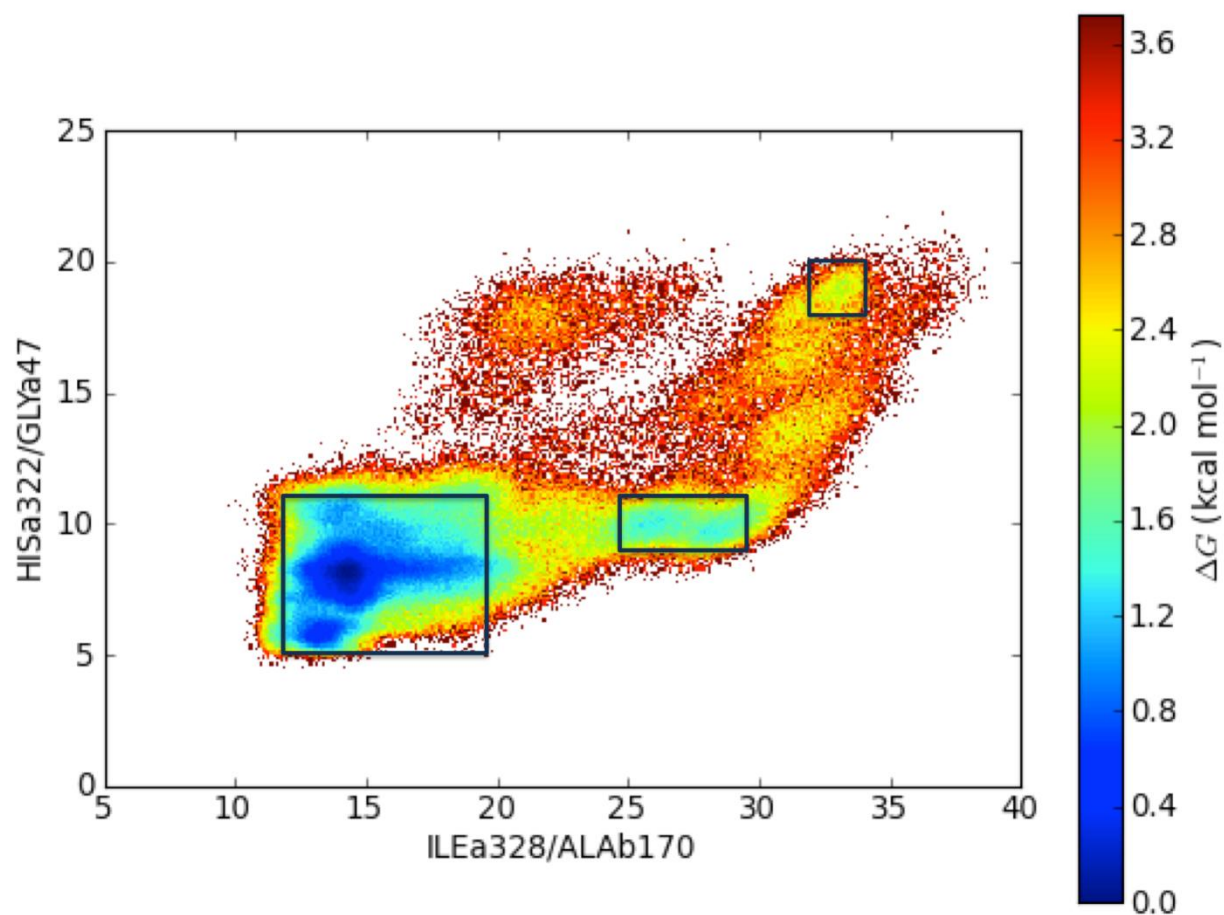
Figure 3-25. Relative free energy diagram constructed based on the separation between ILEα328/ALAβ170 (ILE/ALA) and HISα322/GLYα47 (HIS/GLY). Regions identified as local minima representing the closed, semi-open and wide-open states are boxed.
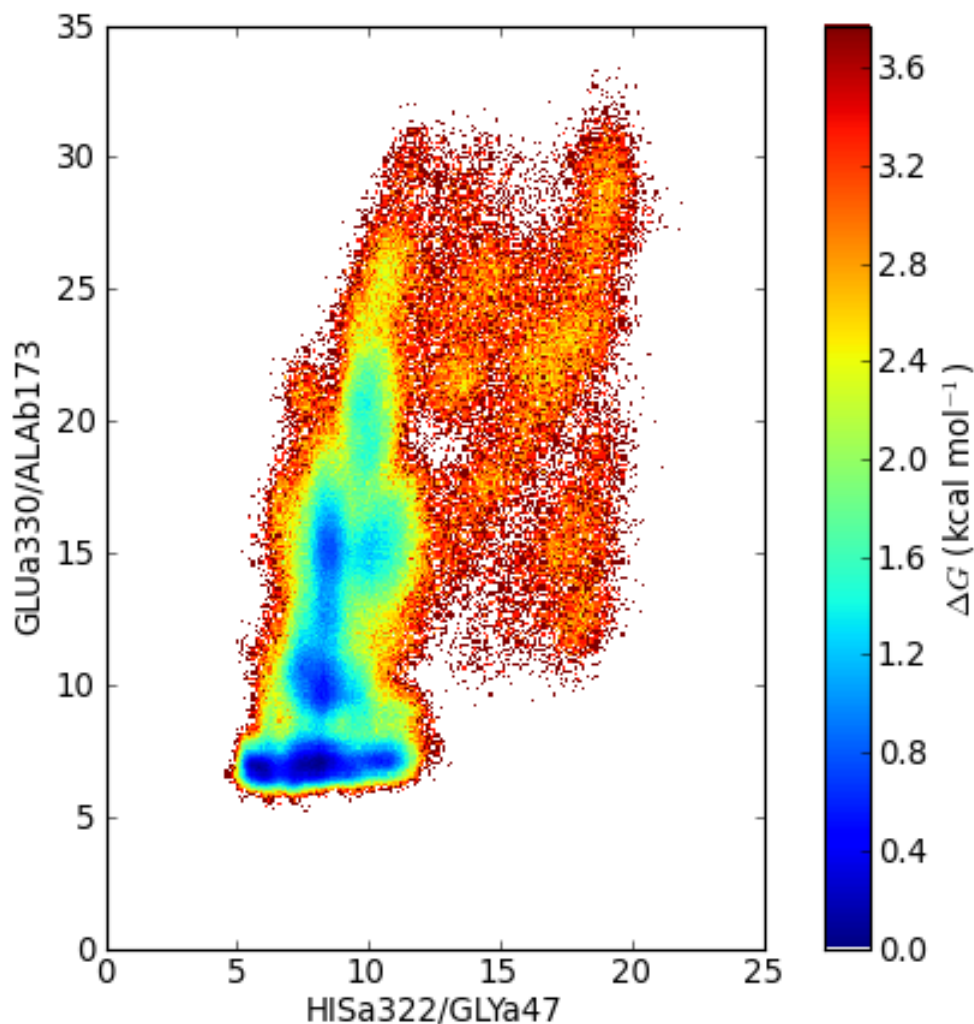
Figure 3-26. Relative free energy diagram constructed based on the separation between HISα322/GLYα47 (HIS/GLY) and GLUα330/ALAβ173 (GLU/ALA).

Figure (3-27) displays the GLUα330/ALAβ173 versus ILEα328/ALAβ170 distance plotted in a free energy map. In this free energy map we can observe at least four distinct regions. One at which both the GLUα330/ALAβ173 (27-32 Å) and ILEα328/ALAβ170 (~30-35 Å) distances are large, another at which both the GLUα330/ALAβ173 (13-15 Å) and ILEα328/ALAβ170 (~6 Å) distances are small, and 2 other regions: one at which the GLUα330/ALAβ173 distance is large (14-16 Å) and the ILEα328/ALAβ170 distance (11-15 Å)
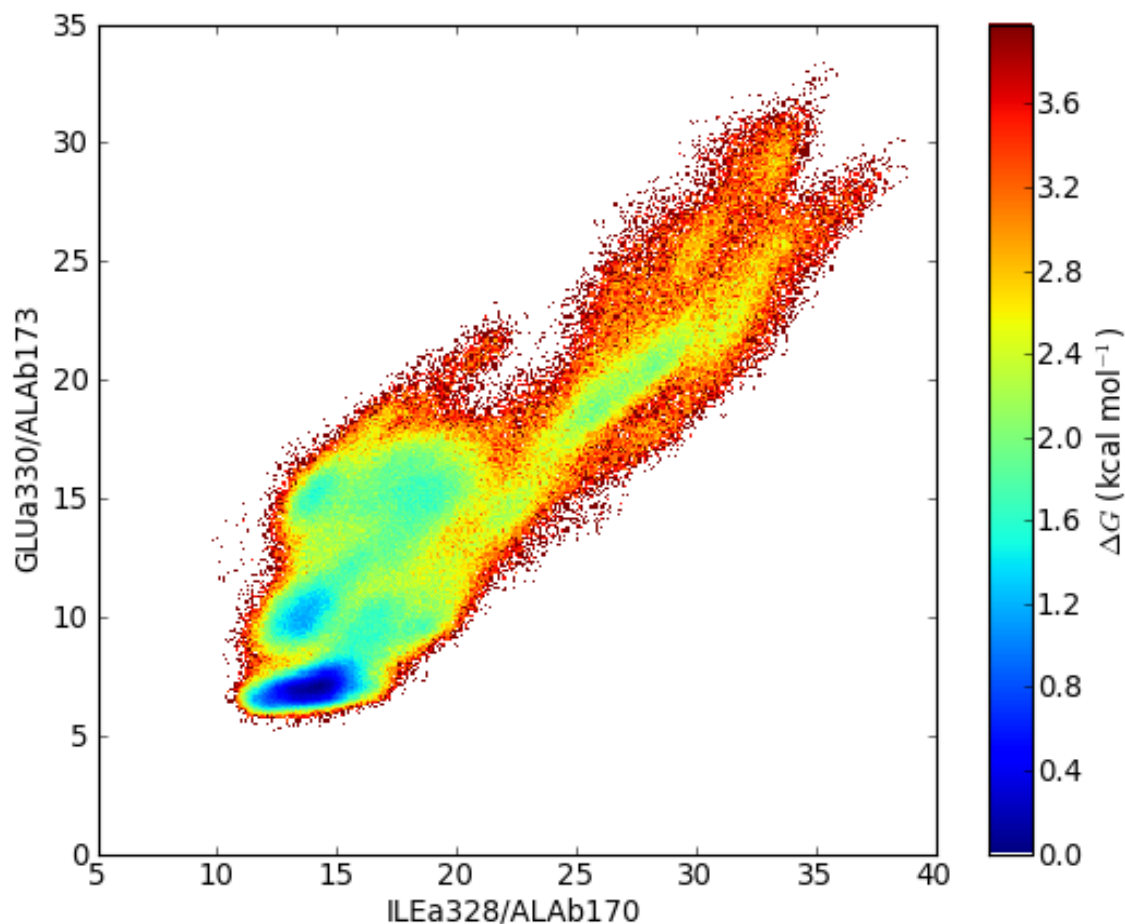
Figure 3-27. Relative free energy diagram constructed based on the separation between ILEα328/ALAβ170 (ILE/ALA) and GLUα330/ALAβ173 (GLU/ALA).

is small, and a second where the ILEα328/ALAβ170 distance is large (~25-27 Å) and the

GLUα330/ALAβ173 distance is intermediate (~17-21 Å).

A final free energy map (Figure (3-28), see SI) was developed based on the separation

between the HISα322-εεN and both nickel ions. In this free energy map we are able to observe

three distinct states, all energetically accessible, that define the closed, semi-open and wide-open

flap states. The different states occur at approximately 6-11, 15-18 and 21-24 Å distances. This

map clearly illustrates the presence of the three flap states.
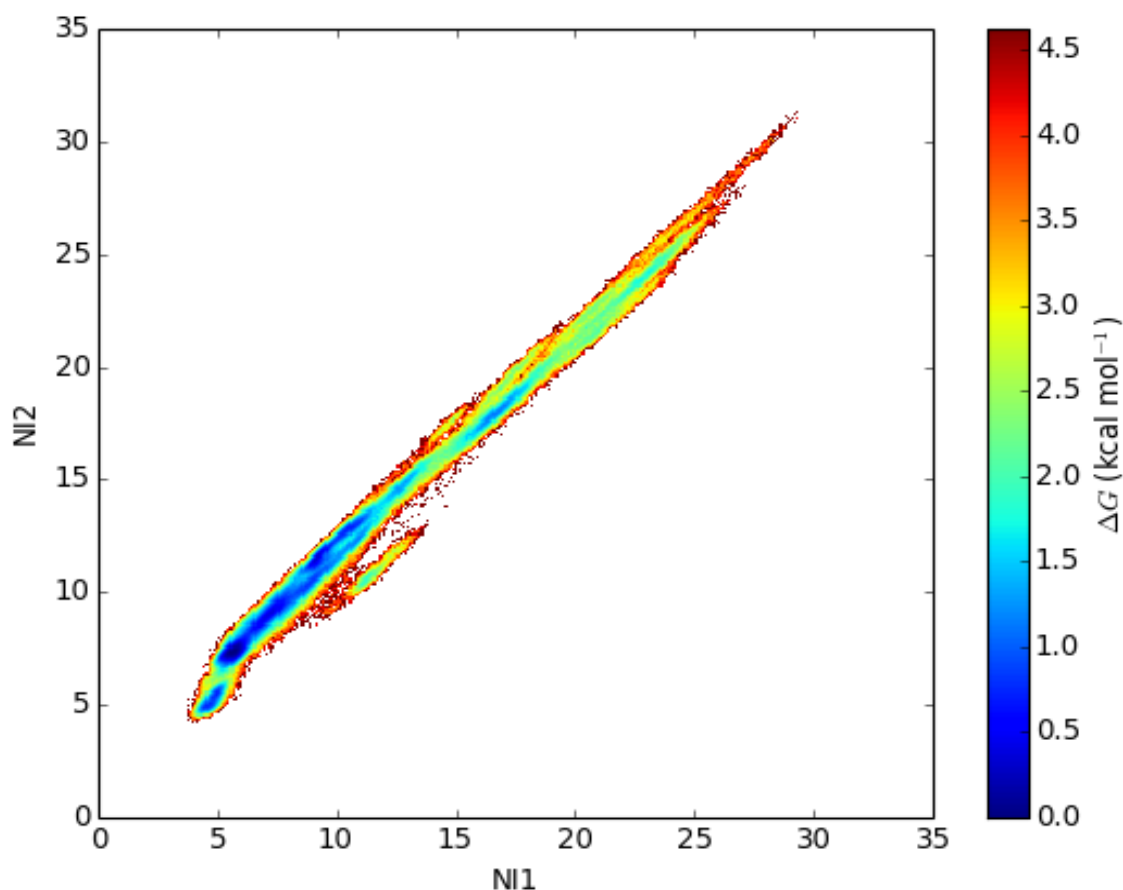
Figure 3-28. Relative free energy diagram HIS322εN/Ni^{2+} constructed based on the separation between (Ni1 HIS and Ni2 HIS).

3.3.7 Additional Structural Features

There exist additional features of the *H. pylori* urease structure that are of interest aside from the aforementioned active site flap. Some of the discussed features below are reaffirmations of, and further elaboration on, the behavior of structural aspects of the enzyme that have been previously described in the crystallographic studies, such as the presence of the interior hollow. Finally, we comment on the distribution of sodium ions during the course of the simulation as ion channeling has been hypothesized to be critical in both the hydrolysis process as well as in the accumulation of ammonia/ammonium in the hollow.

3.3.8 Radius of Gyration

We obtained the radius of gyration for the urease enzyme, as well as the maximum radius (Figure (3-29)). The radius of gyration is the average distance of all atoms from the center of the protein. For *H. pylori* urease the radius of gyration ranged from 60.0 to 61.8 Å over the course of our simulation. At the 230 ns mark the radius leveled off and oscillated between 61.6 and 61.8 Å. The maximum radius, which lists the maximum radius of urease, ranged from 86 to 96 Å and can be partly attributed to the motion of the flaps located on the protein surface.
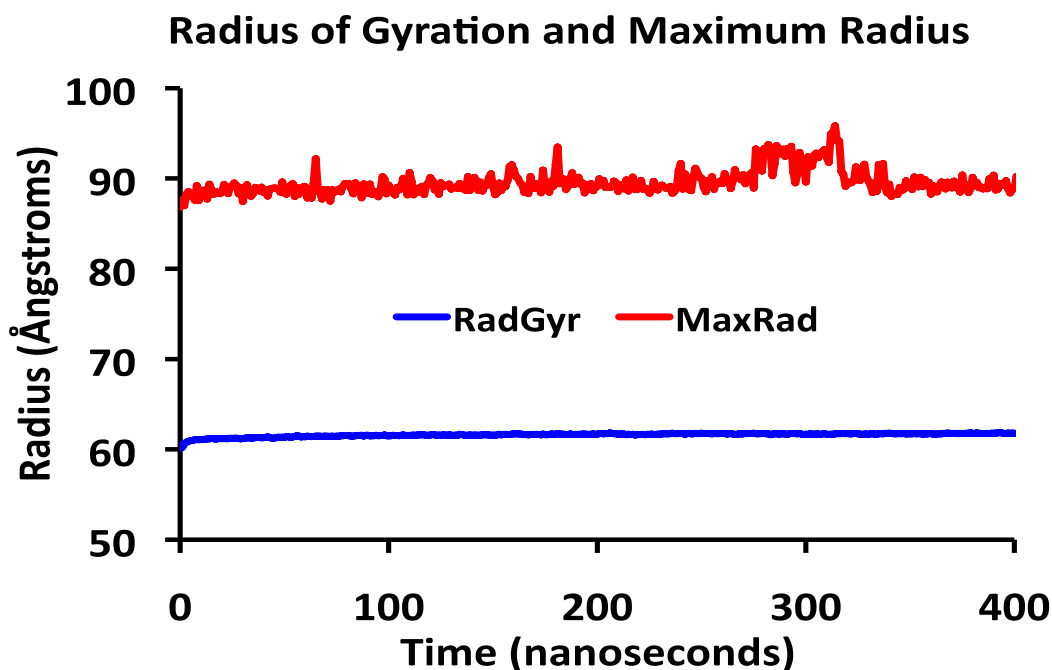
Figure 3-29. Radius of Gyration (blue) and Maximum Radius (red) of *H. pylori* urease.

3.3.9 The Hollow

H. pylori urease has an internal hollow with an estimated volume of ~145,156 Å^3 and it has been hypothesized that this hollow acts as a reservoir of ammonia/ammonium ions.[6] We chose to track the behavior of sodium ions present in our simulation in order to see if this cation accumulates in the hollow as well. First we took the entire trajectory and centered it on the origin. In characterizing the residues outlining the hollow, we observed that the first residues that

we encounter as we move in from the center of the hollow were glutamates - more specifically, the α505 glutamates. These residues do not have high RMSF values based on our "average + one standard deviation" selection criterion, implying a somewhat rigid placement within the hollow. Further analysis was performed by taking the average structure of the enzyme over the entire 40,000 frames produced by the simulation, setting the origin as the protein center of mass, and generating an average PDB structure. While the net charge of all residues within 60 Å of the protein center of mass is -16 (36.4% of total protein charge), the innermost residues of the protein (within 22 Å of the center of mass) are the twelve glutamate α505 residues that place a net -12 charge nearest the hollow center and extension of the radius from the center of mass to 23 Å introduces the first positively charged residue, a single lysine, as well as three additional glutamates for a net charge of -14. This should permit the hollow to accommodate and stabilize a large positive charge. If this is indeed the case, the sodium ion distribution over the course of the MD run should reflect this fact.

The radial distribution function of the Na^+ ions from a dummy atom placed at the origin was calculated over a 95 Å radius with 0.5 Å bins using the average PDB structure. Among the more interesting observations is the fact that of the 44 sodium ions in the system, on average over the entire 400 ns, 30 of these Na^+ ions were found in the hollow within 20.25 Å of the dummy atom located at the protein center of mass as illustrated by the integrated radial distribution function in Figure (3-30). We observe eight distinct peaks within the first 20.25 Å in the radial distribution function at 0.75, 2.25, 6.75, 8.75, 11.25, 15.25, 16.75 and 18.25 Å and there are eight minima within the same span at 1.75, 5.25, 8.25, 10.75, 14.75, 16.25, 17.75 and 20.25 Å and the values of the respective maxima and minima decay over this span until reaching zero at 20.25 Å (Figure (3-31)). Figure (3-32) depicts one hemisphere of the first shell of atoms

encountered from the dummy atom placed at the center of mass, with the glutamate residues

highlighted to illustrate their position as the innermost residues of the hollow. Figure (3-33)

depicts the maxima in the Na^+ ion radial distribution function as shells extending from the center

of mass and terminating at each respective maximum. The highlighted sodium ions (blue

spheres) clearly illustrate the regions of highest probability of locating a sodium ion within the

hollow, and these regions are shown out to 20 Å from the center of mass. This depiction further

supports the conclusion that the hollow is an apt reservoir for positively charged ions.
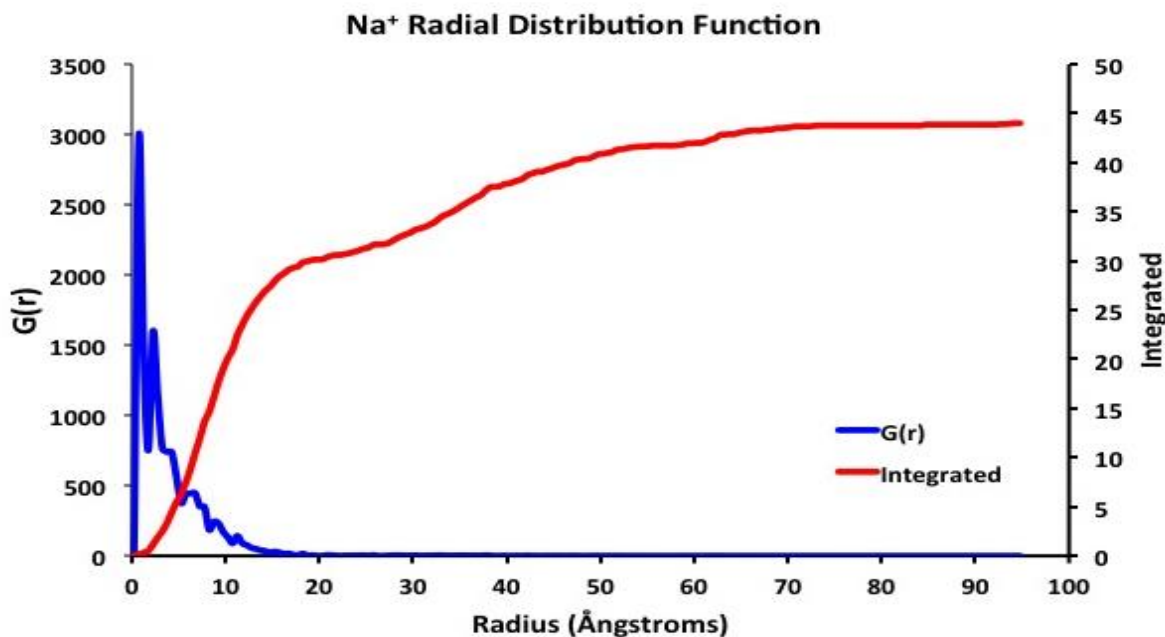
Figure 3-30. Na^+ Radial Distribution Function From Origin.
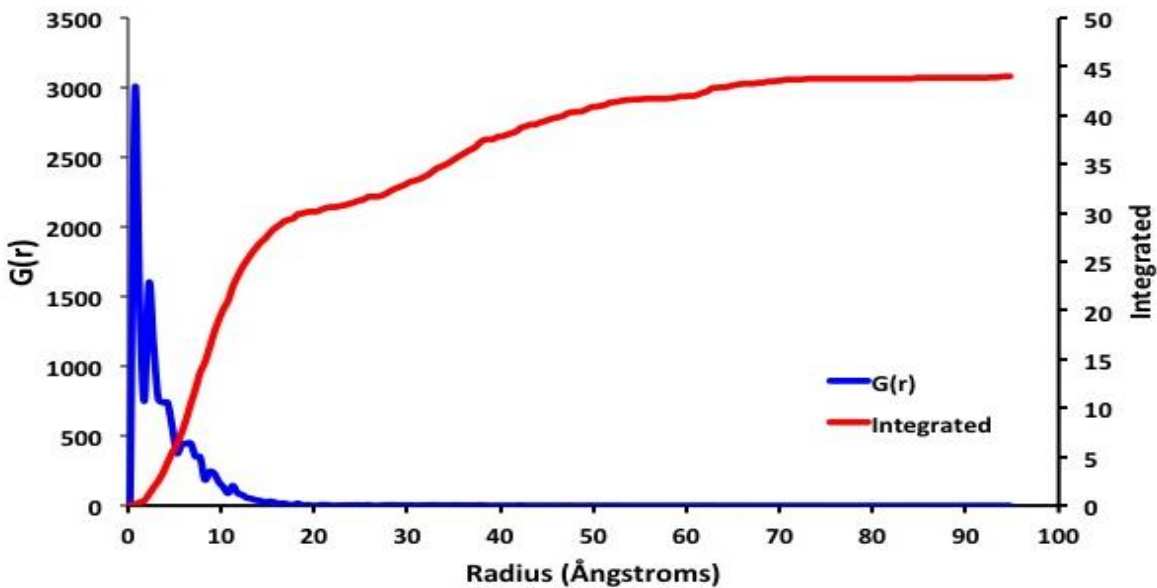
Figure 3-31. Na+ Radial Distribution Function From Origin
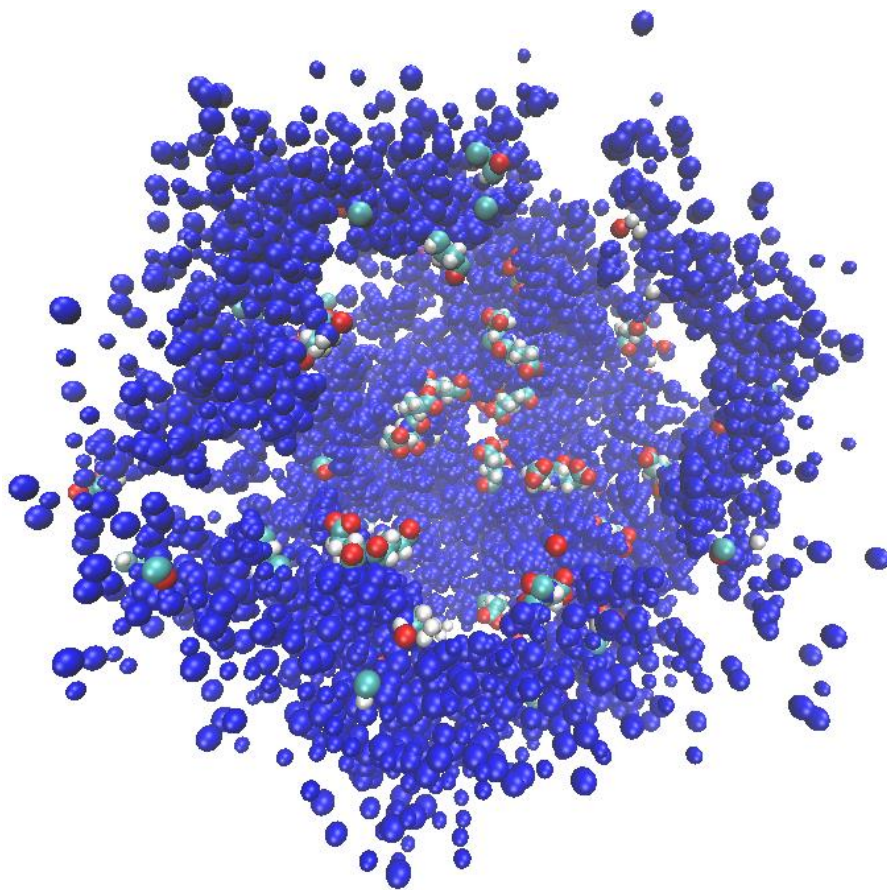
Figure 3-32. First shell of atoms for one hemisphere of HP urease. Glutamate residues are depicted as van der Waals spheres.
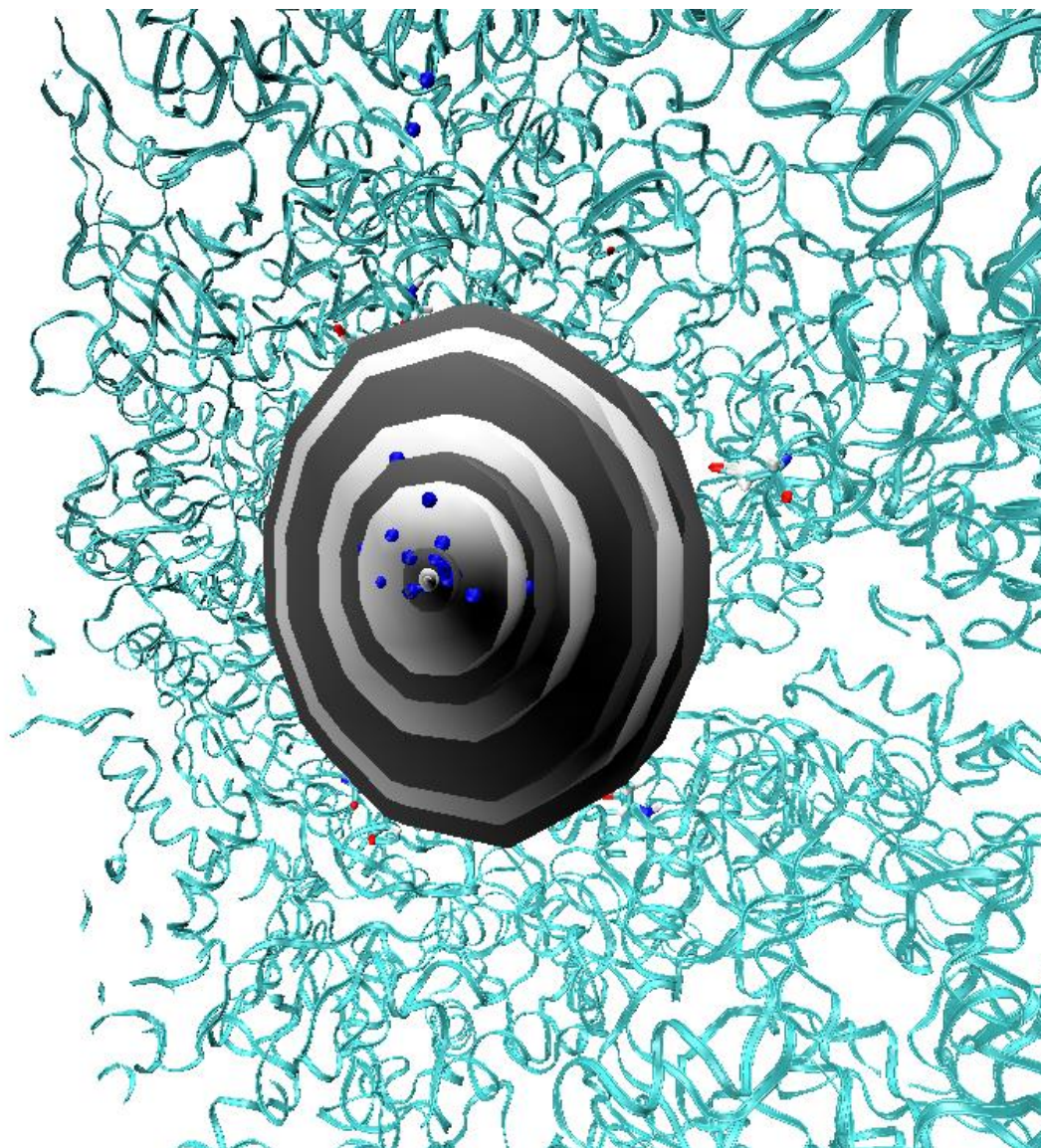
Figure 3-33. Cross-section illustrating the respective maxima in the Na^+ ion radial distribution function up to 20 Å from the protein center of mass using concentric grey and black spheres that terminate at each maximum in the function. Na^+ ions are depicted as blue van der Waals spheres and the residues shown in licorice are residues within 22 Å of the center of mass (all glutamate residues).

3.3.10 Sodium Ion Distribution

The trajectory was further analyzed to determine the sodium ion distribution in our *H. pylori* urease MD simulation. We did this by placing, over each frame of the simulation, a grid with 1 Å^3 cells and counted the number of times a sodium ion appeared in any given cell. This

generated a histogram that provides a representation of the sodium ion distribution over the 40,000 simulation frames. It was shown that most of the sodium ions were found in three major parts of the protein (Figure (3-34)). The first major concentration is observed at the ovoid holes of the two-fold symmetry axes. The second major location where sodium ions were found is by the active site flaps. Flap 1, the closed flap, had the highest concentration of sodium ions out of the 12 flaps. The third major region with a high distribution of sodium ions is the internal hollow and as observed in the radial distribution function, this region had the highest probability of a sodium ion being present. This is consistent with observations made while examining the hollow, as Na^+ ions were observed throughout the simulation in this cavity.
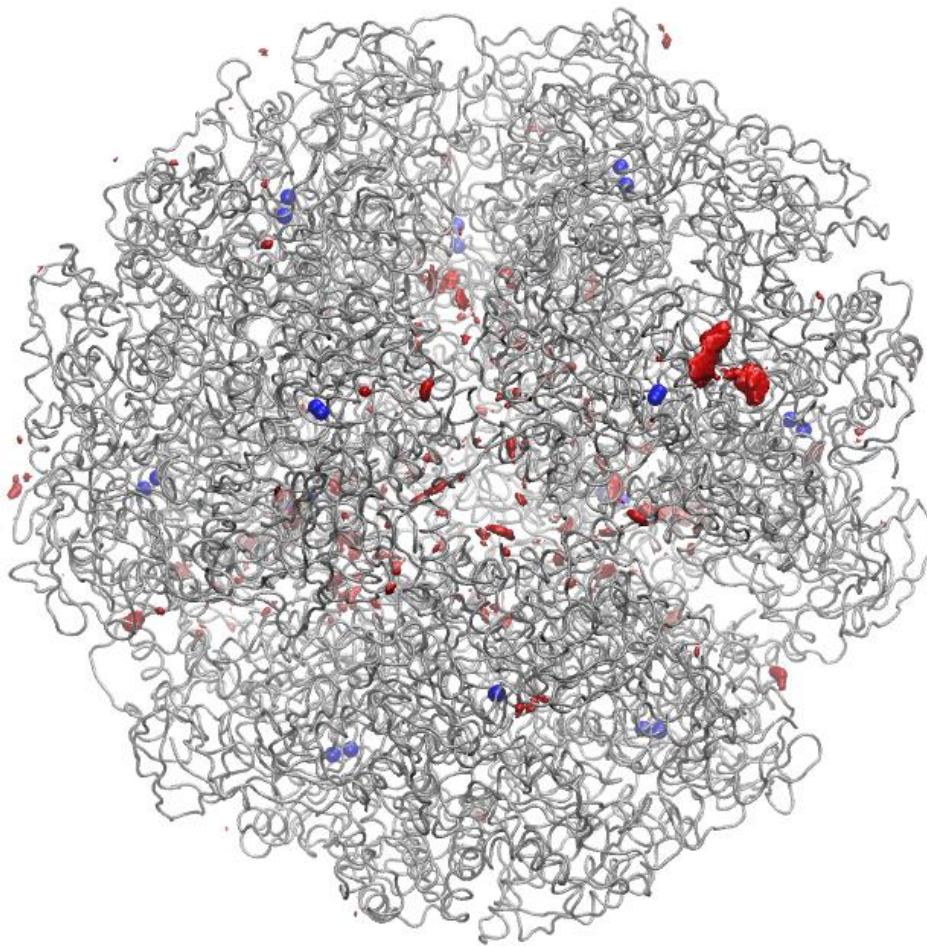
Figure 3-34. Na^+ ion distribution (red).

3.4 Conclusions

We have reported the first MD simulation on the *H. pylori* urease enzyme and made several important observations. First we identified the presence of a third flap state, known as the wide-open flap state, which will provide direction for the development and identification of novel inhibitors of *H. pylori* urease. We have also further elaborated upon a flap located in the posterior of the active site cavity. The simulation reveals that Na^+ ions are capable of channeling through this flap into the hollow, specifically when the active site cover flap is in the closed state. The twelve glutamate α505 residues are, on average, the first residues encountered from the protein center of mass and are the only residues encountered extending out 22 Å from the center of mass of the average PDB structure, providing a net negative twelve charge (negative fourteen when the radius is extended to 23 Å). This accumulation of negative charge (31.8% of total *H. pylori* urease charge at a radius of 23 Å from the average PDB center of mass) helps explain the larger distribution of sodium ions observed in the interior. We also were able to observe that there were distinct regions of the protein in which the protein was very mobile. Furthermore, there were distinct regions with high likelihoods of Na^+ ion presence during the MD simulation including the internal hollow, which clearly exhibited the highest concentration of sodium ions over the course of the simulation. The ovoid holes centered about the two-fold symmetry axes also exhibited high sodium ion concentrations. Additionally, the flap that remained in the closed state had a high Na^+ ion distribution. These findings will aid in understanding the structure and function of the urease enzyme, as well as assist in the strategy implemented in the design of inhibitors of *H. pylori* urease as, particularly the wide-open flap conformation allows for a host of new, larger inhibitor molecules hitherto considered inaccessible to the active site cavity. These larger compounds could be designed in order to

either enter the active site cavity or bind to the flap in order to block opening to the wide open

state from conformational accessibility, or binding and blocking urea entry into the site itself.

CHAPTER 4
MOLECULAR DYNAMICS STUDIES OF HP UREASE IN 10.5M AQUEOUS UREA

4.1 Introduction

Urea is widely known and used as a denaturing agent.[60] Generally, the requirements for protein denaturation by urea are concentrations exceeding 8 M and temperatures over 20 °C. Muthuselvi et al., reported that urea concentrations greater than 10 M were required to denature myoglobin.[61] In order to elucidate the interaction between urea and proteins, numerous MD simulations have been conducted. Independent studies published by Tirado-Rives and Jorgenson[62], as well as Caflisch and Karplus[63] used MD methods to probe the impact of urea as a denaturant on barnase. Both groups reported that observation of denaturing effects on barnase requires a long MD simulation time and an elevated temperature. A recent study by Camilloni et al. reported that urea denatured protein L in MD simulations with temperatures of 400 and 480 K.[64] They also ran a 300 K simulation where the relatively small protein remained in the native conformation after 20 ns. Cai et al. used MD methods in order to probe the effect of urea concentration on protein aggregation.[65] Finally, Moeser and Horinek presented a model for urea denaturation proposing equal side chain and backbone contributions.[66]

Recently, we reported the first full MD simulation on HPU[46], which entailed a 400 ns simulation that revealed a wide-open flap state similar to that observed in an earlier simulation[57] on *Klebsiella aerogenes* urease. Herein we describe the results of our MD simulation on *H. pylori* urease in 10.5 M aqueous urea. To our knowledge this is the first MD study incorporating high concentrations of urea in a simulation on urease. The large concentration of urea was chosen in order to sample the widest possible range of urea/urease interactions without requiring unattainable simulation timescales. We compare these results to a 500 ns trajectory on *H. pylori* urease in aqueous solution (100 ns beyond what was previously reported).[46]

4.2 Methods

4.2.1 Structure Preparation

As in our previous MD efforts on aqueous HP urease, we selected PDB structure 1E9Z[6] as the starting point and utilized PyMOL[55] to perform the reflections required to generate the dodecamer. The H++ online protonation server was used to obtain the correct protonation states of the amino acids throughout the structure at a pH of 7.5[56, 67] following our previous procedure.[46] The bonded model was created with the MCPB component of the tLeap facility (AmberTools version 1.5)[21b] to define the Ni^{2+} coordination site and subsequently the hydrogen atoms and water molecules were added. Neutralization was achieved via the addition of 44 Na^+ ions and the structure solvated by a periodically replicated TIP3P[68] octahedral water box using tLeap. A total of 25,321 water molecules were used to hydrate the structure and 4,788 urea molecules were added to the system. The metal parameters developed for *K. aerogenes* urease by Roberts et al.[53] with the MTK++/MCPB utility of AmberTools and to be consistent with our previous effort[46], were modified utilizing previously developed Lennard-Jones parameters[58] (R* and ε) for Zn^{2+} as parameters for the Ni^{2+} ion.

4.2.2 Minimization and Equilibration

The energy minimization of the structure was performed using the FF99SB[36, 69] force field using a two-stage procedure, where we first minimized the nickel coordination sites by imposing weak harmonic positional restraints of 10 kcal/(mol•$Å^2$) on all atoms outside the Ni coordination sphere. The steepest descent method was first used for $1x10^5$ steps to minimize the active site. Subsequently, the entire structure was relaxed and minimized using steepest descent for $1x10^4$ steps. The initial step length was decreased to $1x10^{-6}$ Å. After minimization the structure was equilibrated in another two-stage process. In the first stage, the temperature was increased from 0 to 300 K over $1x10^6$ steps of MD with a step size of .002 ps (2 fs) in the

canonical (NVT) ensemble, placing a weak harmonic positional restraint on the whole protein. After the system was brought to 300 K, the simulation was run for 10 ps in the isobaric, isothermal (NPT) ensemble after removal of all the harmonic restraints. Temperature control was achieved using Langevin dynamics with a collision frequency, γ, of 2.0 ps^{-1} during the first, and 1.0 ps^{-1} during the second equilibration stage. SHAKE[37] was used to constrain all hydrogen-containing bonds during both equilibration steps. Long-range electrostatic interactions were computed using the particle mesh Ewald method with an 8 Å cutoff.

4.2.3 Molecular Dynamics

A 500 ns MD simulation was run on *Helicobacter pylori* urease in 10.5 M aqueous urea using the PMEMD version of Amber 12[21a, 54] on an M2090 GPU using the FF99SB force field. The production MD run was procured over 500ns in the isobaric, isothermal (NPT) ensemble, which was acquired using 2.0×10^8 steps, with a 0.002 ps time step. The temperature was kept constant at 300 K using Langevin dynamics with a collision frequency of 1.0 ps^{-1} while the pressure was maintained at 1 bar with a pressure relaxation time of 2.0 ps. SHAKE was again used to constrain all hydrogen-containing bonds. For the calculation of the non-bonded interactions, again we made use of the particle mesh Ewald method and a cutoff distance of 8 Å was employed while computing the long-range electrostatic interactions. Frames were saved every 5×10^3 steps (10 ps), providing 5×10^4 frames.

4.2.4 Analysis

All frames were analyzed for the flap state defining residue separations, root mean squared deviation (RMSD), atomic fluctuations (RMSF), and correlation matrices using the PTRAJ utility of AmberTools version 1.5. RMSD, which gives an average deviation of the protein geometry from the reference structure at each point over the simulation time, was obtained for each atom in the entire protein with the first frame of the trajectory as the reference.

We also separated out the residues involved in the motion of the active site-covering flaps and obtained the RMSD for each of these. The RMSF (root mean squared fluctuation) was computed on a per residue basis, using the alpha carbon of each residue as the reference point and providing a time-averaged value. The solvent accessible surface area was determined using NAccess V2.1.1[70] which implements the method of Lee and Richards[71] in the area determination. The secondary structure assignments were made utilizing the DSSP program of Kabsch and Sander.[72] Urea residence times with regions of high RMSF were determined using a code developed to record all urea molecules within 2 Å of any residue of each region of interest based on 501 PDB files (for the equilibrated structure and taken at each ns). We generated histograms for each region and each dimer that show the number of urea that are within 2 Å for over 10 ns. We also plotted for each region of each dimer the total number of urea molecules that reside within 2 Å as a function of time. The chosen distance cutoff is tight in order to account for the great amount of difficulty in determining optimal hydrogen bonding angles for 4,788 urea molecules over 500 PDB files.

Throughout our discussion of the residues, we will identify them both by the residue numbering according to Ha et al. (α and β chains)[6], as well as the residue numbering from the PDB (1-807). For the purposes of this analysis the flap covering the active site was considered to span residues α304-347 (542-585) with the first α-helix composed of residues α304-322 (542-560), the turn composed of α323-329 (561-567) and the second α-helix spans α330-347 (568-585). The α-helices are extended from the helical section described by Ha et al. in order to account for all residues in the flap region that adopt α-helical character at some point during the simulation.

4.3 Results

Over the course of the molecular dynamics simulation of HPU in a 10.5 M aqueous urea solution, the RMSD slowly climbs from 2.0 to 2.5 Å over the first 400 ns and remains at nearly 2.5 Å for the remainder of the simulation. In the active site-covering flaps of dimers 3, 4, 5 and 7 the RMSD spikes in the last 100 to 200 ns. The average radius of gyration levels off at ~300 ns into the simulation and remains at a value between 60.2 and 60.4 Å. The maximum radius, approximately 90 Å, is observed on multiple occasions, at approximately 300, 450 and 500 ns. Over the latter half of the simulation, a radius maximum between 87 and 89 Å is generally maintained.

4.3.1 Protein Denaturation?

The solvent accessible surface area (SASA) of both HPU in water and in aqueous 10.5 M urea was determined for the equilibrated structures and at each 100th ns of the simulation. We did this to determine if the protein was denaturing during our simulation timescale. It has been shown if the protein undergoes denaturation, it experiences an increase in the surface area that is accessible to the solvent.[62-63] In both simulations there was a significant increase in the total SASA (Table 4-1) from the initial stages of the simulation to the 100 ns point. This could be in part due to the protein relaxing from initially being in a crystal conformation. After 100 ns the increase levels off and there is only a modest increase in the SASA over the remainder of the MD simulation. Furthermore, when comparing the simulation incorporating urea to the one featuring only water, the increase is larger in the aqueous simulation indicating the protein is not denaturing (at the timescales employed) due to the 10.5 M concentration of urea found in the solvent. Beyond this analysis it seems likely that urease is less sensitive to urea than other proteins simply because urea is a substrate of this protein and as such urea would locally be expected to be present in concentrations higher than the background.

Table 4-1. Total solvent accessible surface area for the X-ray structure of *H. pylori* urease and evaluated at points along the MD simulations on both the aqueous urease and the urease in aqueous 10.5 M urea.

| | *H. pylori* urease | | *H. pylori* urease + urea | |
	Total SASA (Å^2)	% Increase[a]	Total SASA (Å^2)	% Increase
X-Ray	234281.9	---	234281.9	---
Equilibrated	248967.6	6.3	243999.1	4.1
100ns	281311.5	20.1	273445.3	16.7
200ns	287003.8	22.5	281490.2	20.2
300ns	287774.2	22.8	284452.5	21.4
400ns	291730.1	24.5	287777.2	22.8
500ns	292556.0	24.9	288403.3	23.1

[a]Percent increase in SASA referenced to the X-ray structure

Over the duration of the simulation of the SASA, we find that side chains are much more exposed than the backbone of the enzyme as expected (Table 4-2). A general increase in exposed surface and thereby the percent increase in exposed side chain and main chain occurs over the simulation, with the respective contributions of the side and main chain atoms to the total remaining relatively unchanged. The relative contributions to the total SASA following equilibration shows that the side chain comprised 87.2% of the exposed enzyme surface and the main chain contribution was 12.8%. Following equilibration, the side chain SASA comprised 85.1% of the total, with the remaining 14.9% accounted for by main chain atoms. At the 100ns mark the percentages decline to 84.9% and 15.1% for side chain and main chain, respectively and remains stable at these percentages over the duration of the 500ns simulation, where the respective percentages of exposed side chain and main chain surface area are 84.5% and 15.5%.

These results show that more non-polar atoms than polar are exposed on the surface: 39.4% of the exposed surface is comprised of polar atoms, while 60.6% come from non-polar atoms as evaluated from X-ray structure (Table 4-3). After equilibration, the distribution is 57.0% and 43.0% between non-polar and polar, respectively. The distribution remains relatively unchanged throughout the 500 ns simulation (56.7% and 43.3% at the simulation end).

Table 4-2. Side chain and main chain accessible surface areas for *H. pylori* urease in 10.5 M urea.

	Side chain		Main chain	
	SASA (Å^2)	% Increase[a]	SASA (Å^2)	% Increase
X-Ray	204176.9	---	30105.2	---
Equilibrated	207698.3	1.7	36300.7	20.6
100ns	232283.8	13.8	41162.0	36.7
200ns	238586.8	16.9	42903.6	42.5
300ns	240693.0	17.9	43760.3	45.4
400ns	243522.6	19.3	44255.4	47.0
500ns	243832.3	19.4	44570.5	48.0

[a]Percent increase in SASA referenced to the X-ray structure

Table 4-3. Polar and non-polar accessible surface areas for *H. pylori* urease in 10.5 M urea.

	Non-polar		All polar	
	SASA (Å^2)	% Increase[a]	SASA (Å^2)	% Increase
X-Ray	141933.6	---	92348.2	---
Equilibrated	141806.6	-0.1	102192.4	10.7
100ns	156404.5	10.2	117040.7	26.7
200ns	161719.3	13.9	119771.5	29.7
300ns	163253.5	15.0	121200.0	31.2
400ns	165316.7	16.5	122460.5	32.6
500ns	165990.5	16.9	122411.8	32.6

[a]Percent increase in SASA referenced to the X-ray structure

4.3.2 Root Mean Square Fluctuation

Over the course of the simulation, we observed high RMSF regions that differ slightly from those observed in the aqueous HP urease simulation, and the average RMSF values for the urea run were decreased in comparison to those for the original aqueous HP urease simulation (Figure (4-1)). One check for the denaturation of the protein was observation of the average RMSF at 100 ns intervals (Figures (4-2)-(4-6)). While there is a modest increase in RMSF from 100 to 200 ns, there is no increase in average RMSF for the remainder of the simulation.
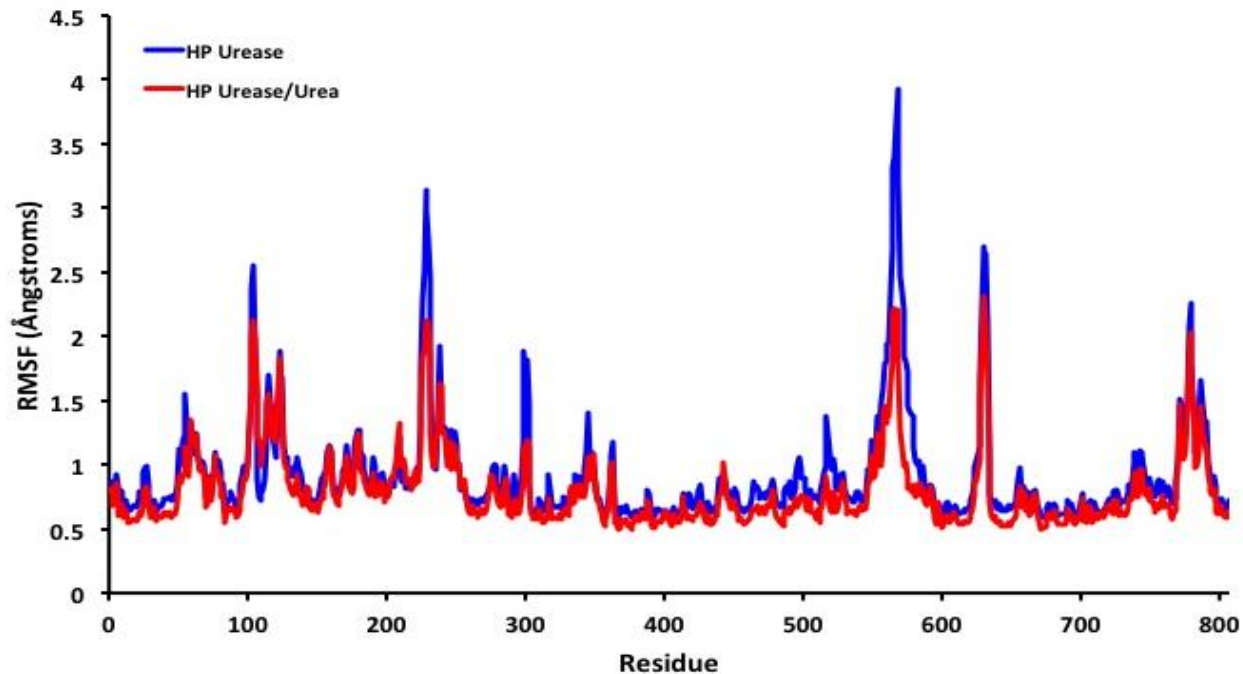
Figure 4-1. Overlay of the average RMSF plots for HP urease and HP urease in 10.5 M aqueous urea 500 ns simulations.
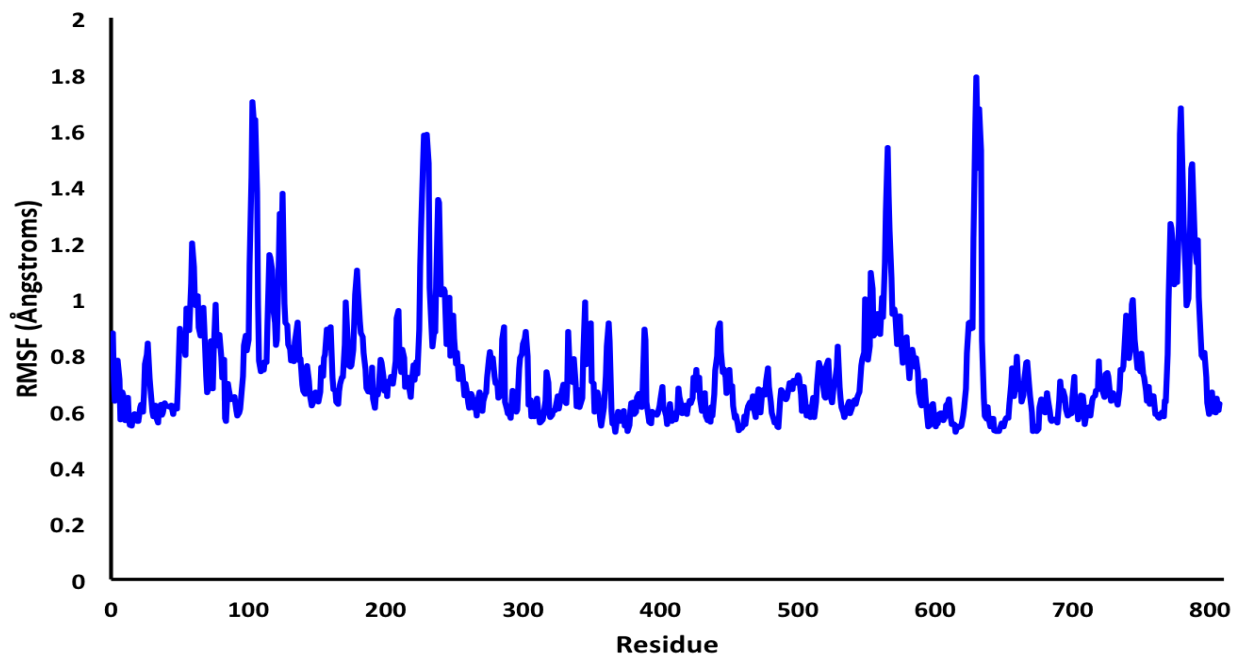
Figure 4-2. Average RMSF per residue of dimers 1-12 (100 ns).
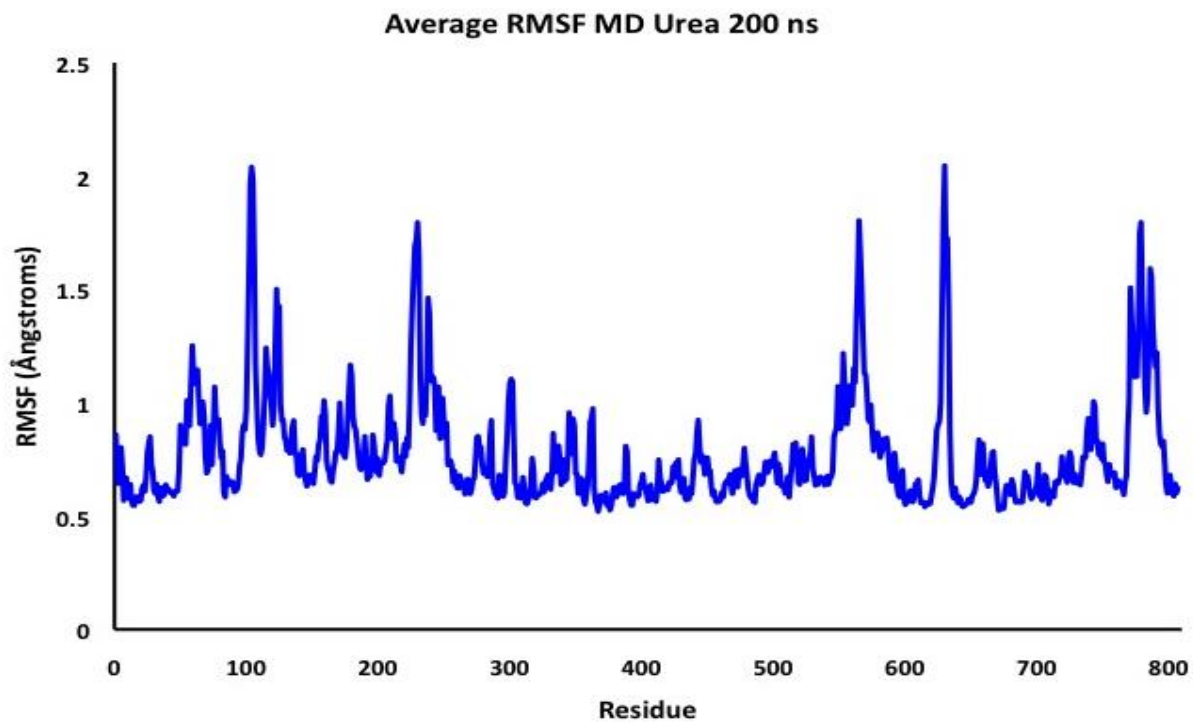
Figure 4-3. Average RMSF per residue of dimers 1-12 (200 ns).
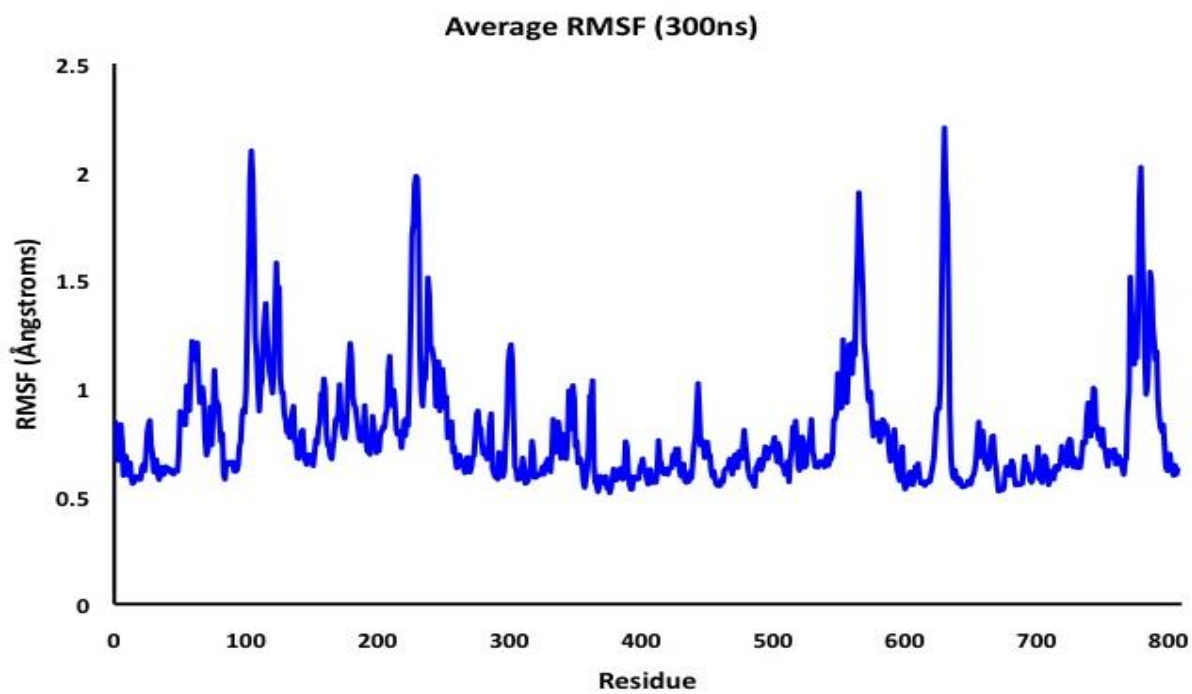
Figure 4-4. Average RMSF per residue of dimers 1-12 (300 ns).
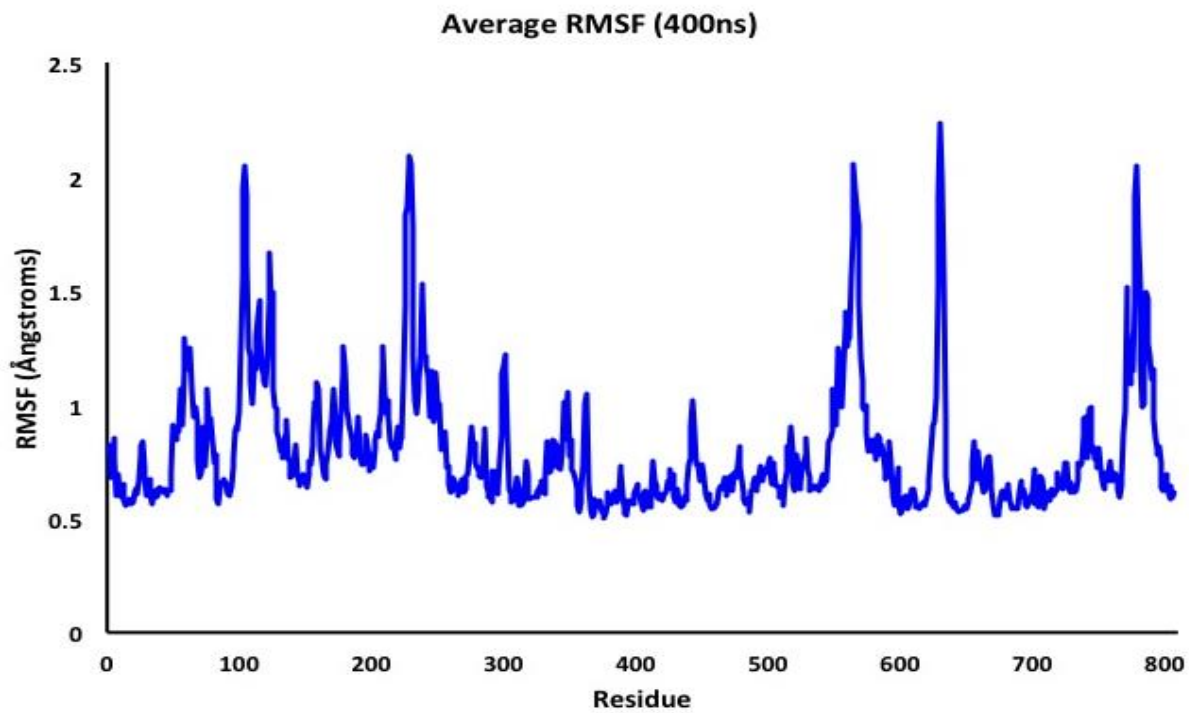
Figure 4-5. Average RMSF per residue of dimers 1-12 (400 ns).
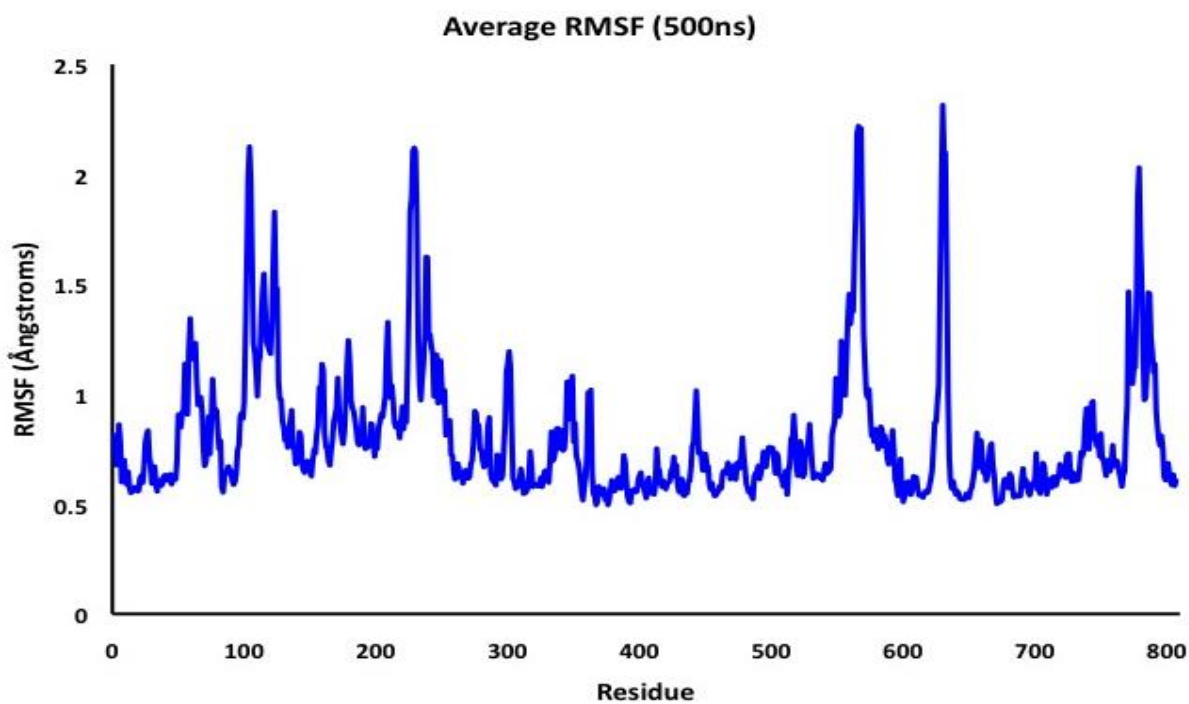
Figure 4-6. Average RMSF per residue of dimers 1-12 (500 ns).

All of the high RMSF regions are found on the surface of the protein and reside mostly on the beta chain, which envelops the protein. These high RMSF regions form horseshoes around the protein and lead to the opening of the active site, of which the loop region is a high RMSF region. These regions contain multiple polar residues and bonds making them good binding candidates for the highly polar urea molecule.

We compared RMSF values of the *H. pylori* run in a 10.5 M aqueous urea solution to our earlier HP urease run in water, and observed that the baseline RMSF value is about the same for both. The high RMSF values are larger for the original run than the run with urea. In the original aqueous run the region that had the highest RMSF value was the loop region, while the region with the highest RMSF value on the urea run was α387-395 (625 to 633), which is the "exhaust flap," whereas in the aqueous simulation, the loop of the active site covering flap possessed the highest RMSF. The exhaust flap of the urease run has slightly higher RMSF than both the loop of the active site covering flap (α326-330 [564-568]) and the ancillary flap (α533-553 [771-791]). In the original run there is a region of high RMSF that resides on the interior of the protein, lining the hollow region, composed of residues α60-65 (298 to 303) that is not observed in the HP urease run in aqueous urea. In conclusion, the urea run has lower RMSF values indicating the protein moves but not as much as in the original run and that it is not being denatured by the urea.[73]

These regions of high RMSF, particularly those of the β-chain, form a horseshoe shaped region on the adjacent dimer that surrounds the loop of the active site covering flap (Figure (4-7)). We find that the horseshoe region attracts molecules of urea (*vide infra*) and contains three points that are equidistant from the loop of the active site covering flap. We hypothesize that this could allow for the shuttling of urea molecules towards the active site for subsequent entry and

possibly facilitate urea transport to the active site region. Urea is a highly proficient enzyme ($k_{cat}=1650\ s^{-1}$), which requires a high flux of substrate molecules reaching the active site.[74] These high RMSF regions are on the surface and we have done extensive studies of the number of urea molecules that interact with these areas throughout the entire simulation (*vide infra*). These horseshoes are symmetrically oriented about the dodecameric structure of *H. pylori* urease as shown in Figure (4-8) of the supporting information.
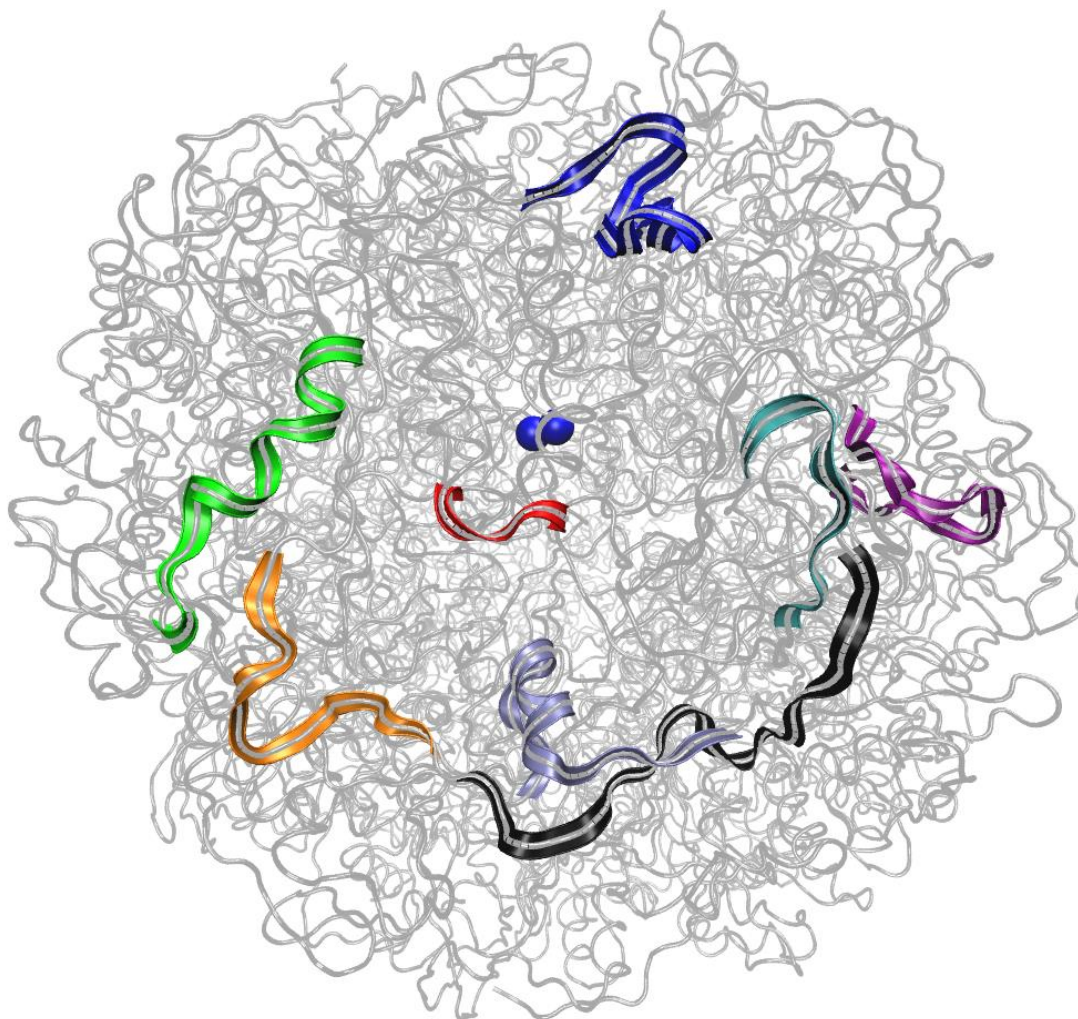
Figure 4-7. Horseshoe of high RMSF regions surrounding the loop (residues 564 to 568) of the active site-covering flap (D4 and D5 refer to the dimer on which the residues are located): 564 to 568 RED (D4); 771-791 BLUE (D4); 55-68 GREEN (D5); 100-111 ORANGE (D5); 113-130 BLACK (D5); 177-184 CYAN (D5) 222-238 PURPLE (D5); 239-255 ICEBLUE (D5).
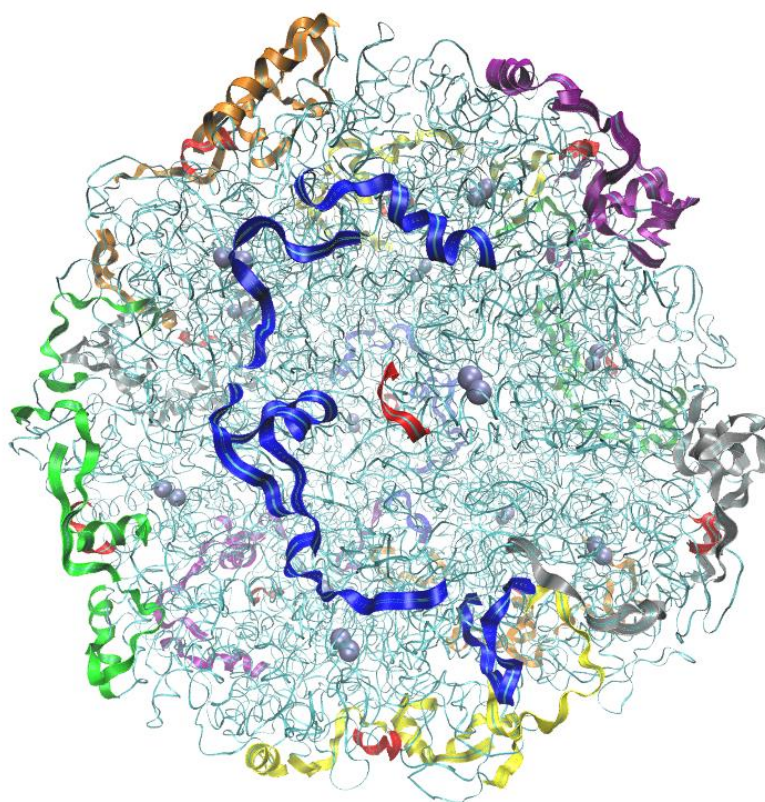
Figure 4-8. All 12 horseshoe shaped regions of high RMSF that individually surround the loops of the active site covering flap.

4.3.3 Active Site Flap Characterization

We previously used a number of characterization techniques to provide evidence for the existence of three well-defined states for the active site flap. These studies focused on the three residue pairs relevant to the characterization of the active site flap: ILEα328/ALAβ170 (ILE566/ALA170), HISα322/GLYα47 (HIS560/GLY285), and GLUα330/ALAβ173 (GLU568/ALA173) (GLU568/ALA173). We first plotted critical separations between these residue pairs to probe the opening of the active site flap, revealing several flaps through the simulation – specifically flap 4 – that exhibit three unique separation states. Subsequent free energy maps revealed clear local minima that were assigned to the unique flap states in the case

of ILEα328/ALAβ170 (ILE566/ALA170). The other two pairs, however, featured no distinct minima at large separations. We suspected that this occurred due to the fact that the corresponding residues are contained within alpha helices that unravel as the flap shifts states. Further study into the changes in secondary structure associated with each residue of interest resulted in observations consistent with this possibility. An expanded discussion of the active site flap is found at the conclusion of the supporting information.

4.3.4 Urea Distribution

To examine the distribution of urea molecules on the surface of the protein we created a 1 $Å^3$ grid around urease and at each point we summed the number of urea molecules present over the entire 500 ns simulation. Overall, urea molecules are found in high density around regions of high RMSF. The second point is that urea is found in higher density at the large ovoid holes to the hollow but at lower density at the holes lined by the α-helices. Urea seems to be found around both hydrophilic and hydrophobic residues. Urea is not found in the hollow at the center of the protein and is not observed until we reach a 19 Å radius from the center of mass of the protein. Urea mostly interacts with loop regions on the protein as opposed to α-helices.

4.3.5 First Solvation Shell

The composition of the first solvation shell was determined for the aqueous HP urease and HP urease plus aqueous urea simulations at 100 ns intervals along the respective trajectories, in addition to the equilibrated structures. The bulk solvent in our urea simulation includes 25,321 water molecules and 4,788 urea molecules, a water to urea ratio of 5.29:1. We chose to select any water or urea molecules within 2 Å of the protein surface as residing within the first solvation shell (Table (4-4)). Following equilibration and at points taken every 100 ns along the trajectory, the water to urea ratio at the protein surface is significantly less than the ratio present in bulk solvent. The highest ratio of 4.55 water molecules per urea molecule was observed

following equilibration and the ratio settled at nearly 4:1 for the remainder of the simulation. The total number of molecules present in the first solvation shell for the aqueous urea simulation was considerably less than the total in the aqueous HPU run, an observation easily accounted for by the relative sizes of the urea and water molecules. While the total number of urea molecules increased post-equilibration, the ratio of water to urea remained much lower than that in the bulk solvent due to a corresponding increase in the number of water molecules. This can be attributed to the "structure-making" property of urea, which possesses eight potential hydrogen bonding sites and can order multiple water molecules while remaining in contact with the protein.

Table 4-4. Composition of the first solvation shell (water or urea molecules within 2 Å of the protein surface) for the aqueous HPU and HPU + aqueous urea simulations.

Entry	HPU Water	HPU + Urea Water	HPU + Urea Urea	HPU + Urea Total	HPU + Urea Water:Urea
Equilibrated	6022	4359	957	5316	4.55
100 ns	7906	5461	1351	6812	4.04
200 ns	8114	5630	1319	6949	4.27
300 ns	8242	5675	1398	7073	4.06
400 ns	8252	5754	1469	7223	3.92
500 ns	8327	5710	1453	7163	3.93

4.3.6 Urea in Active Site Cavities

We further analyzed the trajectory to determine whether urea molecules were approaching any of the twelve active sites. In ten of the twelve dimers, urea molecules are observed within 10 Å of the active site. In the active site of dimer four, we see multiple urea molecules entering beginning at 346 ns, with urea coming within 6.7 Å of the pentacoordinate Ni^{2+} (Figure (4-9)). This accumulation of urea in the active site cavity is in stark contrast to the low amount observed in the closed and semi-open states (Supporting Information, Figures (4-10) and (4-11)). In the closed state at 100 ns, no urea is observed in the active site cavity (Figure (4-10)) as expected. At 250 ns, the active site covering flap has been in the semi-open state for nearly 150 ns, and still only two urea molecules have begun to migrate into the cavity (Figure (4-

11)). Only after the wide-open flap state has been achieved does the rapid migration of urea into the cavity commence for dimer 4. The observation of the large amount of urea in this particular dimer is noteworthy, as this is one of the active site-covering flaps that opens the widest, and opens at an earlier point in the simulation than the other 11. We see the largest influx of urea molecules in dimer 4 and dimer 5, two flaps that reach the wide-open state.
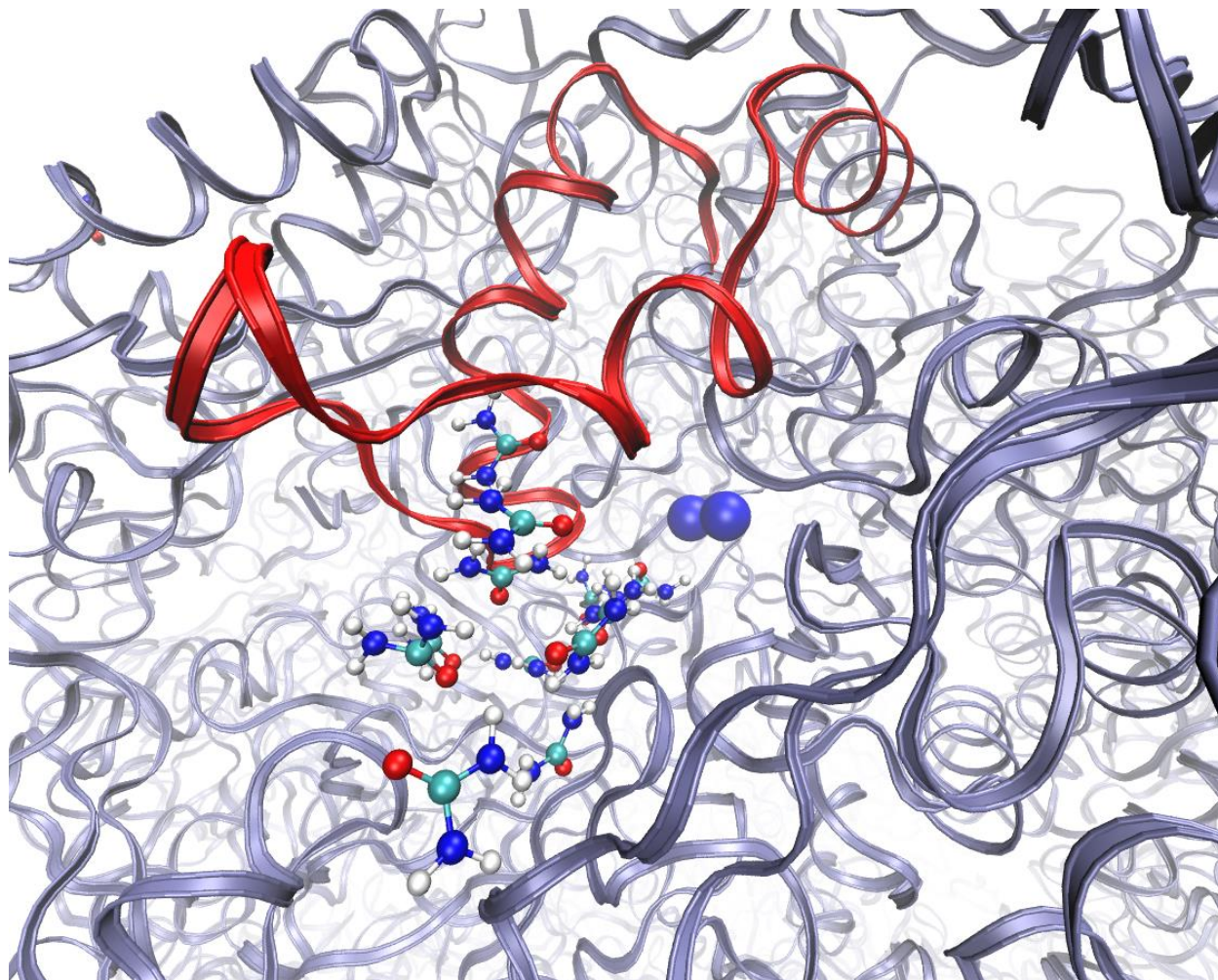
Figure 4-9. Snapshot showing accumulation of urea in the active site and flap of dimer 4 at 400 ns (wide-open flap state). Snapshots at 100 (closed) and 250 (semi-open) are included for comparison (Figures (4-10) and (4-11)).
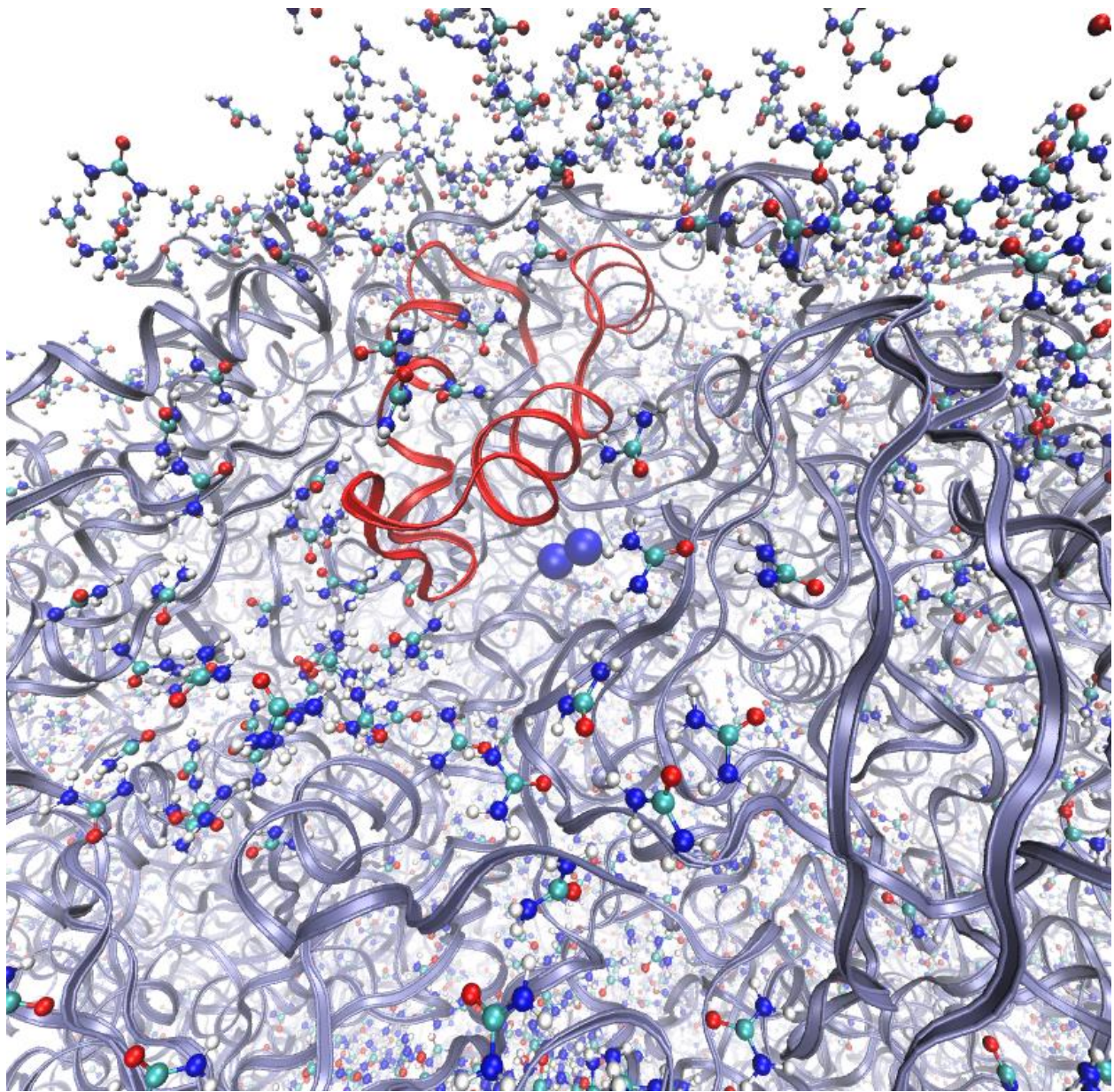
Figure 4-10. Snapshot showing accumulation of urea in the active site and flap of dimer 4 at 100 ns (closed flap state).
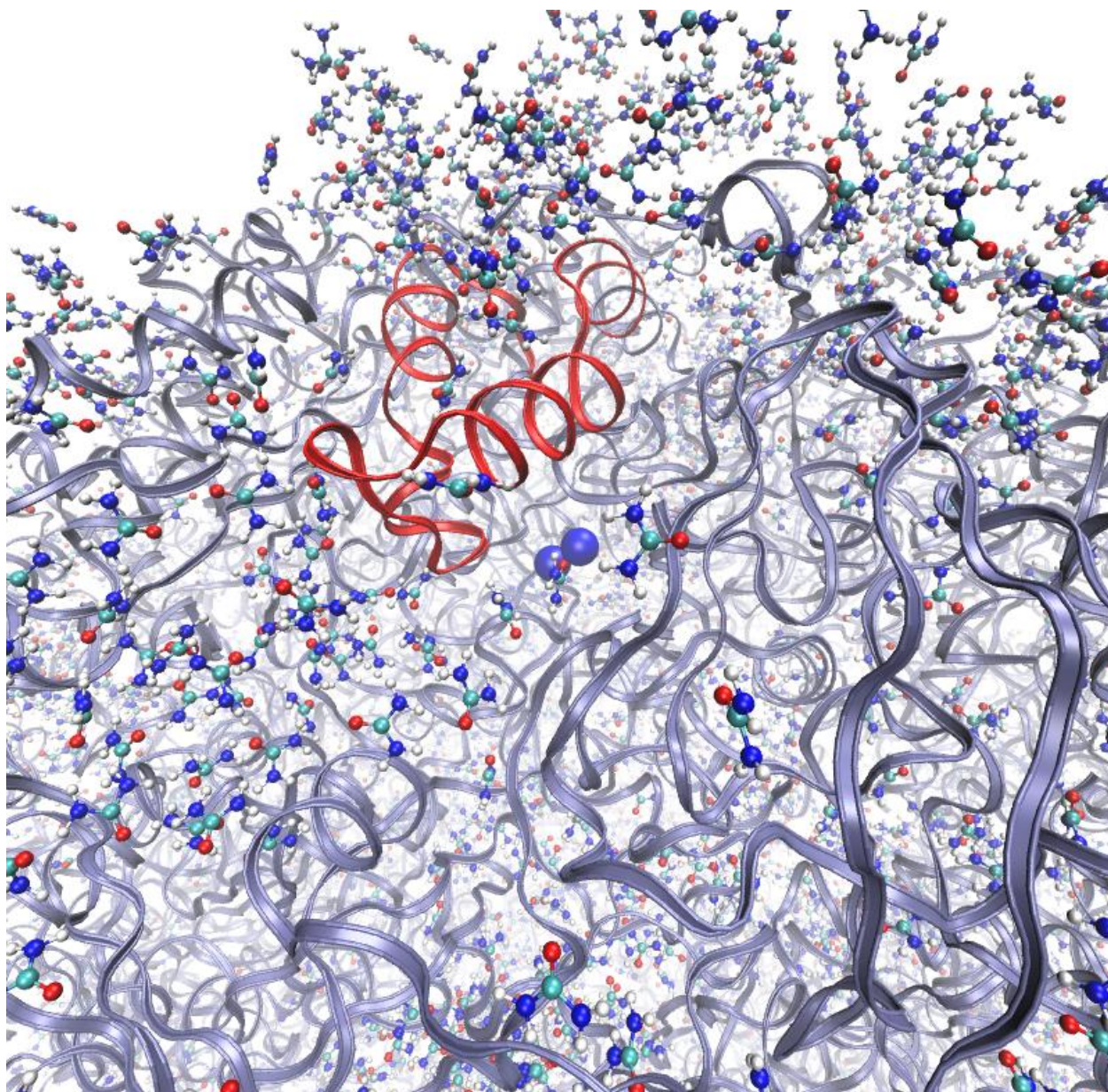
Figure 4-11. Snapshot showing accumulation of urea in the active site and flap of dimer 4 at 250
ns (semi-open flap state).

In dimers 4-8 we observe urea molecules within 7 Å of the pentacoordinate Ni^{2+}, some as

close as 2.85 Å. In conjunction with the active site-covering flap of dimer 4 reaching the wide-

open state at 346 ns, we further observe urea approaching the pentacoordinate Ni^{2+} at a distance

of 6.7 Å (357 ns), a second molecule draws within 3.19 Å and a third urea molecule is 3.85 Å

from this nickel ion near the conclusion of the simulation. There was a large influx of urea within

10 Å of the fifth active site during the simulation, concomitant with the active site-covering flap reaching the wide-open state. For Dimer 6 a urea molecule is within 10 Å of the pentacoordinate Ni^{2+} at 150 ns, and draws within 3.4 Å of the pentacoordinate Ni^{2+} at 347 ns. This urea molecule remains within 10 Å of the metal ion for the remainder of the simulation. Furthermore, in dimer three, urea is observed to approach within 9 Å of the pentacoordinate Ni^{2+}.

Residue separations indicate dimer 7 opens approximately 400 ns into the simulation (ILEα328/ALAβ170 [ILE566/ALA173] separation of 20-25 Å over the final 100 ns), corresponding with an observed influx of urea into the active site cavity over this duration. Flaps 1 and 10 stay closed throughout and there is little urea observed to approach the active site. Flap 12 opens slightly over the final 100 ns (ILEα328/ALAβ170 [ILE566/ALAβ170] separation of 18 Å), however there was little influx of urea into the active site cavity.

In the active sites of dimers 5 through 7, while the quantity of urea observed is lower than in dimer 4, urea molecules do approach the pentacoordinate Ni^{2+} much more closely, at distances of 3.19, 3.44 and 3.23 Å for active sites of dimers five, six and seven respectively. Figure (4-12) depicts the approach of a urea molecule to the 5-coordinate Ni^{2+} with the carbonyl in position for binding to the metal, a posited first step in the urea hydrolysis mechanism.[3, 49, 75]

4.3.7 Urea Residence Times and Saturation of High RMSF Regions

For this analysis we wrote a code that tracked the distance of every urea from every high RMSF region selected and we listed the urea that are within 2 Å of these regions at ns intervals. We chose 2 Å because it is a tight constraint and this is necessary due to complications associated with attempting to detail all hydrogen-bonding angles. Then we took the 9 high RMSF regions on the 12 dimers and binned the urea that resided for 10 nanoseconds or longer. We constructed a histogram that shows the quantity of urea for different time intervals exceeding 10 ns over the entire simulation.
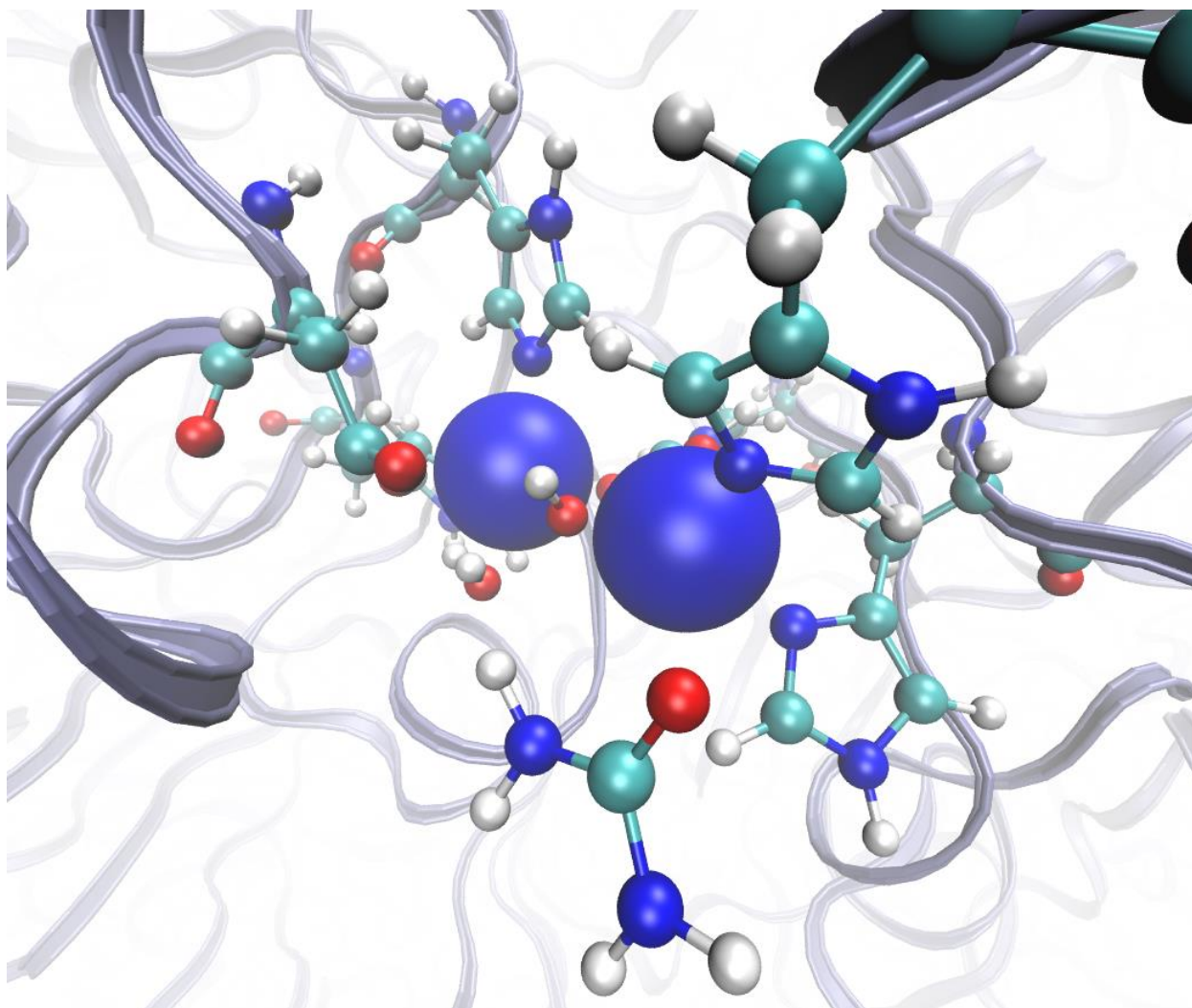
Figure 4-12. Urea molecule (bottom center of image) nearing pentacoordinate Ni^{2+} of dimer 5 at 331 ns.

The region formed by residues β55-β68, which is one of the regions of the horseshoe closest to the loop of the active site-covering flap, does not display a large quantity of urea molecules that interact for an extended duration (Figure (4-13)). Dimer 3 had 11 long-residence urea molecules and dimer 12 had only 1 such urea. Region β100-β111 (100-111) is a loop connecting 2 β-regions that makes up part of the horseshoes. More than one molecule of urea per residue is in close contact with this region (Figure (4-14)). For the region composed of residues β113-β130 (113-130) we see a much higher number of urea interactions (Figure (4-15)). In

dimers 3 and 11 we have over 100 urea molecules interacting for 10 or more ns. This region is the loop region located on the outside of the protein connecting regions of the beta chain found on the surface. Dimer 9 has no urea interacting for over 10 ns. The region β177-β184 (177-184) is one of the regions that are the closest three points of the horseshoe. All dimers except for dimer 11 have values between 1 and 32 long residence urea molecules, typically with only one such molecule (Figure (4-16)). Dimer 11 has 323 long residence urea molecules, which is the highest value observed. Region β222 to β238 (222 to 238) is a region found near the ovoid hole to the entrance of the hollow. There are three dimers with high numbers of urea molecules that interact, namely dimers 3, 7 and 11 which come together to form one of the ovoid holes (Figure (4-17)). Region α□ to α17 (239-255) is the beginning of the alpha chain and is the inner segment of the horseshoe and one of the three closest points to the active site-covering flap. This region is very near the active site flap of the adjacent dimer. In this particular histogram we don't observe very high values of urea residing for over 10 ns (Figure (4-18)). The dimer with highest value is dimer 8, with 24 urea molecules that reside for over 10 ns. Most of the dimers have 13 such urea molecules, with only 2 long residence urea molecules observed for dimer 1, which stays closed. These results are easily explained by the region's proximity to the loop of the active site flap.

We also compared the low RMSF regions of the protein to the high RMSF regions in order to ascertain whether there were as many interactions with urea throughout the simulation as found in the high RMSF regions. Most of the low RMSF regions are not found on the surface of the protein and are therefore not as exposed to the solvating solution as are the high RMSF regions that tend to reside on the enzyme exterior. The few regions that were exposed to the surface of the protein had very few urea molecules within a cutoff of 2 Å.
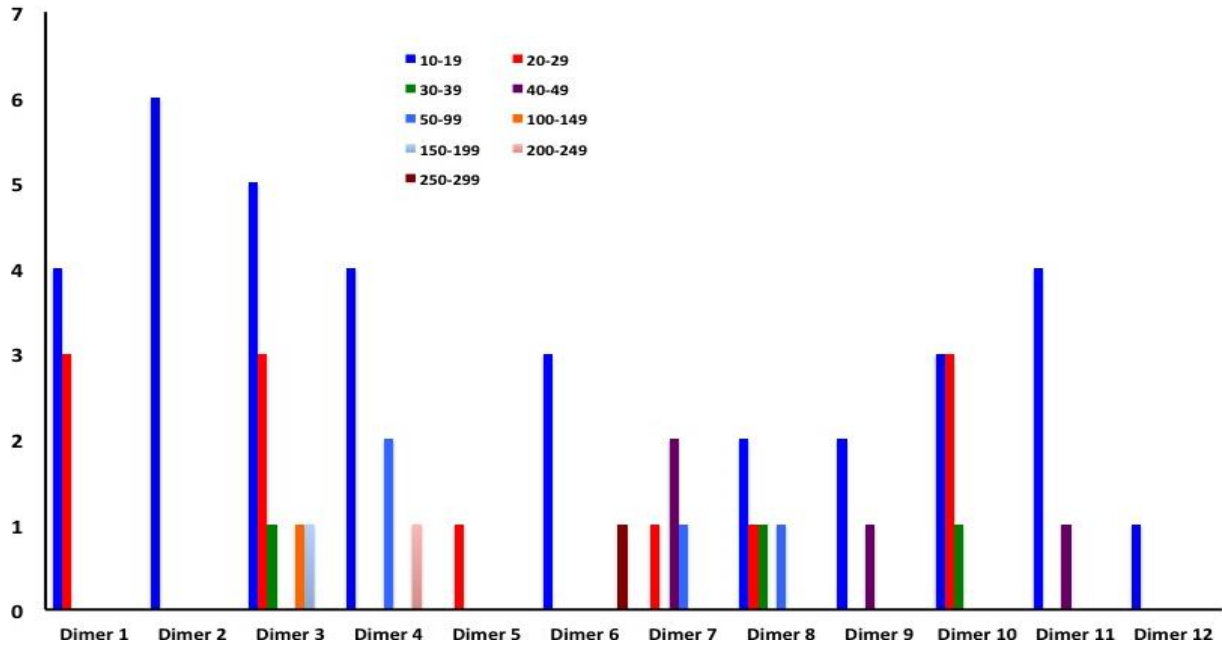
Figure 4-13. Distribution of urea residence times for high RMSF region 55-68.
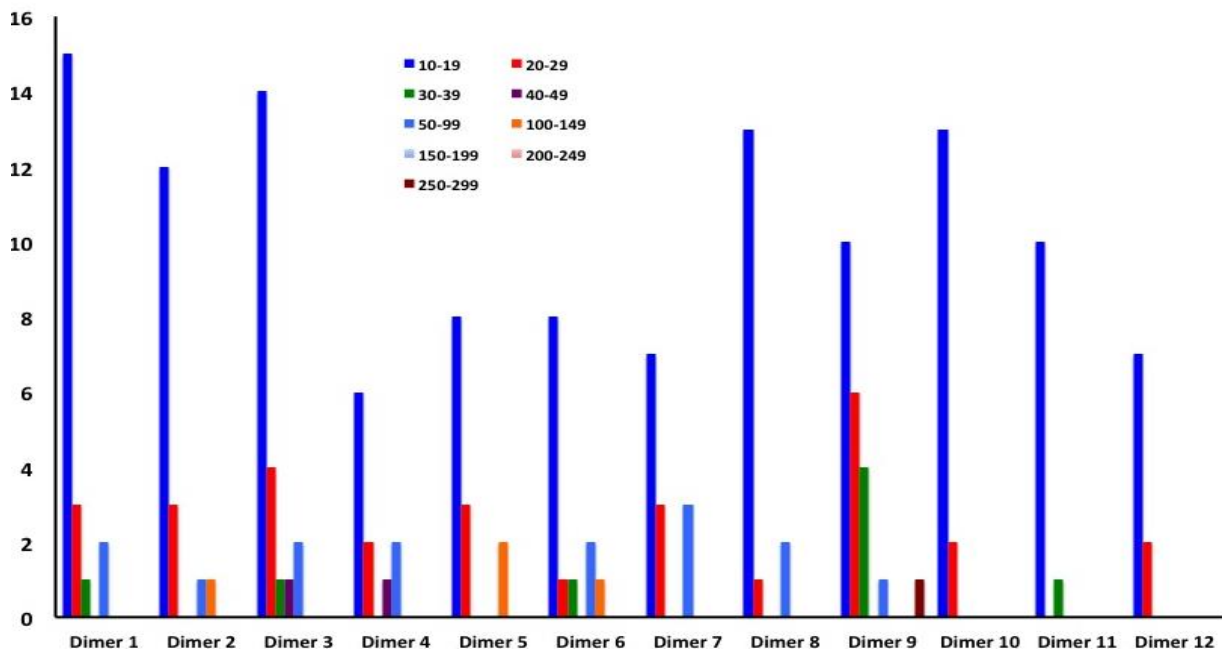
Figure 4-14. Distribution of urea residence times for high RMSF region 100-111.
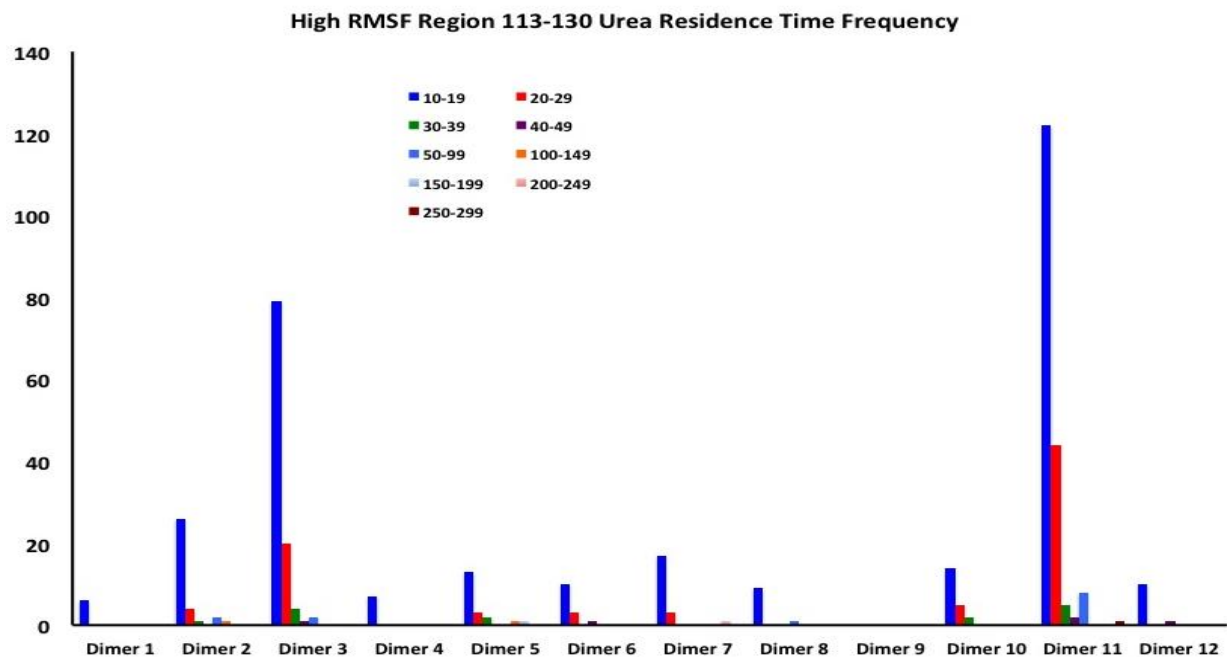
Figure 4-15. Distribution of urea residence times for high RMSF region 113-130.
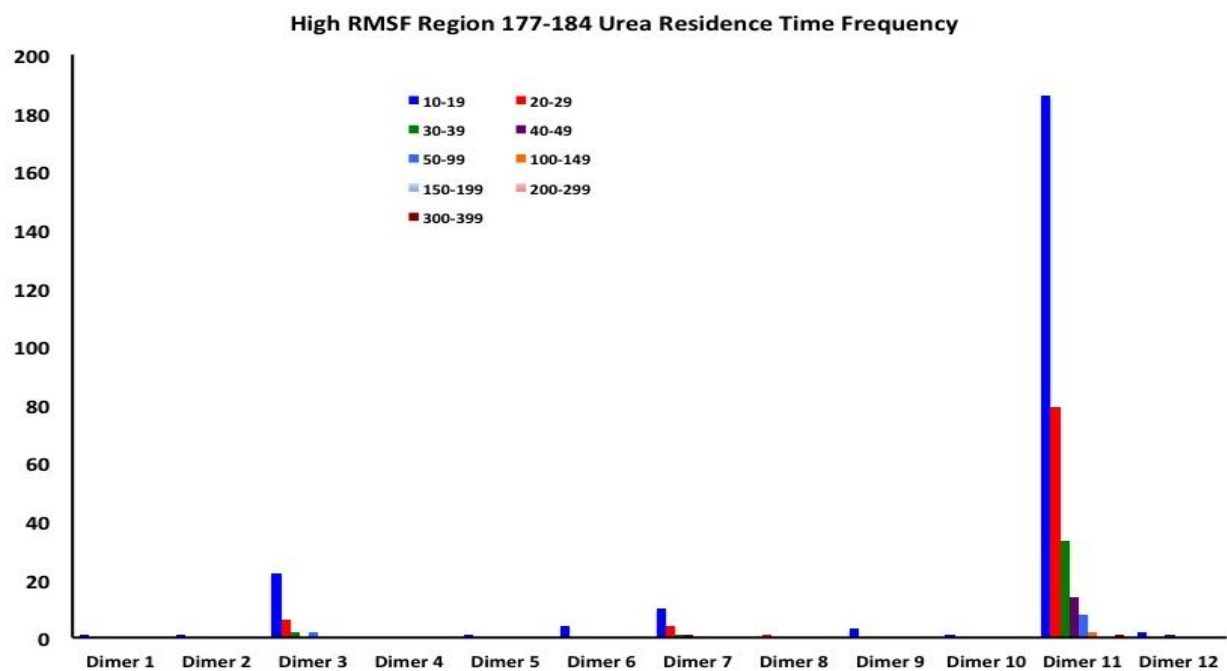
Figure 4-16. Distribution of urea residence times for high RMSF region 177-184.

High RMSF Region 222-238 Urea Residence Time Frequency
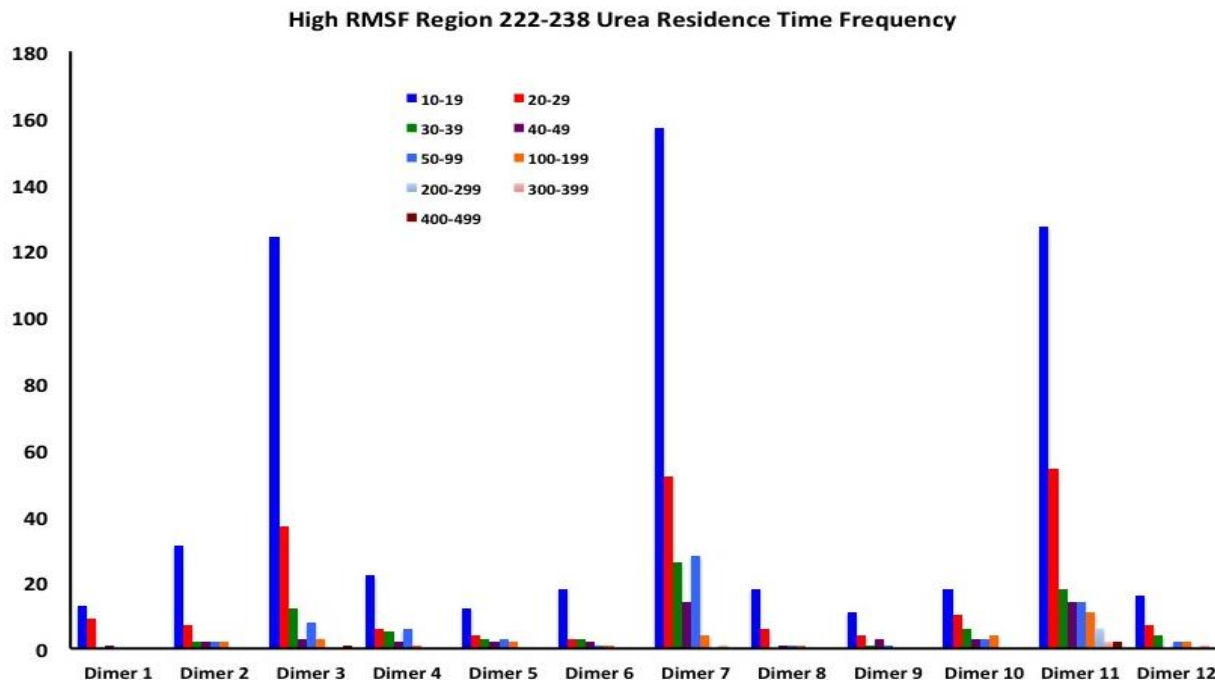
Figure 4-17. Distribution of urea residence times for high RMSF region 222-238.

High RMSF Region 239-255 Urea Residence Time Frequency
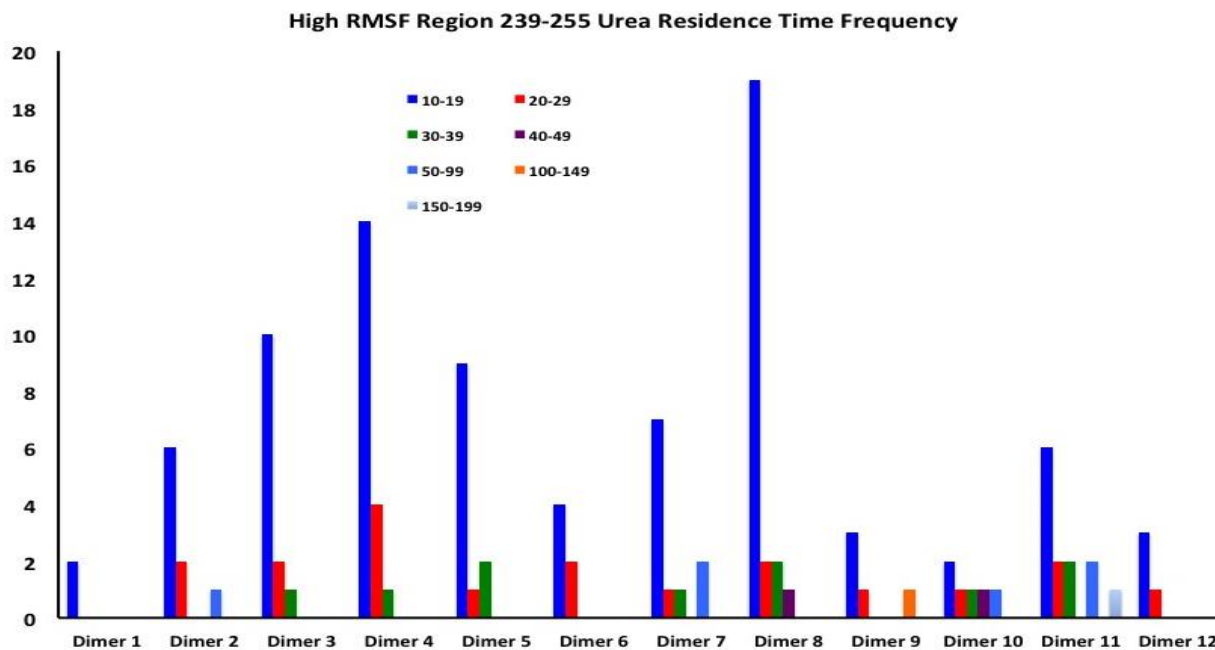
Figure 4-18. Distribution of urea residence times for high RMSF region 239-255

129

For example, within the region β30-38 (2466-2474) of dimer 4, only LEUβ38 (LEU2474) featured any noticeable interactions with urea molecules. Such interactions were few in number; interactions with only 5 unique urea molecules occurred throughout the entire simulation, which is low compared to the total number of urea molecules found in the bulk solvent, which contains around 4788 molecules. Given the high concentration of urea, it is inevitable for a few molecules to be within 2 Å of a residue on the protein's surface.

Region α326 to α330 (564-568) is the loop of the active site covering flap and contains a very small amount of urea per residue (Figure (4-19)). The high RMSF region formed by residues α387-α395 (625-633) corresponds to the loop connecting 2 α-helices found in the back of the active site. This particular region does not seem to attract that many urea molecules throughout the simulation as borne out in Figure (4-20). This is a logical occurrence due to the fact that the loop region is found on the inner segment of the protein. Furthermore, the loop is not near the active site flap and thereby cannot significantly contribute to shuttling urea into the active site; rather it could facilitate exit shuttling of hydrolysis products. In the region composed of residues α533-α553 (771-791) we observe that there are more than 1 urea molecule per residue for most of the dimers, except for dimers four, five and six (Figure (4-21)). Dimer 11 attracts 167 long-residence urea molecules. There seems to be no correlation between the RMSF value and the number of urea that interact. It does make sense for the active site loop not to have urea interact for an extended period of time, as this could attenuate active site entry rather than enhance it. The longest residence time urea molecules on the loop of the active-site covering flap are observed on flaps that remain closed throughout the simulation.
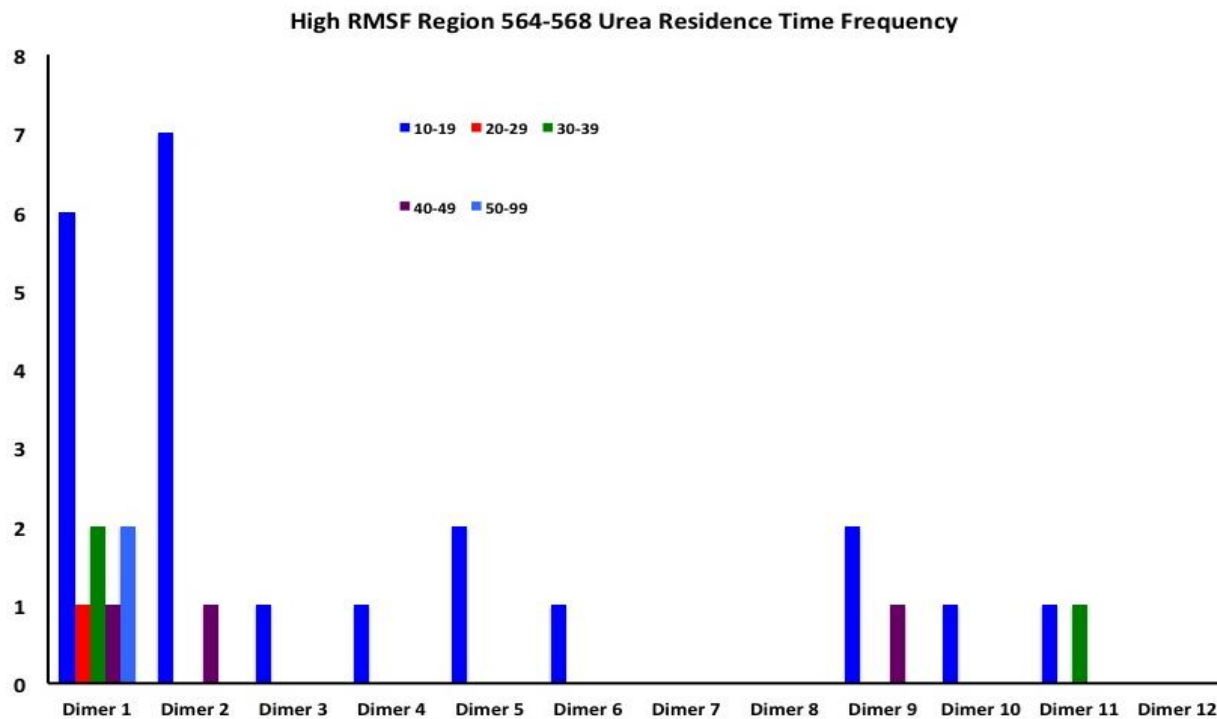
Figure 4-19. Distribution of urea residence times for high RMSF region 564-568.
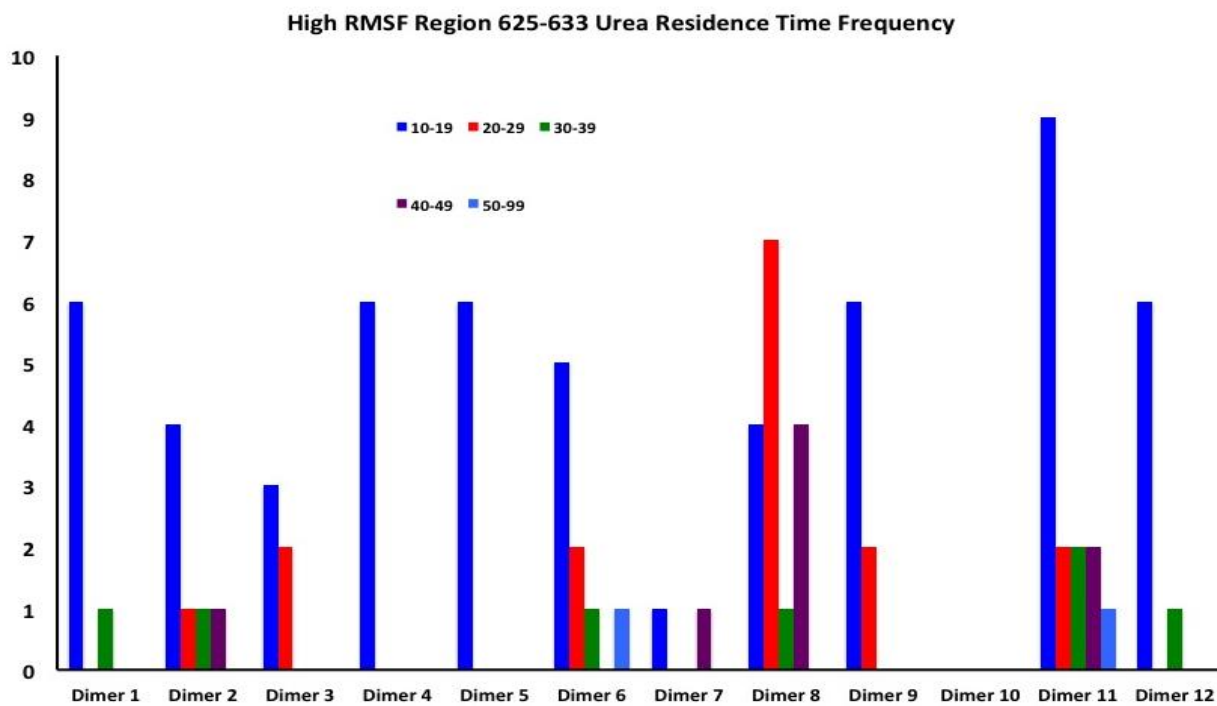
Figure 4-20. Distribution of urea residence times for high RMSF region 625-633.
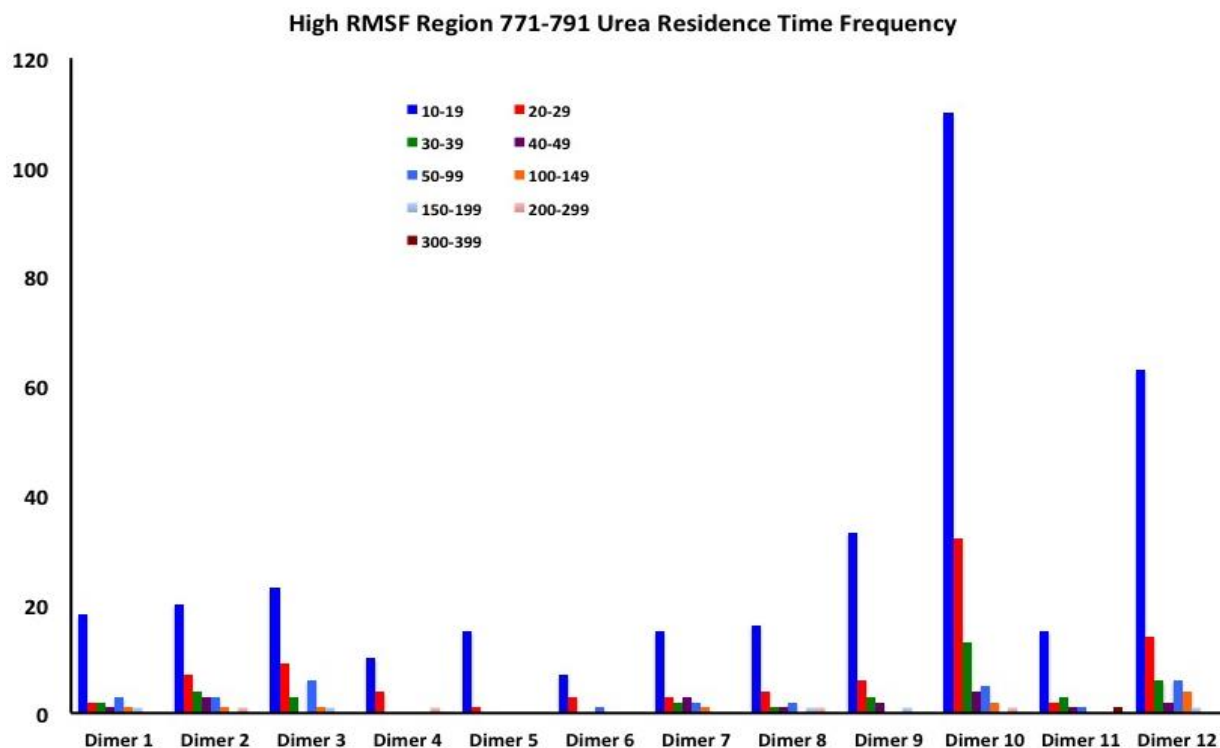
Figure 4-21. Distribution of urea residence times for high RMSF region 771-791.

We observed the number of urea molecules that came within 2 Å of high RMSF regions, including the active site loop and the surrounding horseshoe region, as a function of time. Particular focus was applied to the differences in the number of urea around the wide-open flap of dimer 5 and the closed flap of dimer 6. The loop region of the active site flap of dimer 5 attracts more urea molecules as the flap opens further. The same region for dimer 6, which remains closed for the duration of the simulation, has comparatively fewer urea molecules. Urea molecules approaching the high RMSF regions comprising the horseshoe surrounding the loop of the active site-covering flap, which is located on adjacent dimers, were charted as well. For the wide-open flap there are more attracted urea molecules on average than the horseshoe region surrounding the closed flap state. There are more urea molecules at the termini of the horseshoe surrounding the wide-open flap than that of the closed flap, but the inner regions of the

132

horseshoe have comparably similar urea activity in both dimers with only a slight excess to the regions about the wide-open flap.

4.3.8 High RMSF Regions and Active Site Entry

Having recognized the entry of urea into the active site and identifying specific urea molecules that draw near to the pentacoordinate Ni^{2+}, we monitored the position of these molecules before their entry into the active site. One common feature of the pathway undertaken by these urea molecules that accessed the active site is periods of interaction with multiple high RMSF regions on the exterior of the protein. We hypothesize that the outer residues of the protein function in a dual-fold manner, first to attract urea molecules and then to shuttle the substrate molecules toward the active site covering flap. We will detail the travels of two urea molecules as examples supporting our hypothesis, however these are not the only occurrences of this phenomenon.

Urea 20177 spends a significant portion of the simulation traversing the high RMSF regions. Following periods of interaction with other regions of high RMSF, this molecule begins to interact with β-chain residues 222-238 of chain 1 at 318 ns before shuttling to residues 564-568 (α326-α330) of α-chain 1 at 322-324 ns. Subsequently, the urea molecule enters the active site where it remains for the duration of the simulation in close proximity to the pentacoordinate Ni^{2+} ion. Snapshots of this sequence are shown in Figure (4-22). Of particular note in this sequence is that while the urea escapes from the loop of the active site-covering flap upon its initial contact, the horseshoe recaptures the molecule and shuttles it back for active site entry. Our second molecule, urea 13099, is observed to behave similarly to the previously described urea although with more observed escape and recapture events between the loop of the active-site covering flap and the high RMSF regions of the horseshoe over the final 104 ns of the simulation. First it is drawn to the inside portion of the horseshoe at 396 ns, after which time

occurs a series of transfer events in which the urea molecule bounces between the interior portion of the horseshoe and the loop of active site-covering flap through 406 ns. Subsequently the urea molecule is shuttled to other high RMSF regions comprising the horseshoe. The final urea transfer to the loop of the active site covering flap occurs at 446 ns and following this transfer the molecule descends into the active site cavity for the remainder of the simulation. These travels are displayed in Figures (4-23)-(4-28).

4.4 Conclusions

Herein we have presented the results and analysis of a molecular dynamics simulation of HP urease in a 10.5 M aqueous urea solution over 500 ns. We compared the results to an extension of our previous study, a 400 ns MD simulation[46] of aqueous HP urease that we lengthened by 100ns. The flap motion that was previously observed, including the identification of a wide-open flap state, was also observed in the active site covering flaps over the course of our aqueous urea simulation. A direct correlation between the flaps opening and the entry of urea is observed, where a wider opening of the flap results in the entry of significantly more urea in to the active site cavity. It appears that urea cannot access the active site using the semi-open state observed experimentally[6], suggesting that the wide-open state observed computationally is necessary for active site ingress by urea. We comment on our observations regarding the regions of high RMSF, which form horseshoes about the loops of the active site-covering flaps. We also analyzed the first solvation shell to ensure the protein is not denaturing. In the bulk solvent the water to urea ratio is 5.29:1, however the first solvation shell (2 Å from the protein surface) generally has a ratio of 4:1 over the duration of the run.

The most important finding in this study is the horseshoe surrounding the active site-covering loops that is created by the high RMSF regions of the β-chain, and the mechanism *via* which urea is shuttled into the active site. There are three points specifically on the horseshoe

that are closest to the active site flap that seemingly guide urea to the loop of the active site flap and allow subsequent entry into the active site cavity. From this new observation we can now target other regions of the protein that may negatively affect the enzymatic hydrolysis of urea. It is feasible that binding or otherwise reducing the urea attracting capability of these three closest points could attenuate urea entry into the active site. Further disruption or severing of the horseshoe composing residues, either enzymatically or via other means, could affect the urea hydrolysis process as well. This could be further evaluated by creating a mutated version of *H. pylori* urease by altering one or more of the high RMSF residues comprising the horseshoe regions.
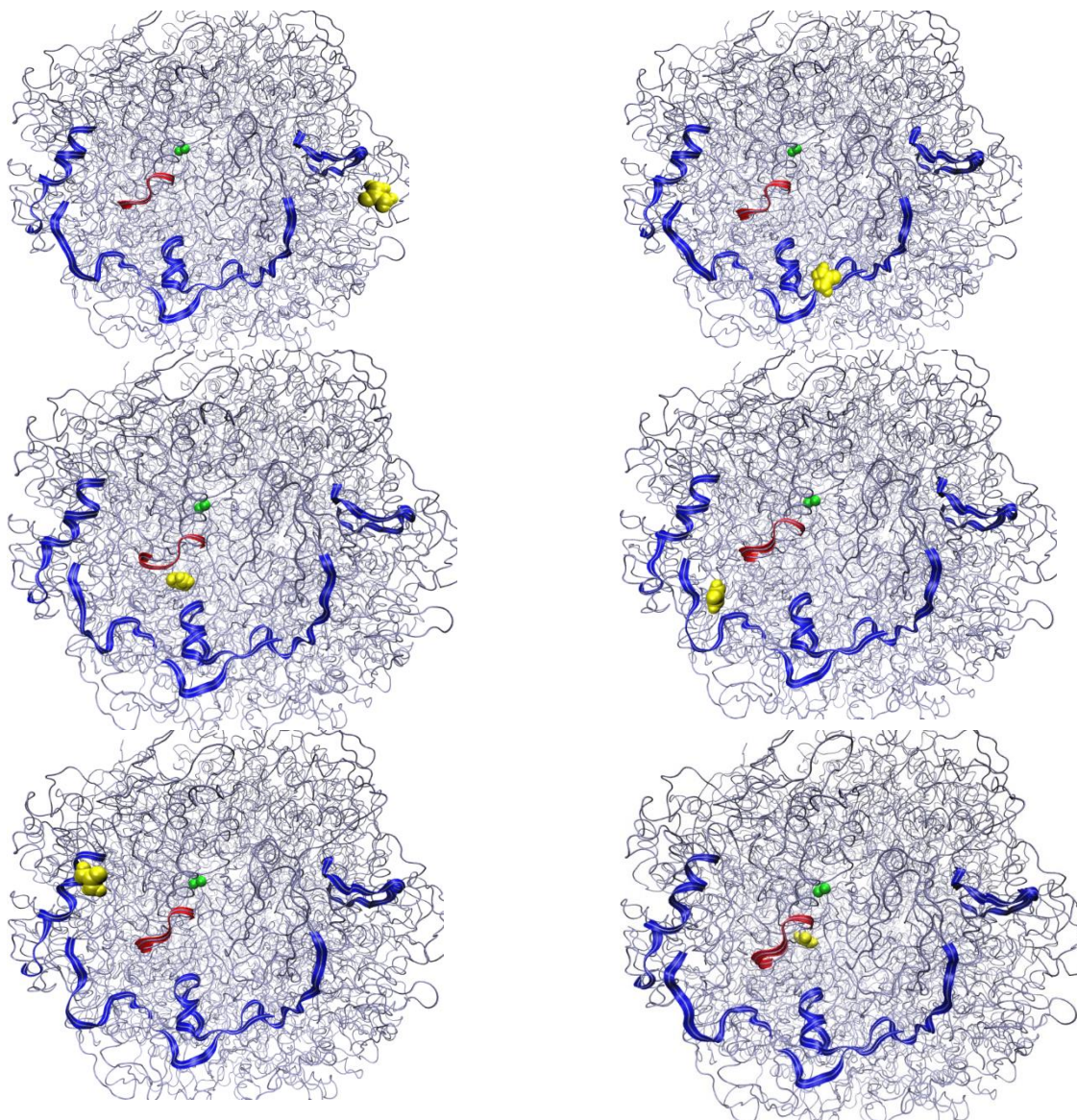
Figure 4-22. Snapshots of urea molecule 20177 interacting with regions of high RMSF from 318-328 ns before descending into the active site cavity. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 318 ns (top left), 319 ns (top right), 321 ns (middle left), 322 ns (middle right), 324 ns (bottom left) and 328 ns (bottom right).
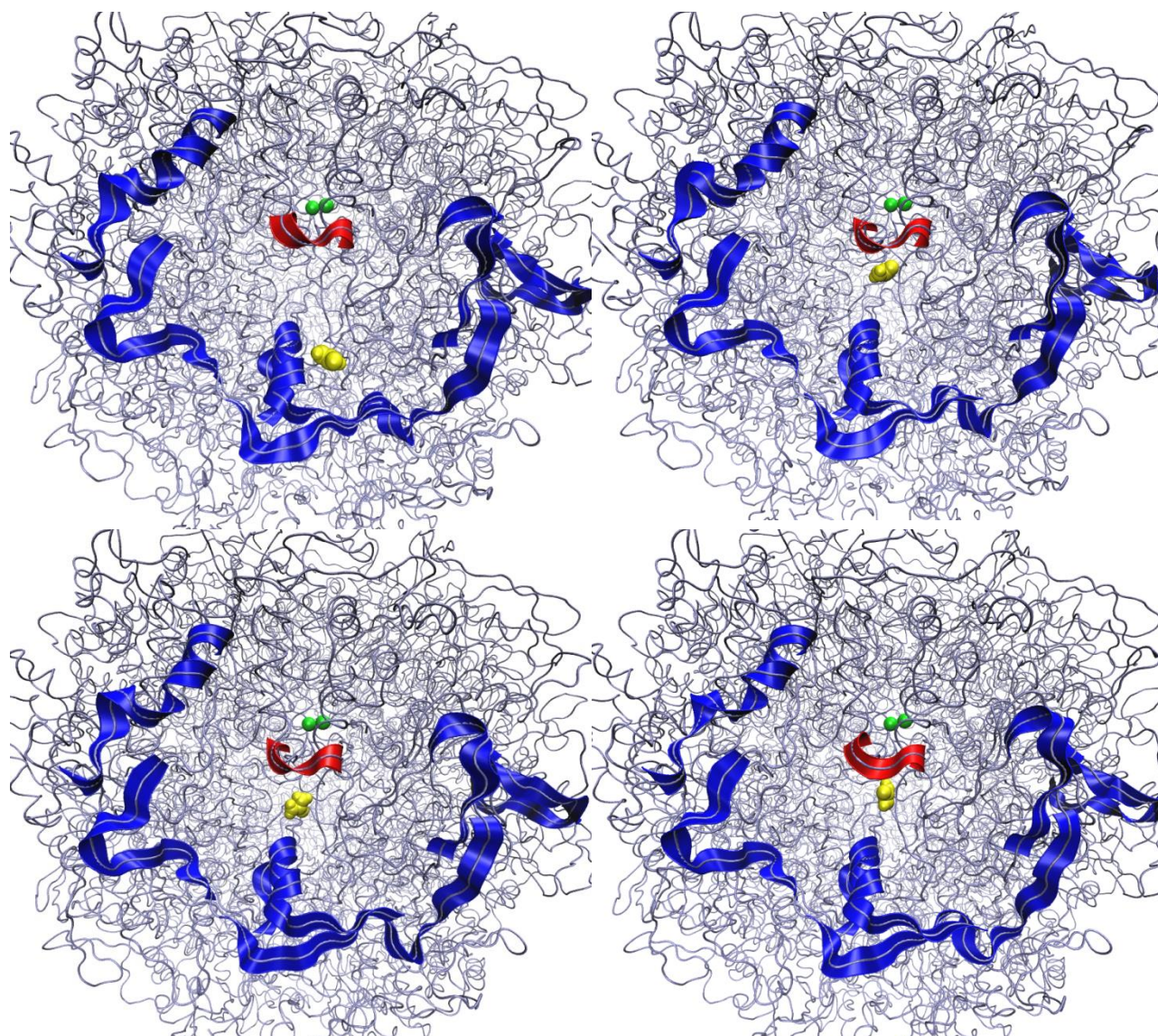
Figure 4-23. Snapshots of urea molecule 13099 interacting with regions of high RMSF from 396-403 ns. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 396 ns (top left), 397 ns (top right), 398 ns (bottom left) and 403 ns (bottom right).
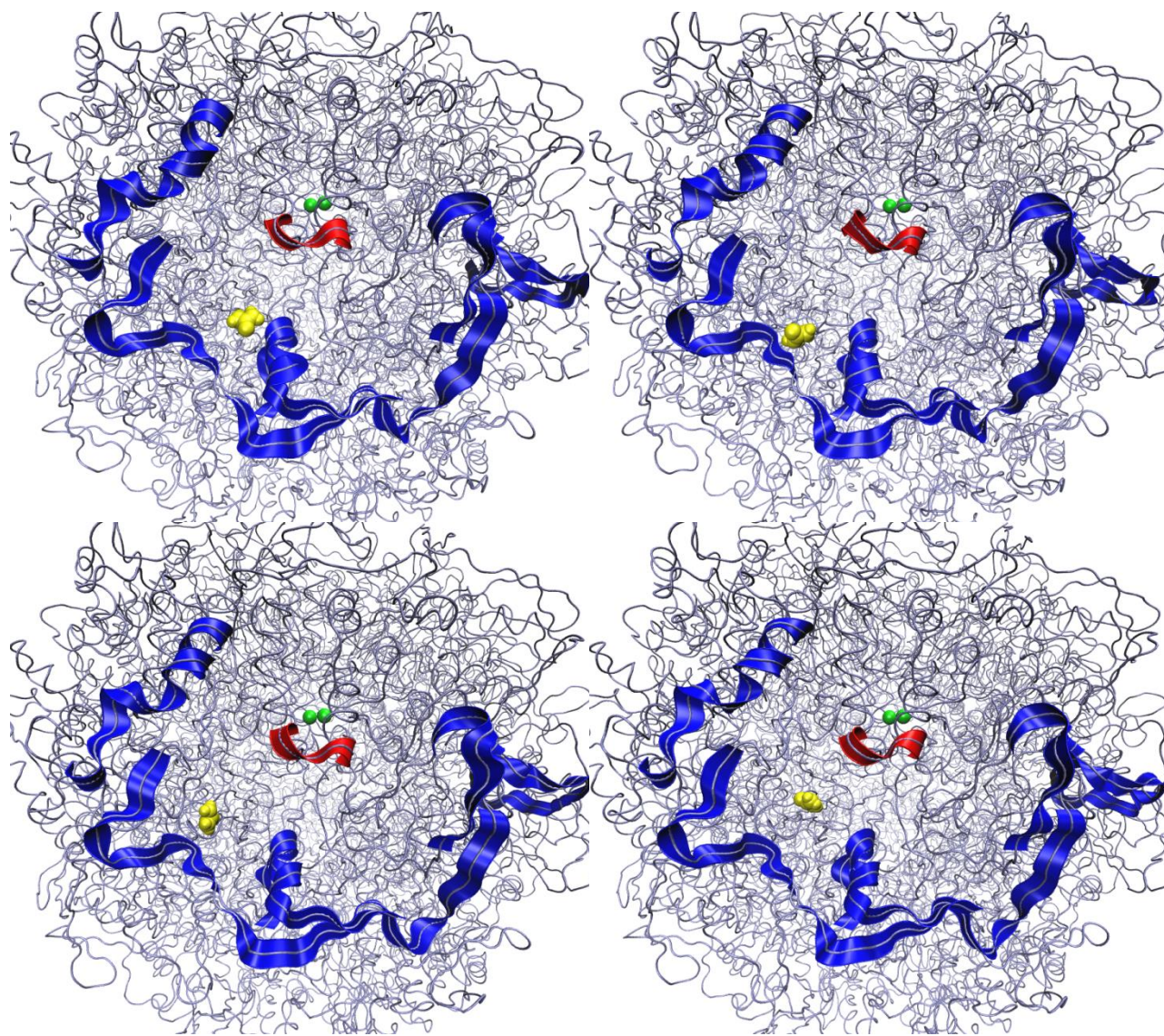
Figure 4-24. Snapshots of urea molecule 13099 interacting with regions of high RMSF from 406-417 ns. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 406 ns (top left), 410 ns (top right), 416 ns (bottom left) and 417 ns (bottom right).
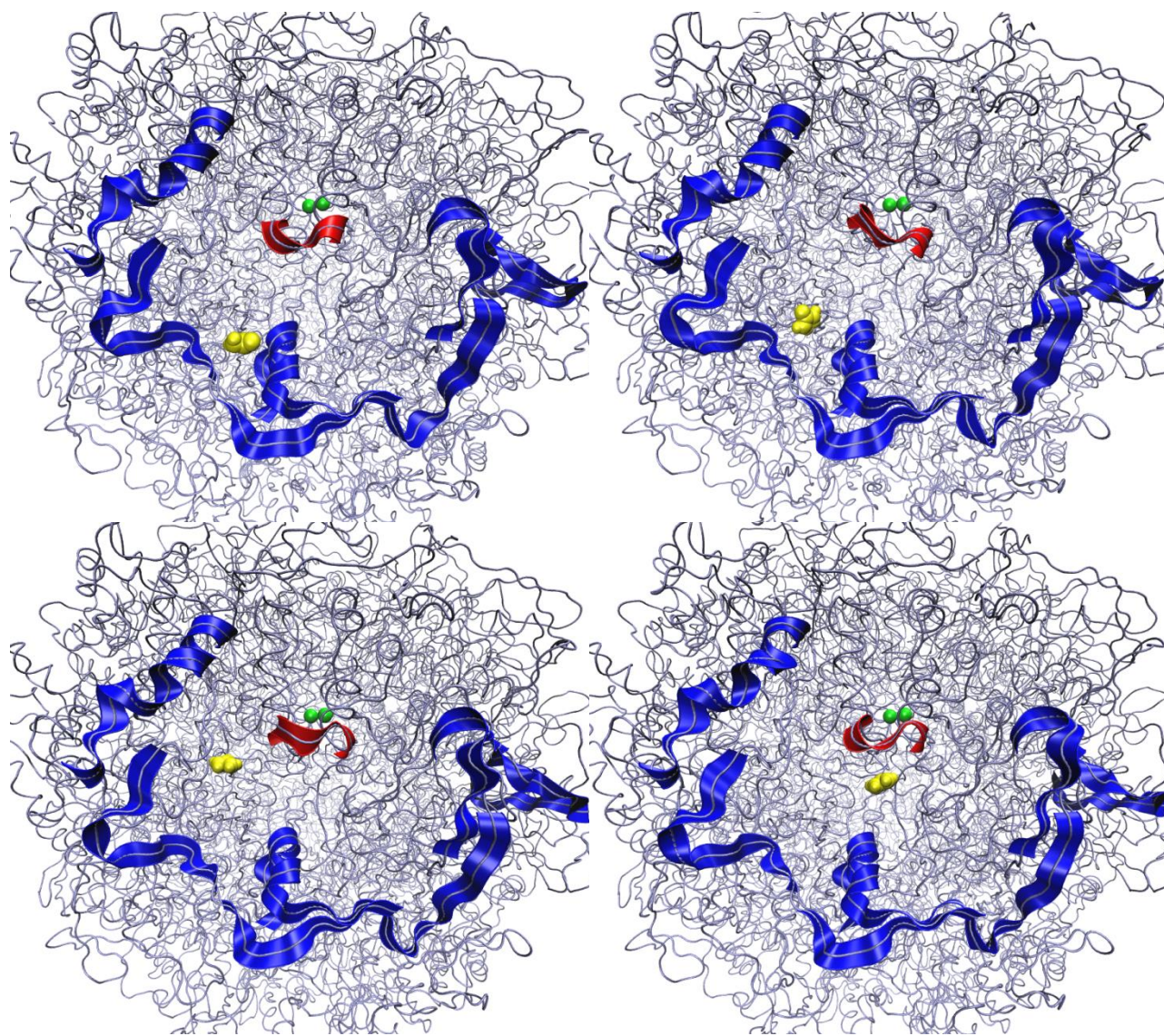
Figure 4-25. Snapshots of urea molecule 13099 interacting with regions of high RMSF from 419-424 ns. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 419 ns (top left), 422 ns (top right), 423 ns (bottom left) and 424 ns (bottom right).
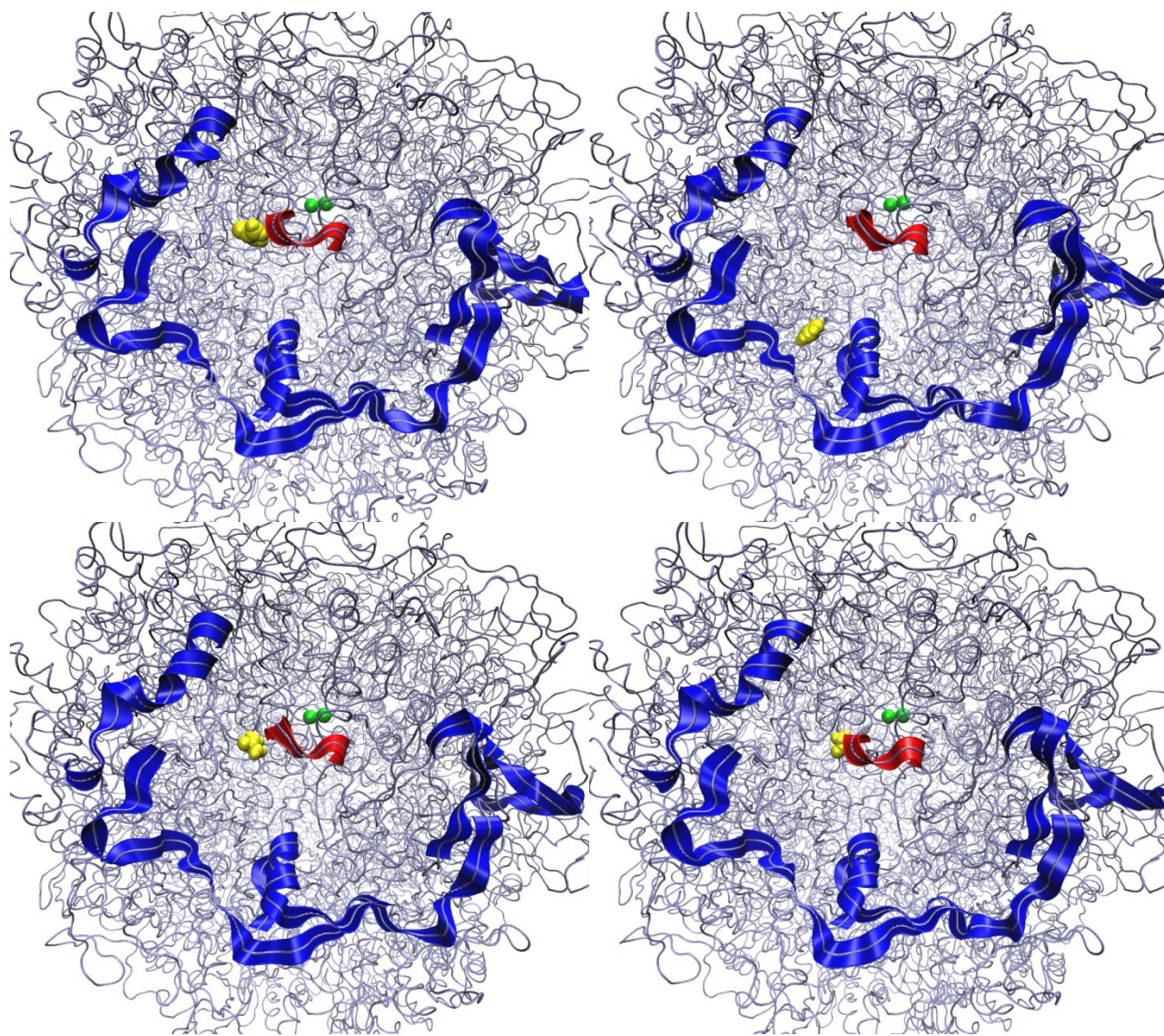
Figure 4-26. Snapshots of urea molecule 13099 interacting with regions of high RMSF from 432-447 ns. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 432 ns (top left), 438 ns (top right), 446 ns (bottom left) and 447 ns (bottom right).
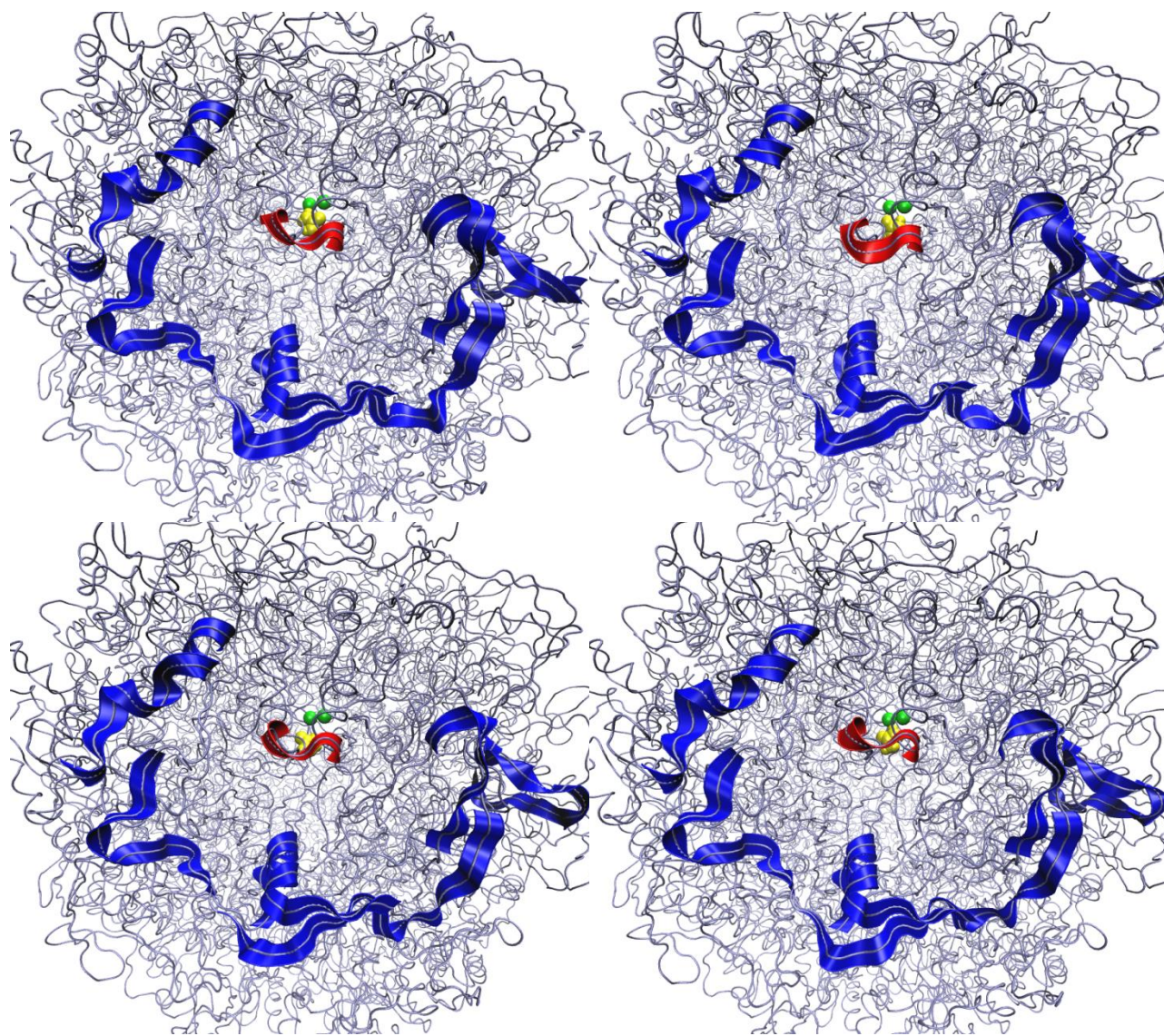
Figure 4-27. Snapshots of urea molecule 13099 interacting with regions of high RMSF from 450-480 ns. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 450 ns (top left), 467 ns (top right), 475 ns (bottom left) and 480 ns (bottom right).
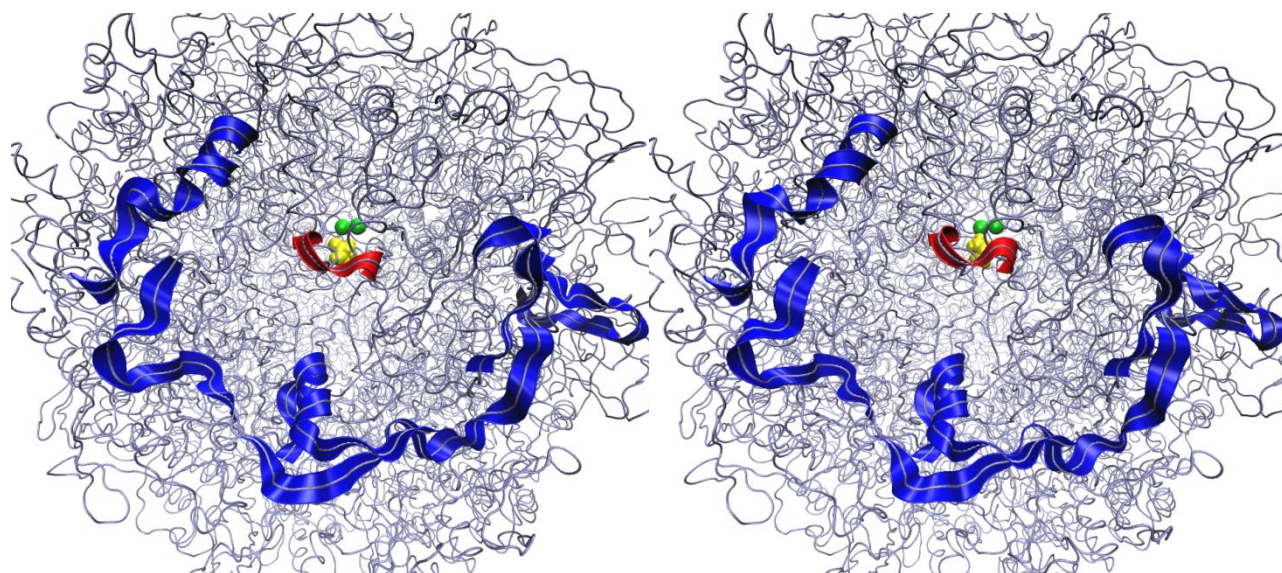
Figure 4-28. Snapshots of urea molecule 13099 interacting with regions of high RMSF from 490-500 ns. High RMSF regions comprising the horseshoe are shown as blue ribbons. The loop of the active site-covering flap is depicted as a red ribbon. The Ni^{2+} ions are depicted as green van der Waals spheres and the urea molecule as a collection of yellow van der Waals spheres. The snapshots correspond to 490 ns (left) and 500 ns (right).

CHAPTER 5
MOLECULAR DYNAMICS STUDIES OF HP UREASE IN 150 mM AQEUOUS NH$_4$Cl

In this chapter, we outline the effect that a 150 mM solution of ammonium chloride has on the HP urease enzyme. We are interested in studying the effect of ammonium chloride on the MD simulation of HP urease because of the natural process of urea hydrolysis that produces ammonium ions. *H. pylori* survival is dependent upon the pH of the environment. The low-pH environment of the human stomach could instantaneously destroy the bacteria, which thrive between pH values of 6 and 8, if it were not for the presence of HP urease. We are interested in observing the patterns and locations of ammonium ions that are stored within the protein, as well as uncovering any yet-unknown pathways involving these ions.

5.1 Methods

5.1.1 Structure Preparation

As in our previous MD efforts on aqueous HP urease and HP urease in 10.5 M aqueous urea, we selected PDB structure 1E9Z[6] as the starting point and utilized PyMOL[55] to generate the dodecamer. The H++ online protonation server was used to obtain the correct protonation states of the amino acids throughout the structure at a pH of 7.5[56, 67] following our previous procedure.[46] The bonded model was created with the MCPB component of the tLeap facility (AmberTools version 1.5)[21b] to define the Ni^{2+} coordination site and subsequently the hydrogen atoms and water molecules were added**.** The structure was solvated by a periodically replicated TIP3P[68] octahedral water box using tLeap. A total of 53,093 water molecules were used to hydrate the structure. In order to run this simulation in a 150 mM solution of ammonium chloride, we needed to neutralize the system by placing the appropriate excess of NH$_4^+$ ions. HP urease has a total charge of -44. Calculating the molarity of the system to measure out ammonium ions and chloride ions, we placed 143 ammonium ions and 99 chloride ions; the ion

discrepancy addressed the charge of HP urease, neutralizing the overall charge of the system. The metal parameters developed for *K. aerogenes* urease by Roberts et al.[53] with the MTK++/MCPB utility of AmberTools were modified utilizing previously developed Lennard-Jones parameters[58] (R* and ε) for Zn^{2+} as parameters for the Ni^{2+} ion; this was done to be consistent with our previous effort.[46]

5.1.2 Minimization and Equilibration

The two-stage energy minimization of the structure was performed using the FF99SB[36, 69] force field. Ni^{2+} coordination sites were optimized by imposing weak harmonic positional restraints of 10 kcal/(mol•$Å^2$) on all atoms outside the Ni coordination sphere. The steepest descent method was first used for $1x10^5$ steps to minimize the active site. Subsequently, the entire structure was relaxed and minimized using steepest descent for $1x10^4$ steps. The initial step length was decreased to $1x10^{-6}$ Å. After minimization the structure was equilibrated in another two-stage process. In the first stage, the temperature was raised from 0 to 300 K over $1x10^6$ steps of MD with a step size of 0.002 ps (2 fs) in the canonical (NVT) ensemble, placing a weak harmonic positional restraint on the whole protein. After the system was brought to 300 K, the simulation was run for 10 ps in the isobaric, isothermal (NPT) ensemble after removal of all the harmonic restraints. Temperature control was achieved using Langevin dynamics with a collision frequency, γ, of 2.0 ps^{-1} during the first, and 1.0 ps^{-1} during the second equilibration stage. SHAKE[37] was used to constrain all hydrogen-containing bonds during both equilibration steps. Long-range electrostatic interactions were computed using the particle mesh Ewald method with an 8 Å cutoff.

5.1.3 Molecular Dynamics

A 500 ns MD simulation was run on *Helicobacter pylori* urease in 10.5 M aqueous urea using the PMEMD version of Amber $12^{21a, 54}$ on an M2090 GPU using the FF99SB force field.

The production MD run was procured over 500ns in the isobaric, isothermal (NPT) ensemble, which was acquired using 2.0×10^8 steps, with a 0.002 ps time step. The temperature was kept constant at 300 K using Langevin dynamics with a collision frequency of 1.0 ps^{-1} while the pressure was maintained at 1 bar with a pressure relaxation time of 2.0 ps. SHAKE was again used to constrain all hydrogen-containing bonds. For the calculation of the non-bonded interactions, we again made use of the particle mesh Ewald method, and a cutoff distance of 8 Å was employed while computing the long-range electrostatic interactions. Frames were saved every 5×10^3 steps (10 ps), providing 5×10^4 frames.

5.1.4 Analysis

All frames were analyzed for the flap state defining residue separations, RMSD, atomic fluctuations (RMSF), and correlation matrices using the PTRAJ utility of AmberTools version 1.5. RMSD (root mean squared deviation) was obtained for each atom in the entire protein with the first frame of the trajectory as the reference and gives an average deviation of the protein geometry from the reference structure at each point over the 2 site-covering flaps and obtained the RMSD for each of these. The RMSF (root mean squared fluctuation) was computed on a per residue basis, using the alpha carbon of each residue as the reference point and providing a time-averaged value. The solvent accessible surface area was determined using NAccess V2.1.1[70] which implements the method of Lee and Richards[71] in the area determination. The secondary structure assignments were made utilizing the DSSP program of Kabsch and Sander.[72]

Throughout our discussion of the residues, we will identify them both by the residue numbering according to Ha et al. (α and β chains)[6], as well as the residue numbering from the PDB (1-807). For the purposes of this analysis, the flap covering the active site was considered to span residues α304-347 (542-585) with the first α-helix composed of residues α304-322 (542-560), the turn composed of α323-329 (561-567) and the second α-helix spans α330-347 (568-

585). The α-helices are extended from the helical section described by Ha et al.[6] in order to account for all residues in the flap region that adopt α-helical character at some point during the simulation.

5.2 Results

To understand the function of HP urease in the presence of ammonium ions, a urea hydrolysis product, and identify potential inhibition targets, we analyzed all relevant elements of the MD simulation. Analysis of the active site provided a starting point for analysis. Beyond this, we studied the solvent-accessible surface area to exclude the possibility of denaturation as a result of solvent and ionic composition of the solution, as well as the first solvation shell to determine incorporation of ammonium and chloride ions relative to the bulk solvent ratios. We also observed the behavior within the hollow region throughout the simulation. Other regions of interest include the high-RMSF regions, as well as the postulated exhaust flaps, which may serve as an ammonium exhaust conduit. Another point of interest involved the interaction of the ammonium ions themselves with other regions and residues; interactions with LYSα109 (LYS347) were given particular attention. Finally, we compared the ammonium chloride trajectory with that of past simulations to identify any nonspecific differences related to solvent.

5.2.1 Basic Statistics

Temperature remains constant at 300 +/- .5 K over the entire MD simulation, and the pressure fluctuates from -450 to 400 bars. The system energy remains constant at 580,000 kcal/mol. The average radius of gyration levels slightly over 61.5 Å while the radius reaches a maximum of 94 Å. This average radius is in accord with that observed in our aqueous simulation[46] and is nearly 1.5 Å larger than that observed in our simulation in aqueous urea. The maximum radius exceeds that of our previous simulation in aqueous urea by 4 Å, while it is 2 Å less than the maximum observed in the aqueous urease simulation (Figure (5-1)). The overall

RMSD of the ammonium chloride simulation levels off early in the simulation at nearly 2.5 Å, which is in accord with previous urease simulations and indicates that the structure is not undergoing a rapid structural expansion and degradation. The RMSD of most of the individual flaps levels off at a value slightly less than 3.5 Å. Flap 4 seems to be gradually increasing at the end of the simulation. Flap 2, while the RMSD starts out level, is increasing at the end of the simulation. Still, flap 2 has a max RMSD of 2.2 Å.
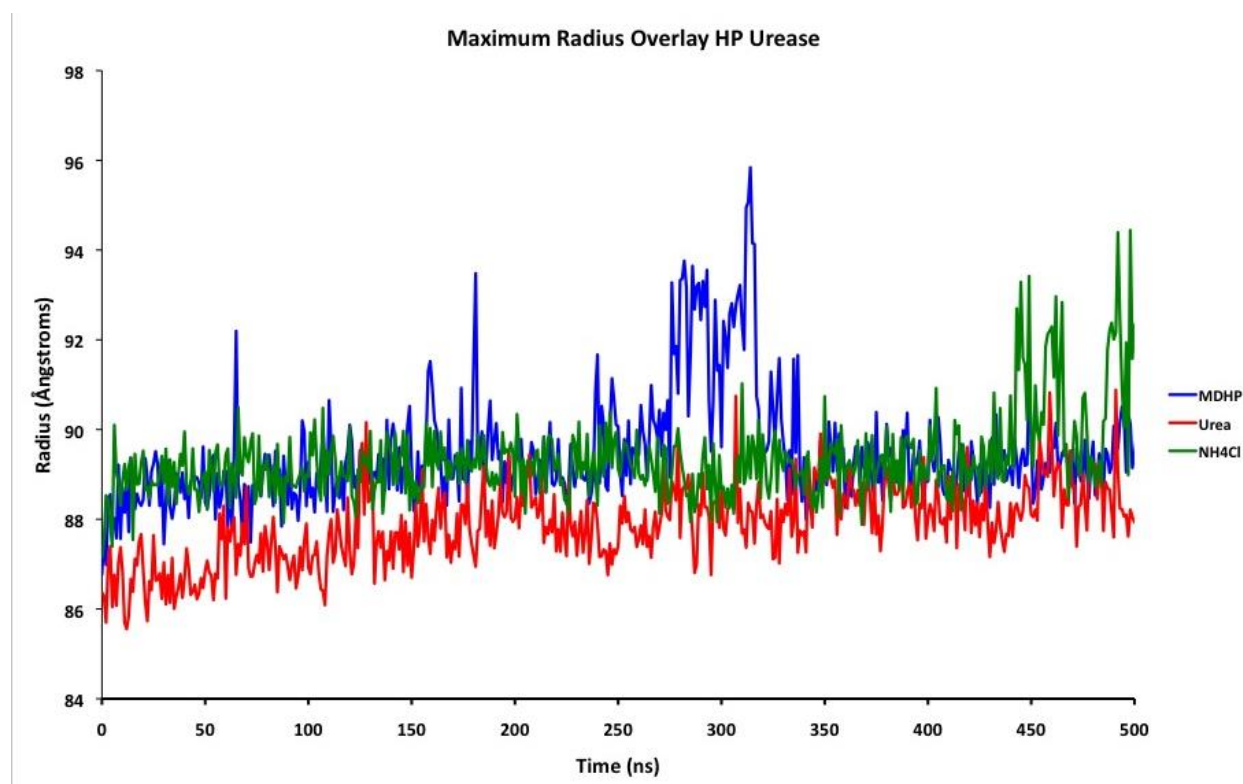
Figure 5-1. Overlay of maximum radius of gyration over time of aqueous HP urease (blue), HP urease in 10.5 M aqueous urea (red) and 150 mM NH₄Cl (green).

5.2.2 Active Site Flap Motion

We analyzed the distances from three residues on the active site flap to three separate residues at the base of the protein in order to characterize the opening of the flap. These specific residue pairings highlighted the separation between ILEα328/ALAβ170 (ILE566/ALA170), HISα322/GLYα47 (HIS560/GLY285) and GLUα330/ALAβ173 (GLU568/ALA173). We

147

observed more movement of the flap than in our previously reported simulations. One rationalization for this increase in motion is that the ammonium ions disrupt hydrogen-bonding interactions that occur between the flap and the protein. This increased motion manifests as more noise in the data. We found that flaps 3, 4 and 8 entered the wide-open state during this time interval. We define the wide-open state as reached when the separation between ILEα328 and ALAβ170 (ILE566 to ALA170) (IA) reaches 27 Å or higher. In flap 4 we see the largest value for the IA distance reaching 30 Å around 200 ns after starting from a closed state with IA distance 12 Å (Figure (5-2)). After several fluctuations, Flap 4 closes around 350 ns but then opens to the semi-open state with an IA distance of 26 Å around 400 ns. Flap 3 opens completely within the first 50 ns and returns to the semi-open state around 350 ns. We also observed flaps 2, 5, 6, and
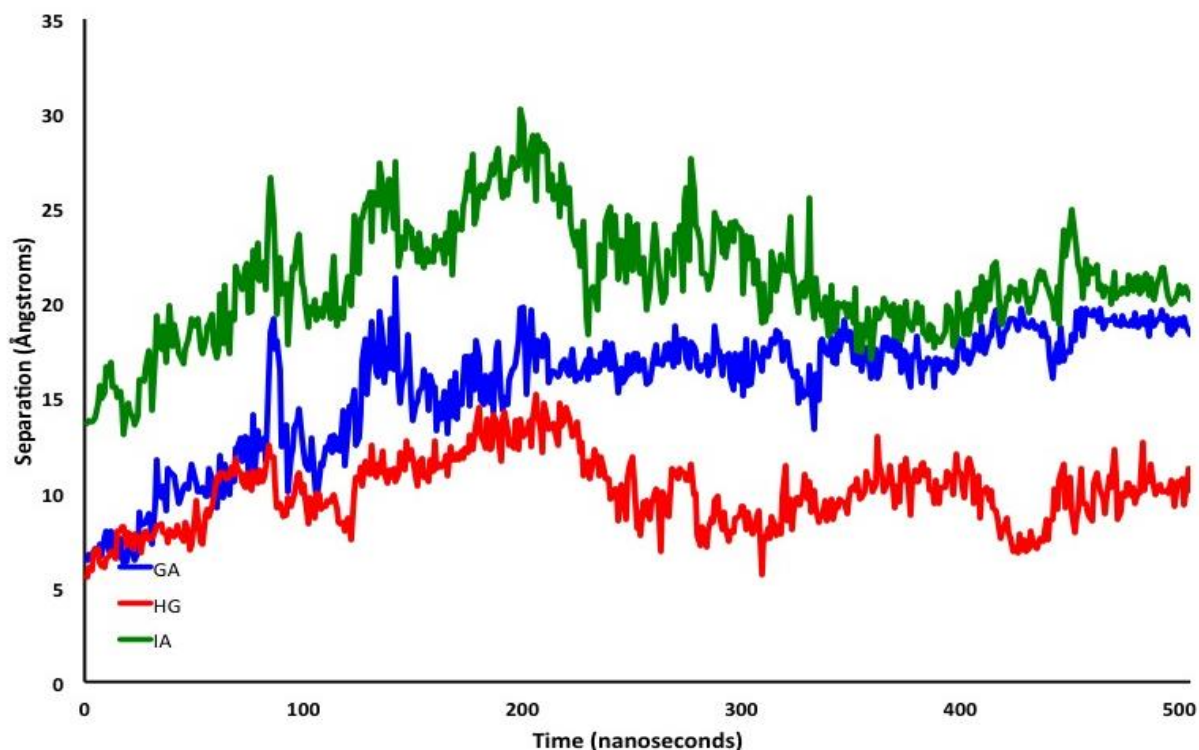
Figure 5-2. Separations (Å) between GLUα330/ALAβ173 (GA, blue), HISα322/GLYα470 (HG, red) and ILEα328/ALAβ170 (IA, green) in dimer 4 for HP urease in 150 mM aqueous NH$_4$Cl.

7 enter the semi-open state during the course of the simulation. The semi-open state is defined by IA separations of 23 to 26 Å. The remaining flaps do not open as wide, except for fleeting periods of simulation time. For example, in flap 10 the IA separation starts at 12 Å, but at around 300 ns the distance spikes to 21 Å before quickly dropping back to 12 Å.

We constructed relative free energy diagrams based on the residue separations outlined above, as described in our previous reports,[46, 59, 76] in order to visualize local energy minima corresponding to various flap states. One of the biggest differences found in these free energy maps is that there are more minima at larger distances than observed in the other 2, especially along the IA distance. There are minima that reach as high as 25 Å. For this particular system there are no distinguishable patterns other than the fact that there are minima at much larger distances and at the original distances found in the other simulations.

5.2.3 Solvent Accessible Surface Area

The total solvent accessible surface area (SASA) for urease in 150 mM aqueous NH_4Cl increases 7.5 percent from the crystal structure compared to the 6.3 percent increase observed in the aqueous HP urease simulation (Table (5-1)). At 100 ns, the total SASA in the ammonium chloride run increased 23.7 percent from the crystal structure, while it increased 20.1 percent in the aqueous run. Both increases are comparatively large placed beside the 16.7 percent increase in SASA over the first 100 ns of the MD simulation of urease in 10.5 M aqueous urea. While the original aqueous run only increased in SASA by about 4.8 percent over the remainder of the simulation, the ammonium chloride run saw a nearly 7 percent increase in the final 100 ns, placing it at a total of 32.2 percent above that of the X-ray.

The equilibrated structure possessed a 23.5 percent increase in main chain SA exposure compared to the X-ray structure, while the side chain exposed SA only increased 5.1 percent (Table (5-2)). The ratio of exposed side chain to main chain is about 7 to 1. Side chain exposure

slowly increased up to 25.8 percent from X-ray. Following 100 ns of MD, the main chain

exposed SA increased dramatically to a total of 42.5 % over X-ray structure. This exposed area

continued to climb to 45.1 percent after 400 ns, and over the final 100 ns, there was a further 31

percent increase with respect to X-ray. At the conclusion of the simulation, the exposed percent

change reached 76 percent, resulting in an exposed SA increase from 30105.2 Å^2 to 52999.5 Å^2,

leaving the resulting ratio at 5 to 1.

Table 5-1. Total solvent accessible surface area for the X-ray structure of *H. pylori* urease and
evaluated at points along the MD simulations on both aqueous urease and urease in
aqueous 150 mM NH$_4$Cl.

Structure	*H. pylori* urease		*H. pylori* urease + NH$_4$Cl	
	Total SASA (Å^2)	% Increase[a]	Total SASA (Å^2)	% Increase[a]
X-ray	234,281.9	---	234281.9	---
Equilibrium	248,967.6	6.3	251740.4	7.5
100 ns	281,311.5	20.1	289697.9	23.7
200 ns	287,003.8	22.5	291354.3	24.4
300 ns	287,774.2	22.8	292701.0	24.9
400 ns	291,730.1	24.5	293978.4	25.5
500 ns	292,556.0	24.9	309812.7	32.2

[a]Percent increase in SASA referenced to the X-ray structure

Table 5-2. Side chain and main chain accessible surface areas for *H. pylori* urease in 150 mM
NH$_4$Cl.

Structure	Side Chain		Main Chain	
	SASA (Å^2)	% Increase[a]	SASA (Å^2)	% Increase
X-ray	204,176.9	---	30,105.2	---
Equilibrium	214,575.2	5.1	37,166.0	23.5
100 ns	246,785.7	20.9	42,912.1	42.5
200 ns	248,170.9	21.5	43,182.8	43.4
300 ns	249,425.2	22.2	43,276.6	43.8
400 ns	250,310.8	22.6	43,668.2	45.1
500 ns	256,814.3	25.8	52,999.5	76.0

[a]Percent increase in SASA referenced to the X-ray structure

The exposed non-polar surface area increases 1.3 percent while the polar increases 16.9

percent relative to the crystal structure upon equilibration (Table (5-3)). The exposed non-polar

surface area increased by 13.6 percent following 100 ns of MD and gradually increased by a total

15.6 percent at 400 ns. The SASA in polar residues increased to 39.1 percent larger than the

crystal structure after 100 ns and to 40.7 percent higher following 400 ns. Both the exposed non-polar and polar surface area increased sharply over the final 100 ns with the former being 21.6 percent higher than the X-ray structure and the latter 48.6 percent greater at the simulation's end. These data support the conclusion that the simulation conditions are not in fact leading to denaturing of the enzyme as in our previous work.

Table 5-3. Polar and non-polar accessible surface areas for *H. pylori* urease in 150 mM NH$_4$Cl.

Structure	Non-polar		All polar	
	SASA (Å^2)	% Increase[a]	SASA (Å^2)	% Increase[a]
X-ray	141,933.6	---	92,348.2	---
Equilibrium	143,769.5	1.3	107,971.2	16.9
100 ns	161,221.8	13.6	128,476.5	39.1
200 ns	162,025.7	14.2	129,327.6	40.0
300 ns	163,461.2	15.2	129,241.2	39.9
400 ns	164,085.6	15.6	129,893.4	40.7
500 ns	172,585.8	21.6	137,227.3	48.6

[a]Percent increase in SASA referenced to the X-ray structure

5.2.4 The Hollow

The internal hollow of *H. pylori* urease has been hypothesized to act as a reservoir of ammonia/ammonium ions, and has a volume of ~145,156 Å.[6] Limited motion is observed in the hollow-defining residues over the course of the simulation, which suggests there is a relatively constant volume in the internal cavity. We chose to track the behavior of ammonium ions in order to ascertain whether this cation accumulates in the hollow as hypothesized in our previous MD work on aqueous urease, where sodium ions were tracked.[46] In characterizing the residues outlining the hollow, we observed that the first residues that we encounter as we move in from the center of the hollow were glutamates and, more specifically, the GLUα505 (GLU743). These GLUα505 (GLU743) residues were largely stationary, with the exception of the GLUα505 (GLU743) of dimer 11, which featured an RMSF of 2.0. Snapshots of the structure at 100 ns intervals, with the origin set as the protein center of mass, were used to generate PDB structures.

These structures were used to track the number of ammonium ions present in 1 Å increments from the origin. While the net charge of all residues within 60 Å of the protein center of mass is -16 (36.4% of total protein charge), the innermost residues of the protein (within 22 Å of the center of mass) are the twelve GLUα505 (GLU743) residues that place a net -12 charge nearest the hollow center. Extension of the radius from the center of mass to 23 Å introduces the first positively charged residue, a single lysine, as well as three additional glutamates for a net charge of -14. This should permit the hollow to accommodate and stabilize a positive charge. If this is indeed the case, the ammonium ion distribution over the course of the MD run should reflect this fact, which is the case here.

As seen in Table (5-4), we are able to observe that the initial positions of the ammonium ions were not in the hollow as seen with the equilibration column of data. As time progresses, at 100 ns we observe a total of 9 ammonium ions within 20 Å of the center, with most of them located 16 to 20 Å from the center, in the sphere of influence of the GLUα505 (GLU743) residue's charge. Then, as we progress to 200 ns, it decreases to 5 ammonium ions, which indicates that some of the ammoniums have left the hollow. At 300 ns, the number goes up to 7, decreasing to 6 at 400 ns and then again increasing to 11 ammonium ions after 500 ns. This shows that there is mobility of the ammonium ions and that the hollow acts like a reservoir for ammonium ions prior to their exit into the surrounding solution.

Table 5-4. Ammonium ion count in hollow of HP urease at 100 ns intervals as a function of distance from center of mass.

Radius (Å)	Equilibrium	100 ns	200 ns	300 ns	400 ns	500 ns
5	0	0	0	1	0	0
10	0	0	0	0	1	0
15	0	3	0	2	4	4
20	0	6	5	4	1	7

5.2.5 First Solvation Shell

The number of ammonium ions, chloride ions and water molecules that resided within 2 Å of the protein were considered to comprise the first shell of solvation. The shell composition was obtained using PDB files generated from the trajectory following equilibration and at the after 100, 200, 300, 400, and 500 ns of the simulation and the results are summarized in Table (5-5).

Table 5-5. Ion and water molecule count at 100 ns intervals in first solvation shell.

Point	AMM	Cl^-	H_2O
Equil.	84	1	5995
100 ns	86	2	7913
200 ns	92	4	8157
300 ns	86	1	8223
400 ns	76	4	8260
500 ns	77	4	8302

This MD simulation contained 143 ammonium ions, 99 chloride ions and 53,093 water molecules. This corresponds to a 371:1 ratio of water molecules to ammonium ions in the bulk solvent, with a corresponding 536:1 ratio of water to chloride ions. A total of 84 ammonium ions were observed in the first solvation shell following structure minimization and equilibration. This increased to 92 ions following 200 ns then decreased to 76 ions at 400 ns. At the conclusion of the simulation the first shell contained 77 ammonium ions, representing 53.8 percent of the total ammonium ion count, and represents an over 5-fold decrease in the water to ammonium ion ratio observed in the bulk solvent at 71:1. A comparatively miniscule number of chloride ions were observed in the first solvation shell. After equilibration we observe 1 chloride ion out of the total 99 chloride ions. From there, the quantity never increases higher than 4 chloride ions at 200, 400 and 500 ns while dipping to 1 ion at 300 ns. At each snapshot of the simulation this corresponds to a significant *increase* compared to the water to chloride ion ratio observed in the bulk solvent. The water composition of the first solvation shell following equilibration was 5,995

molecules followed by a drastic increase to 7,913 water molecules after 100 ns of MD. Over the final 400 ns of the simulation, there is a slight increase in water content of the first solvation shell, reaching 8302 water molecules at 500 ns.

5.2.6 Root Mean Square Fluctuation

We analyzed the high RMSF regions in the ammonium chloride run. The average RMSF value was 0.88 Å with a standard deviation of 0.42 Å and we choose to define high RMSF regions as the average plus one standard deviation (1.3 Å). There is a clearer contrast between the mobile and immobile parts of the protein. The mobile areas are very distinct and consist of residues β50-63 (50-63), β100-106 (100-106), β113-130 (113-130), β166-183 (166-183), and β224-238 (224-238). Just as observed in the urea run of HP urease, the beta chain, which envelops the protein on the surface, is very mobile and enables the shuttling of the urea to the active site (Figure (5-3)). As for the highly mobile regions on the alpha chains, there are four distance regions. One such region is α295-302 (533-540), which is the region within the hollow near the entrance to the hollow. α315-341 (553-579) is the legendary active site flap. Interestingly enough, in all 12 dimers, residues α387-396 (625-634) are highly mobile. This has been hypothesized before as being part of an exhaust flap consisting of two alpha helices and a loop, in which residues α387-396 (625-634) make up the loop. This could explain their involvement in the pathway by which ammonium and carbamate are expelled into the hollow. Finally we have residues α532-553 (770-791), which are the residues that are part of the ancillary flap on the surface of the protein. The purpose of this flap is not yet explained.
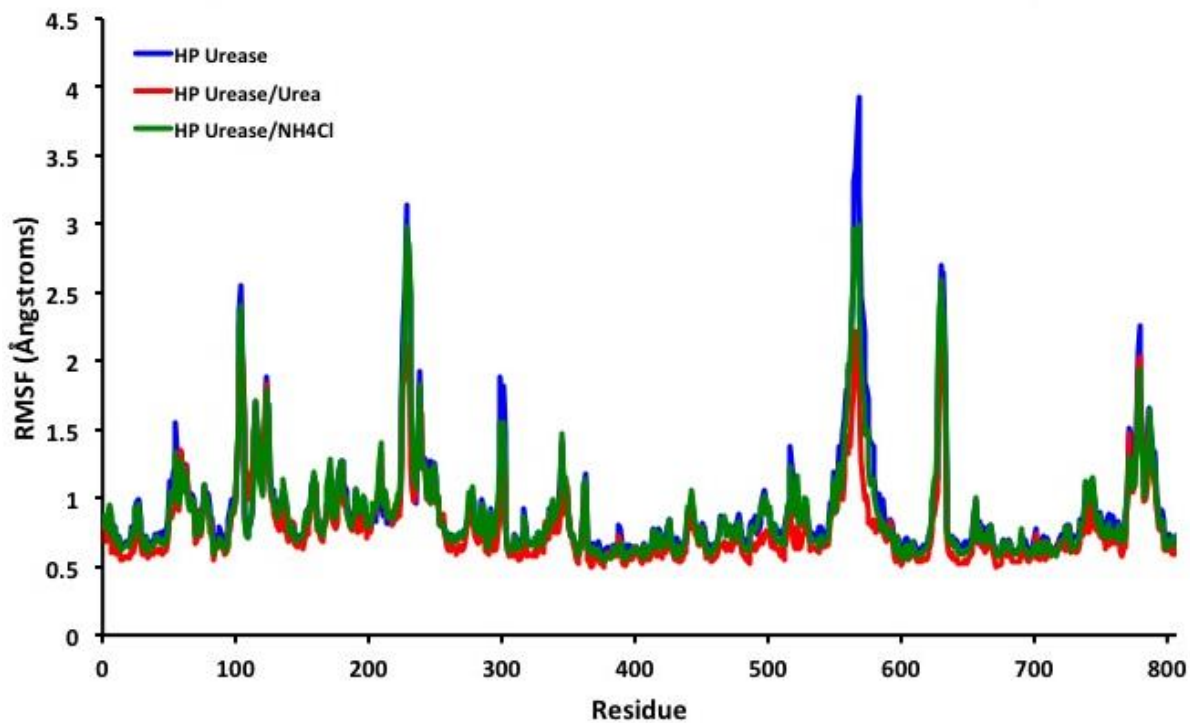
Figure 5-3. Overlay of average RMSF for aqueous HP urease (blue), HP urease in 10.5 M aqueous urea (red) and HP urease in 150 mM NH₄Cl (green).
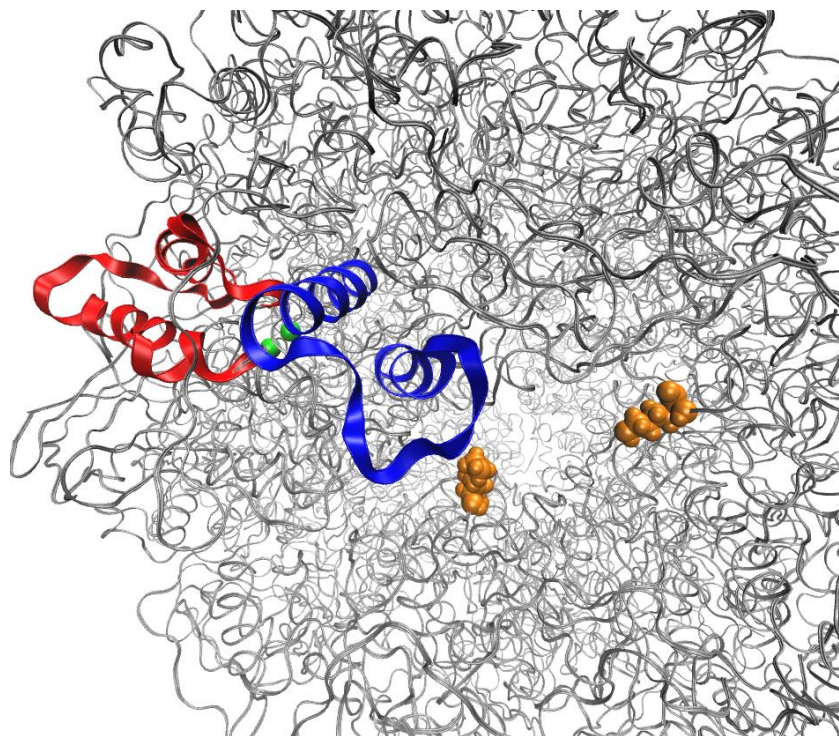
Figure 5-4. Relative positions of active site covering flap (red), exhaust flap (blue) and LYSα109 (LYS347) residues. Ni2+ ions are shown for reference as green spheres.

5.2.7 Ammonium Ion Interactions

In order to postulate plausible pathways by which HP urease expels ammonium ions, we conducted analyses of these ions as they interact with different components of the enzyme. Specifically, we observed interactions of the ions with the active sites, postulated exhaust flaps, nickel ion pairs, high-RMSF regions, and specific residues of interest, such as LYSα109 (LYS347). In doing so, we sought to determine a potential conduit by which ammonium ions are expelled into the surrounding environment by the urease.

The number of ammonium ions that interacted with high RMSF region α533-553 (771-791) was collected for each of the twelve dimers. We looked at all NH_4^+ ions that were within 6 Å of any given atom within that region, and we found several interesting results. Most notably, of the 456 interaction instances counted within the trajectory, 240 were found to occur with GLUα541 (GLU779), or 52.6% of the total interactions. GLUα537 (GLU775) was involved in a distant second number of interactions, with 74 interactions (16.2%); beyond that, ASPα548 (ASP786) featured 60 interactions (13.2%). These three residues, all negatively charged, accounted for the vast majority of all interactions. Interestingly, despite the large quantity of GLUα541 (GLU779) interactions, the first α533-553 (771-791) dimer contained only 1 observed ammonium interacting with this residue, an unusual anomaly.

Another interesting result was found by observing the actual distances of the counted ammonium ions. We immediately noticed that most of these interactions featured distances indicative of tight binding; in dimer 1, for example, interactions with GLUα541 (GLU779) OR GLUα537 (GLU775) were within the 1-3 Å range. By reducing the distance cutoff to 4 Å, 97.6% of the ion-residue interactions were kept, indicating tight binding throughout nearly all interactions in the simulation. Most of these interactions, however, existed for 1 ns; few multi-frame interactions existed for more than 5 ns.

We also analyzed the number of ammonium ions that interacted with the active site flap, quantifying the number of ammoniums that appeared within 6 Å of the active site flap. The obtained quantities ranged between 12 and 40 ammonium ions, depending on the dimer observed. Most of these ammonium ions were located near the alpha helices. We also narrowed down the search range to look at ammonium ions that were within 4 Å; in doing so, we obtained a range from 2 to 19 ions per dimer within that distance cutoff.

We quantified the number of interactions to see which amino acid within the flap interacted most often with ammonium ions. ASPα336 (ASP574) featured the most interactions with 92, followed by (SERα337) SER575 with only 27, and then ARGα340 (ARG578) with 24 interactions. The rest of the observed trends have been documented in Table (5-6).

Ammonium ions usually only remained for 1-2 ns of interaction time. However, if an ammonium ion interacts with the alpha helix, it tends to remain for between 5 to 15 ns; examples include the majority of interactions with ASPα291 (ASP574) (ammonium 9986 remained during the 115-138 ns frames of the simulation, and ammonium 9989 remained during the 205-223 ns frames, for instance), ASPα336 (ASP1386) (at least 5 ammonium ions remained within 4 Å for at least 5 ns of simulation time), and ASPα336 (ASP7882) (at least 6 ammonium ions remained within 4 Å for at least 5 ns of simulation time).

We also analyzed the surroundings of the nickel pairs, searching for ammonium ions that were within 10 Å of each pair. This search identified 15 total ammoniums within 10 Å of any nickel pair, out of the 52 total ammoniums studied. There were 34 separate ammoniums found between 18 and 30 Å from any nickel pair, some of which were in the direction of nearby LYSα109 (LYS347) residues and exhaust flaps; examples include ammonium ions 9783 (near

the exhaust flap α378-413 [616-651]), 9963 (near LYSα109[LYS347] dimer 4-9 [LYS2783-LYS6843]), and 9833 (near LYSα109 (LYS347) dimer 2-7 [LYS1159-LYS5219]).

In this particular study, we were able to form a hypothesis outlining a viable pathway through which ammonium is deprotonated into ammonia and then expelled into the surrounding environment. In our aqueous urease MD simulation, we observe a highly mobile flap-like structure found in the back of the active site.[46] This flap-like structure has shifted closer to the surface in this simulation, with the loop region positioned slightly over the ovoid holes that reside on the two-fold symmetry axes. Upon further observation, we also noticed the presence of pairs of deprotonated lysine residues, LYSα109 (LYS347), which also frame the ovoid hole. The exhaust flaps sit across each other on one axis, and the LYS residues are positioned across each other on the perpendicular axis. As the simulation continues, we observe the side chains of this lysine residue sweeping back and forth, covering nearly half of the distance of the ovoid hole entrance along the transverse diameter. We also observe some of the ammonium ions travelling from the direction of the nickels, interacting with LYSα109 (LYS347) for approximately 5 to 10 ns, then proceeding to interact with the exhaust flap for a 1 to 2 ns duration before moving and exiting into the bulk solvent. This was observed multiples times; specific examples include ammonium ions 9789 (360-380 ns) and 9843 (460-480 ns) travelling through the hole framed by LYS1159 and LYS5219, as well as the exhaust flaps.

With regard to the exhaust flaps, we quantified ammonium ion interactions by quantifying ions that came within 3 Å of any flap residue. On average, we observed approximately 20 to 25 ammonium ions interacting with the flaps over the course of the simulation. By quantifying these interactions, we determined GLUα122 (GLU360), which is located on the tip of the flap, to be most involved with ammonium ions. In all, there were 70 total

interactions across all exhaust flaps. The second most common residue was GLUα393

(GLU631), followed by LYSα391 (LYS629). Table (5-6) displays the quantity of ammonium

ion/exhaust flap interactions by dimer. LYSα109 (LYS347), which is deprotonated, is within 3 Å

of 10 to 16 ammonium ions interacting throughout the simulation for each LYSα109 (LYS347).

Table 5-6. Number of ammonium ions within 3 Å of exhaust flap residues.

Residue	Dimer												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
GLU630	6	3	3	12	2	3	3	6	11	8	6	7	70
GLU631	2	0	1	3	5	3	0	0	5	6	1	0	26
LYS629	1	1	1	1	1	1	0	3	3	0	0	0	12
LYS632	6	2	1	1	1	1	2	1	0	2	0	0	17
ASP634	3	3	7	1	2	2	6	6	5	7	2	4	48
GLU624	2	3	3	2	0	1	1	0	1	1	2	3	19
GLY633	2	2	0	4	3	2	1	0	2	2	7	1	26

5.2.7 Lysine 347

One of the actions we took to understand and study the movement of the ovoid holes and

the possible mechanism by which ammonium ions are deprotonated and then expelled into the

surrounding solvent was to observe the behavior of the LYSα109 (LYS347) residues that framed

the holes. Specifically, we measured the distance between the side chain nitrogen found on one

of the LYSα109 (LYS347) and its opposing partner's nitrogen and studied the distance

differences. The literature contains examples of environment influencing the pKa of the lysine

side, and our calculations using H++ suggest that the structure of the enzyme and the

environment of the lysine residue can adjust the pKa from 6.8 to 11.8, making this discussion

reasonable.

From a visual perspective, we observed that the sidechains of LYSα109 (LYS347), as

previously stated, appeared to have a wiper-like motion, sweeping back and forth, spanning

around half of the largest diameter of the ovoid holes at particular points in time (Figure (5-5)).

There are 6 pairs of these residues, one for each ovoid hole. Two of these pairs (dimer pairs 1-8,

11-12) have a difference in minimum-maximum distance up to 20 Å, while the other four pairs (pairs 2-7, 3-10, 4-9, 5-6) have a difference in distance between 10 to 13 Å. For the second set of pairs, not as much movement was observed.
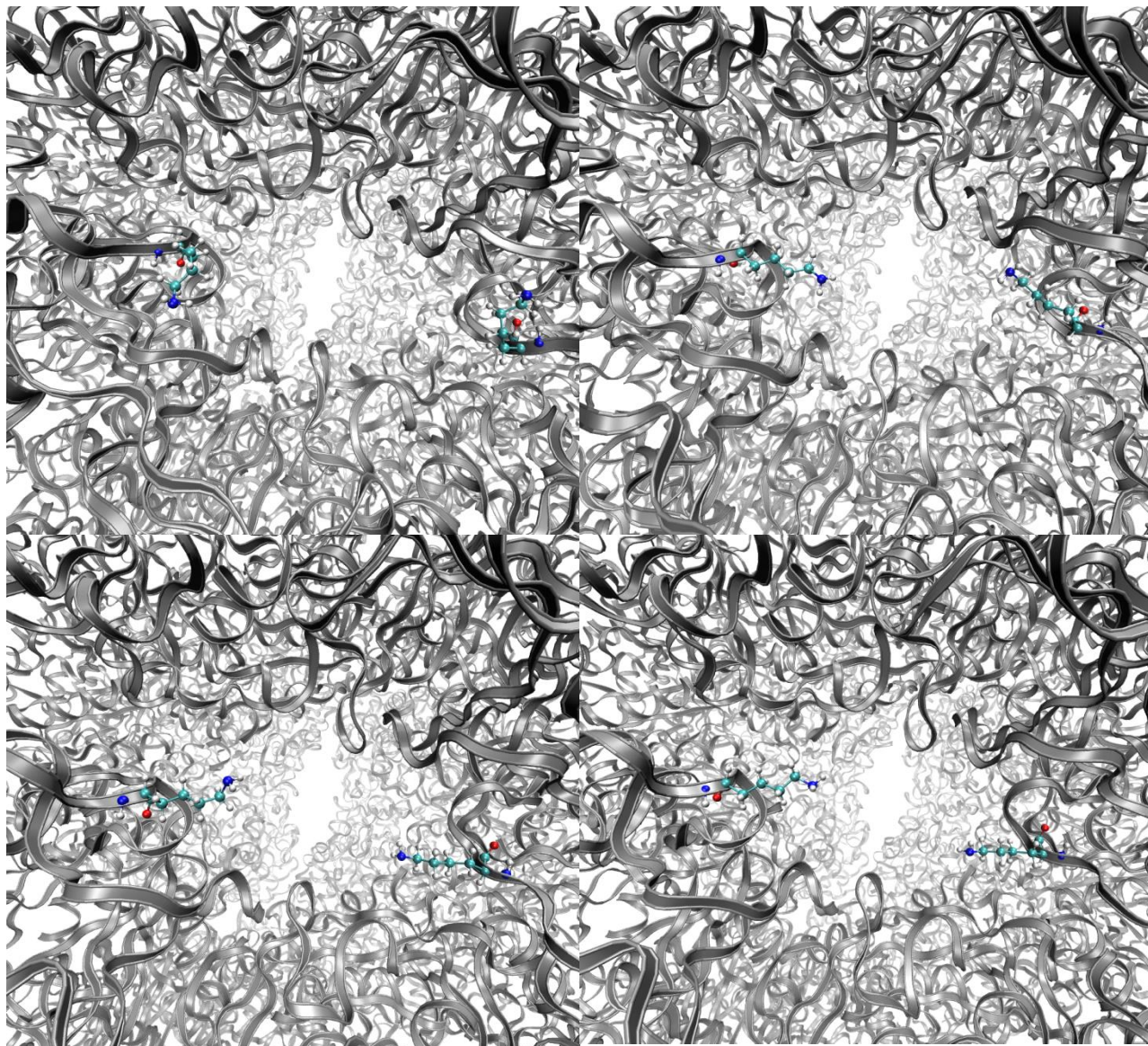
Figure 5-5. Sweeping action of LYSα109 (LYS347) side chain.

We attempted to find a correlation between the increased motion of the LYS side chains and the width of the open active site flap. However, we were unable to determine a pattern, likely due to the fact we did not have enough simulation time to be able to conclude a correlation.

One interesting pattern we found, however, was that, for each pair, the distance

measurements seemed to fluctuate in a sinusoidal manner. This is coupled with increased

sinusoidal amplitude as time increases; this explains the observation that the largest difference in

distance typically occurs toward the end the simulation, as expressed most clearly in the plots of

dimer 2-7 (minimum at 323 ns, maximum at simulation's end) and 1-8 (minimum at 359 ns,

maximum at 445 ns). Only the separation between these lysine residues on dimers 5 and 6

featured little discernible sinusoidal behavior, with a maximum that occurred toward the

beginning of the simulation. Table (5-7) lists the dimers of the paired residues, their respective

minimum and maximum separations, and the differences, which reflect sidechain area swept.

The plots of these dimer pair separations are also included.

Table 5-7. Minimum and maximum separation between LYSα109 (LYS347) side chain nitrogen
atoms (Å).

Dimer Pair (Residue #)	Maximum (Å)	Time (ns)	Minimum (Å)	Time (ns)	Difference (Å)
1-8 (347-6031)	32.84	445	12.76	359	20.08
2-7 (1159-5219)	26.13	323	15.54	501	10.59
3-10 (1971-7655)	27.54	321	13.92	251	13.62
4-9 (2783-6843)	25.18	424	13.93	291	11.25
5-6 (3595-4407)	32.84	445	12.75	359	20.09
11-12 (8467-9279)	33.32	375	14.17	191	19.15

5.2.8 Discussion of Pathway

We tracked the movement of a few ammonium ions to observe the path by which they

exited the protein. One of the ions we observed, ammonium 9887 (Figure (5-6)), remained in the

hollow for the first 48 ns of simulation time. Then, at 49 ns, it exited the protein via the exhaust

flap loop, interacting with the loop region within the 50-54 ns frames. At this juncture, the ion

then travelled to the deprotonated LYSα109 (LYS347) residue – located further away from the exhaust flap – and interacted with the residue for the next 22 ns of the simulation. The ammonium ion then travels into the bulk solvent at the 73 ns frame. Ammonium 9843, the second ion we observed, follows a similar exit pathway; similar images containing these observations can be found in Figure (5-7).
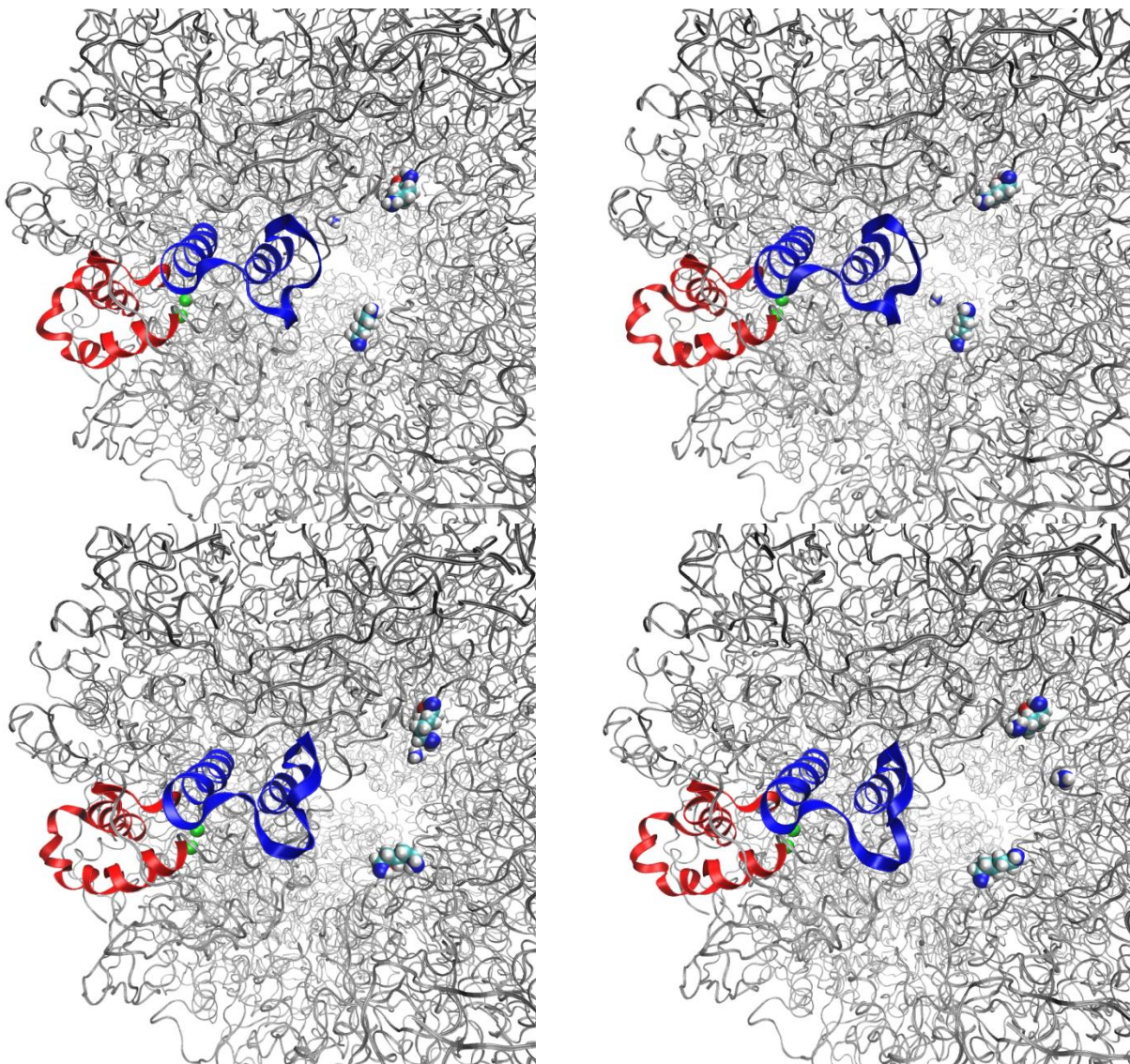
Figure 5-6. Graphical representation of ammonium ion 9887 exiting ovoid hole, interacting with exhaust flap and being swept by LYSα109 (LYS347) before departing.
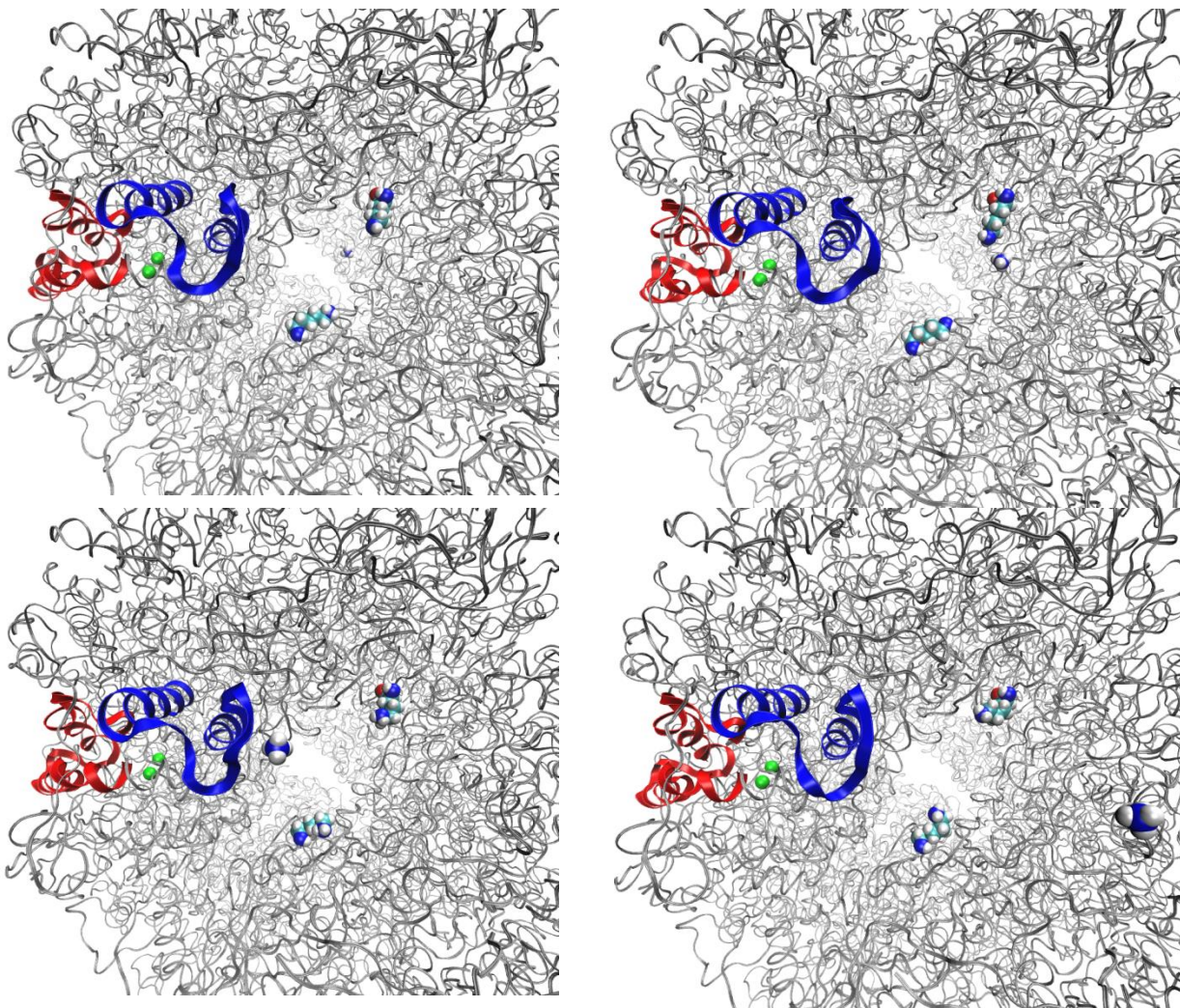
Figure 5-7. Graphical representation of ammonium ion exiting ovoid hole, interacting with exhaust flap and being swept by LYSα109 (LYS347) before departing.

5.2.9 Trajectory Comparison

We compared the differences in structure between the original run of HP urease to that of

the structure of HP urease simulated in 150 mM aqueous ammonium chloride. One of the most

important differences is in the motion of the side chain of LYSα109 (LYS347). In the original

run the side chain did not sweep back and forth or extend as much, resulting in decreased surface

area coverage by the side chain. In the run with ammonium chloride, however, there is a clear

extension of the side chain and a wide sweeping movement, resulting in a larger coverage of the

163

ovoid hole entrance. This is also observed in the plots found in xxx where the distance of the 2 nitrogen atoms on opposing LYSα109 (LYS347) residues were measured and compared.

Another key structural difference between the two runs is in the exhaust flap. In the original run the loop region of the flap and some of the alpha helix regions are exposed, while in the run with the ammonium chloride the entire flap is exposed and protrudes further from the remainder of the bulk exterior. It also exhibited more movement then that of the original run.

In both simulations, three of the four alpha helix-defined holes, each lying along the three-fold symmetry axes, have very defined alpha helices. In the original run the fourth hole has a partially unraveled alpha helix, which occurred during the minimization and equilibration stages. Similar behavior was exhibited in the simulation with ammonium chloride, although the alpha helixes were more unraveled. This also occurred in the equilibration and minimization stages. The remainder of the protein exhibited similar behavior for the two simulations.

5.3 Conclusions

In this study, we simulated the behavior of HP urease in 150 mM ammonium chloride, observing the ensuing effects. This was done with the intent of uncovering a possible pathway or mechanism explaining the means by which ammonium ions are expelled from the active site. Additionally, we sought to determine if there existed a viable method of ammonium ion deprotonation to allow the potent base ammonia to enter and neutralize the surrounding environment. We also studied the exhaust flap found by the high RMSF region in the first paper to determine its function – should it possess one – in this simulation.

First, we observed ammonium ions moving within and subsequently exiting the protein's internal hollow region, with the exit into the surrounding environment occurring primarily via the ovoid holes framed by the exhaust flaps. We also noticed that, along the pathway of the ammonium ions in the trajectory, we can observe the side chain of the deprotonated LYSα109

(LYS347) sweeping back and forth, more so than in any other simulation. This sweeping motion allows the LYS residues to span half of the diameter of the ovoid holes. This movement, coupled with the observed interactions of the residue with ammonium ions exiting via the hole, could suggest a possible mode of action by which ammonium ions are deprotonated to form ammonia. The exhaust flap, meanwhile, was observed to have significant movement with regard to the loop region, and is largely exposed outward on the protein's surface, more so than in any other simulation performed. Mutation of the *H. pylori* urease enzyme to remove the LYSα109 (LYS347) residue and replacement with a neutral or acidic residue and subsequent analysis of the effect of that urease on acidic solutions would provide some insight into the importance of this ovoid hole-framing residue.

CHAPTER 6
DOCKING STUDIES PERFORMED ON *KLEBSIELLA AEROGENES* UREASE

The following chapter has been adapted from *Reduction of Urease Activity by Interaction with the Flap Covering the Active Site,* published in the Journal of Chemical Information and Modeling in 2015 (http://pubs.acs.org/doi/abs/10.1021/ci500562t)[77], with the permission of the American Chemical Society.

6.1 Introduction

To identify potential inhibitors to test experimentally, I conducted docking studies upon a large volume of ligands from the ZINC ligand database[78], split into multiple subsets based on origin or compound type. Specific subsets included the Biogenic Compounds, Drug Database, Natural Product, ZINC In Man, and Lead-Like subsets. Additionally, I selected multiple individual ligands that had been previously shown to possess good binding affinities for *H. pylori* urease in the literature.[12] These individual ligands, discussed in greater detail further into this chapter, were compiled and treated as a separate subset for these docking studies.

After selecting the ZINC library subsets, we conducted our docking studies via Glide[45], utilizing standard docking precision. We evaluated these ligands by docking them to the active site of *Klebsiella aerogenes* urease, an enzyme whose active site shares an identical nickel coordination site to that of *H. pylori* urease.[79] The exhaustive quantity of research discussing the mechanistic and functional details of *K. aerogenes* urease in the literature makes the isozyme an even more suitable model for these docking studies. The *K. aerogenes* urease active site, much like that of *H. pylori* urease, features two Ni^{2+} ions coordinated by several species; the first ion is coordinated by two HIS side chains and a water molecule, whereas the second is coordinated by an ASP residue in addition to the HIS residues and water molecule. Additionally, the ions are bridged by a hydroxide, and a carbamylated LYS residue.[6, 12, 79b, 80] The urease active site also

features a flexible loop covering (THR308 – ARG336 in the case of *K. aerogenes*)[53] that features a residue – CYS319 – that may play a catalytic role.[79b] For these docking studies, the flap was configured into a "wide-open" state, exposing the active site for ligand binding. After performing these studies with this active site, we ranked the compounds by docking score and selected the best 10 for further analysis; the ligands selected in each database will be discussed in the following sections.

6.2 ZINC Biogenic Compounds
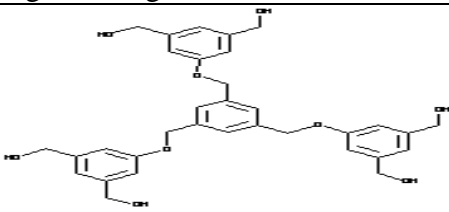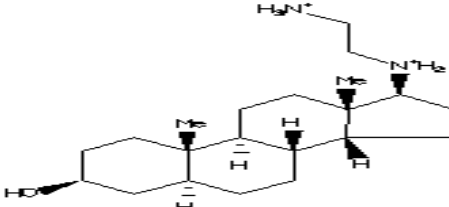
The ZINC Biogenic Compounds (ZBC) subset consists of all commercially available primary metabolites and secondary natural products. It has been constructed through data acquired from public domain sources and information supplied by vendors. The ZBC subset contains 180,313 entries, the most of any of the subsets we chose to analyze. This includes stereochemical guesses for each ambiguous compound, up to a maximum cap of four guesses per compound. Table (6-1) catalogs the compounds from the ZBC subset that produced the highest scores in docking studies, while Figure (6-1) depicts the specific interactions of these ligands with active site residues. The top-scoring ligand, with a docking score of -9.711, features 4 hydrogen bonds to the protein backbone, 5 hydrogen bonds with sidechains, one π-cation interaction with ARG 2082, and is within 4 A of both the active site and the flap residues. The 2nd-ranked ligand, ID 79209932, features 3 hydrogen bonds to both sidechains and backbones (one of each involving a flap residue) and π-cation interaction with ARG 2082, which is simultaneously involved in a side-chain hydrogen bond interaction. The 3rd-ranked ligand, with a score of -9.242, features 3 hydrogen bonds with the backbone as well as a π-cation interaction; however, no bonding interactions involving the sidechain are depicted. The 4th-ranked ligand, with a score of -9.193, features 3 backbone hydrogen bonds – one uniquely involving the LEU 2062 flap residue – and 2 sidechain hydrogen bonds, but no π-cation interactions. The 5th-ranked

ligand, with a score of -9.174, features 3 hydrogen bonds with both the sidechain and backbone, as well as a π-cation interaction with ARG2082, itself again involved in a side chain interaction. The 6th-ranked ligand, with a score of -8.956, features 3 backbone hydrogen interactions and, notably, 5 sidechain hydrogen interactions, 2 of which are involved with ARG2082, itself involved in a π-cation interaction. Interestingly, this ligand also features 2 πstacking interactions with HIS 1912, both involving a single naphthol. The 7th-ranked ligand, with a score of -8.92, features 3 side-chain and 4 backbone hydrogen bonds, but no πinteractions or hydrogen bonds with any flap or active-site residues. The 8th-ranked ligand, with a score of -8.877 features 4 side-chain interactions – one of which involves ARG2082, itself involved in π-cation interactions – and 3 backbone hydrogen interactions, as well as a single π-π interaction between HIS1912 and a naphthol group. The 9th-ranked ligand, with a score of -8.757, involves 5 side-chain and 3 backbone hydrogen interactions – none occurring with an active site or flap residue – and a single π-cation interaction with an otherwise-uninvolved ARG2082. Finally, the 10th-ranked ligand – the only one to lack any form of aromaticity – features a score of -8.743 and contains a single side-chain hydrogen interaction and two backbone hydrogen interactions, one occurring with the backbone of PHE 2078. Overall, the ligands commonly featured a fairly even spread of bonding activity across the ligand, a general presence of aromatic groups, and π interactions with either ARG2082 or HIS1912, Within the group, while the quantity of hydrogen interactions and molecular weight varied independently from docking score, higher-scoring ligands typically featured interactions with MET 2061.

Table 6-1. Top 10 ZBC Ligands by Docking Score

Rank	Name	ZINC ID	MW	Score	Ligand Image
1	[(2S,3R,4R,5R,6S)-3,5-dihydroxy-2-[(1S,2R)-3-hydroxy-2-[(1S)-1-hydroxy-2-[(E)-(4-hydroxy-1-naphthyl)	79209931	692.718	-9.711	
2	[(2S,3R,4S,5R,6S)-3,5-dihydroxy-2-[(1S,2R)-3-hydroxy-2-[(1S)-1-hydroxy-2-[(E)-(4-hydroxy-1-naphthyl)	79209932	692.718	-9.496	
3	[2-[(1R)-1-amino-2-(4-hydroxyphenyl)ethyl]-6-ethyl-3-(4-methylthiazol-2-yl)-4-oxo-chromen-7-yl]	70664754	585.682	-9.242	
4	[(2S,3R,4S,5R,6S)-3,5-dihydroxy-2-[(1S,2R)-3-hydroxy-2-[(1S)-1-hydroxy-2-[(E)-(2-hydroxy-1-naphthyl)	79204475	694.734	-9.193	
5	[(2S,3R,4S,5S,6S)-3,5-dihydroxy-2-[(1S,2R)-3-hydroxy-2-[(1S)-1-hydroxy-2-[(E)-(4-hydroxy-1-naphthyl)	79209940	692.718	-9.174	
6	[(2S,3R,4R,5R,6S)-2-[(1S,2R)-2-[(1S)-2-[(E)-(2,4-dihydroxy-1-naphthyl)methylene amino]-1-hydroxy-etho	79210084	724.716	-8.956	
7	[(2S,3R,4R,5R,6S)-2-[(1S,2R)-2-[(1S)-2-[(E)-(2,4-dihydroxy-1-naphthyl)methylene amino]-1-hydroxy-etho	79209936	692.718	-8.92	
8	[(2S,3R,4R,5S,6S)-2-[(1S,2R)-2-[(1S)-2-[(E)-(2,4-dihydroxy-1-naphthyl)methylene amino]-1-hydroxy-etho…	79210091	724.716	-8.877	

Table 6-1. Continued

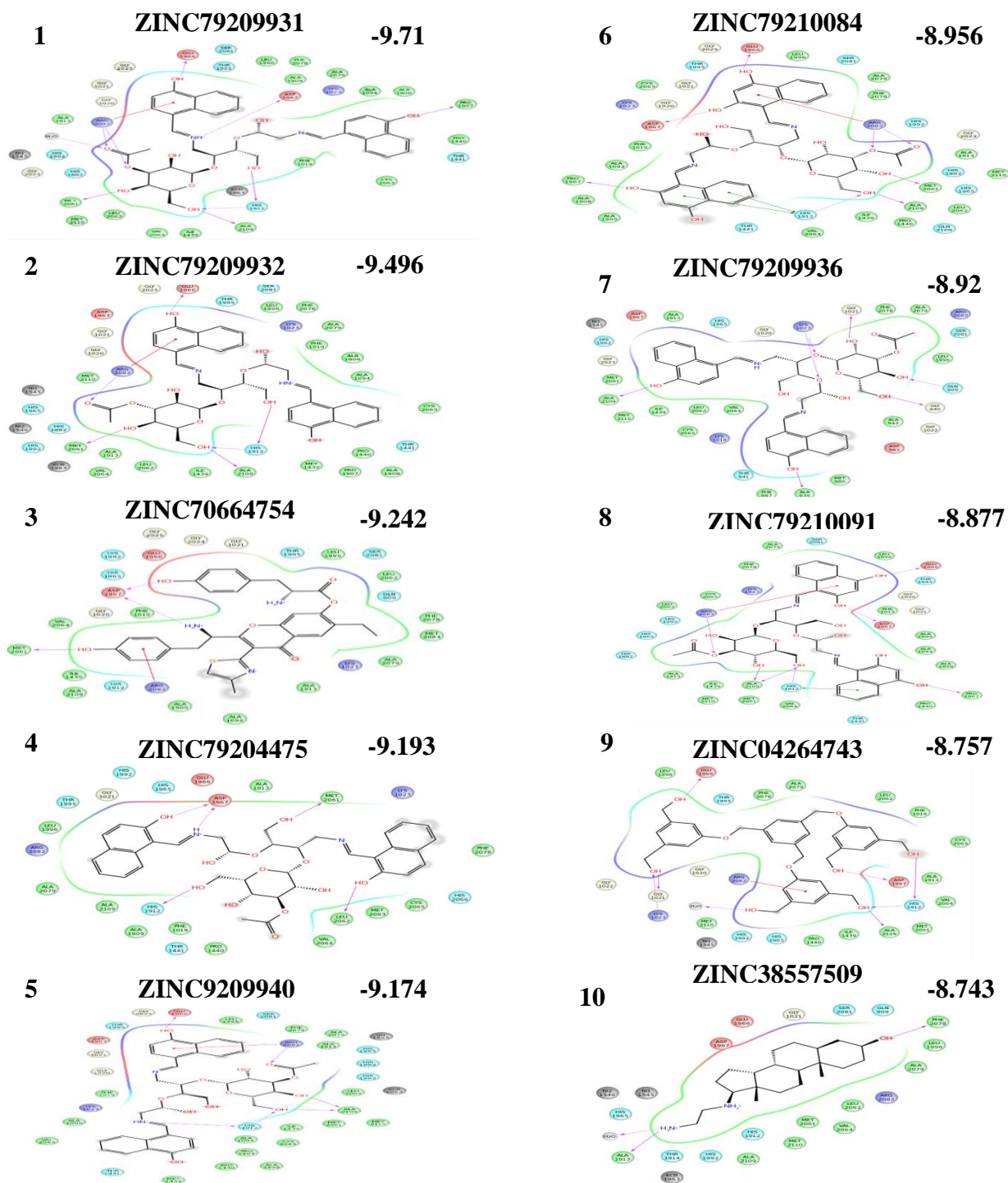| Rank	Name	ZINC ID	MW	Score	Ligand Image
9	[3-[[3,5-bis[[3,5-bis(hydroxymethyl)phenoxy]methyl]phenyl]methoxy]-5-(hydroxymethyl)phenyl]methanol	04264743	576.642	-8.757	
10	(3S,5S,8R,9S,10S,13S,14R,17S)-17-(2-aminoethylamino)-10,13-dimethyl-2,3,4,5,6,7,8,9,11,12,14,15,16,1…	38557509	336.564	-8.743	

Figure 6-1. Top ZBC ligands interacting with the residues of the active site. Species are colored by type or name: HIS is colored in light blue, GLY in white, hydrophobic residues green, charged residues red (-) or violet (+), and all other species in gray. Interactions with amino acids are shown as solid (backbone) or dashed (sidechain) lines.

6.3 ZINC Drug Database Subset

The ZINC Drug Database (ZDD) consists of all compounds that have been approved for use as drugs by the FDA and any equivalent entity. This subset was built entirely through publicly accessible information, and only contains molecules that can be purchased as pure compounds. These criteria return a subset size of 2,924 ligands, considerably less than that of the ZBC subset. As in the ZBC subset, all stereochemical guesses of a compound are included in the database, up to a maximum of four per molecule. Table (6-2) contains the 10 top-scoring ligands observed after docking studies were performed on the ZDD subset (note that interaction images for ligands 3-5 were missing or not available; these ligands have been skipped as a result), while Figure (6-2) displays the interactions between active site residues and the ligands. The top-ranking ligand in this set, the adrenergic antagonist labetalol, features a docking score of -8.1 and was within 4 A of both the active site and flap residues, and featured both 3 backbone hydrogen bonds and 2 sidechain hydrogen bonds, one of which involved ARG 2082. No form of π interaction was observed. The 2nd-ranking ligand, pentamidine isethionate, is a symmetrical ligand with a score of -8.077 that possesses 4 side-chain hydrogen interactions, but no backbone interactions. Furthermore, it features a π-cation interaction with ARG 2082. The 6th-ranked ligand, NADH, features a score of -7.847 and includes 1 backbone hydrogen bond and 4 sidechain hydrogen bonds, but no π interactions. The 7th-ranked ligand, alcaftadine, has a score of -7.847 and features only 2 side-chain hydrogen bond interaction. The 8th-ranked ligand, a stereoisomer of labetalol, has a score of -7.795, featuring 3 sidechain and 1 backbone hydrogen bond, as well as a π-cation interaction with ARG 2082. The 9th-ranked ligand, also a stereoisomer of labetalol, features 2 sidechain and 2 backbone hydrogen interactions, but lacks a π-cation interaction. The 10th-ranked ligand, kanamycin, features 3 sidechain hydrogen bonds and 1 backbone hydrogen bond, with no π-cation interactions. The 11th-ranked ligand, a third

172

stereoisomer of labetalol hydrochloride, has a score of -7.667 and features two backbone and three sidechain hydrogen interactions, one of which involved ASP2082. A π-π stacking interaction was observed with HIS 1912. The 12th-ranked ligand, histamine dihydrochloride, has a score of -7.654 and features a single backbone hydrogen bond interaction, a sidechain hydrogen interactions, a hydrogen interaction with a water molecule, and a π-cation interaction with ARG 2082. The 13th-ranked ligand, apogen, features a docking score of -7.597 and consists of two sidechain and 2 backbone hydrogen bonds; all but one of the main-chain bonds originate from a single primary amine. Overall, much fewer π interactivity was observed among the top ligands in the ZDD library when compared to the ZBC subset; however, the lack of correlation between hydrogen bond count or molecular weight and docking score was similar. Interaction with ARG2082 – whether through hydrogen bonding or π-cation interactions – remained a defining feature among the ligands.

Table 6-2. Top 10 ZDD Ligands by Docking Score
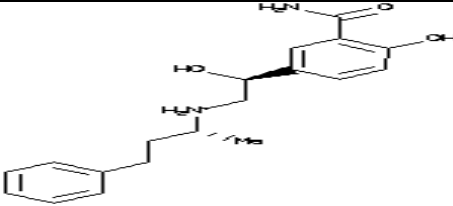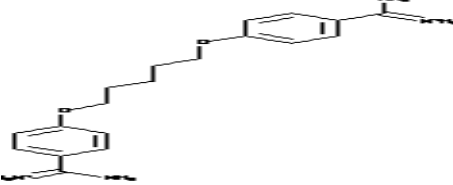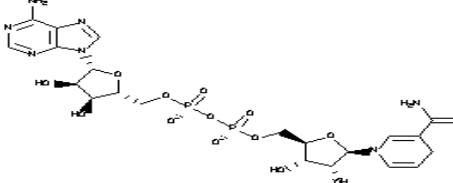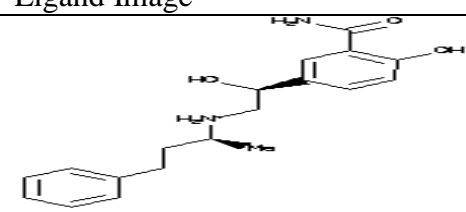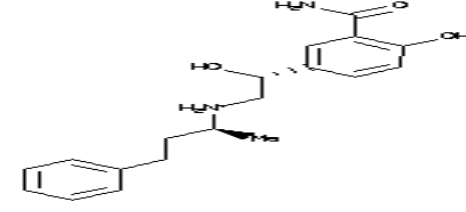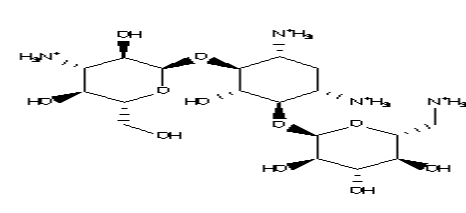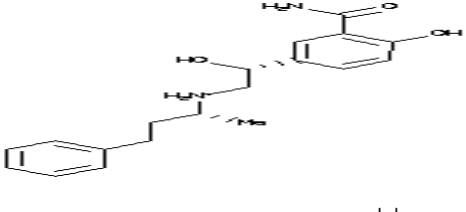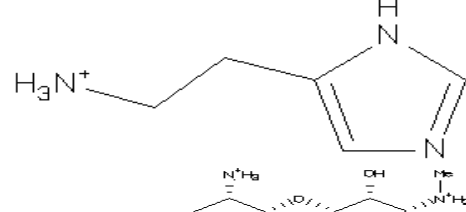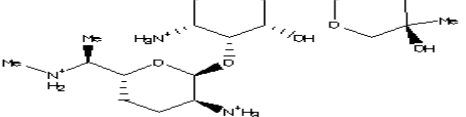
Rank	Name	ZINC ID	MW	Score	Ligand Image
1	Labetalol Hydrochloride	00000416	329.42	-8.1	
2	Pentamidine Isethionate	01530775	342.443	-8.077	
6	NADH	53682927	663.43	-7.847	

Table 6-2 Continued

Rank	Name	ZINC ID	MW	Score	Ligand Image
8	Labetalol hydrochloride	00403011	329.42	-7.795	
9	Labetalol hydrochloride	00403010	329.42	-7.748	
10	Kanamycin A Sulfate	08214590	488.535	-7.722	
11	Labetalol hydrochloride	00004319	329.42	-7.667	
12	Histamine dihydrochloride	00388081	112.156	-7.654	
13	Apogen	08101116	482.643	-7.597	
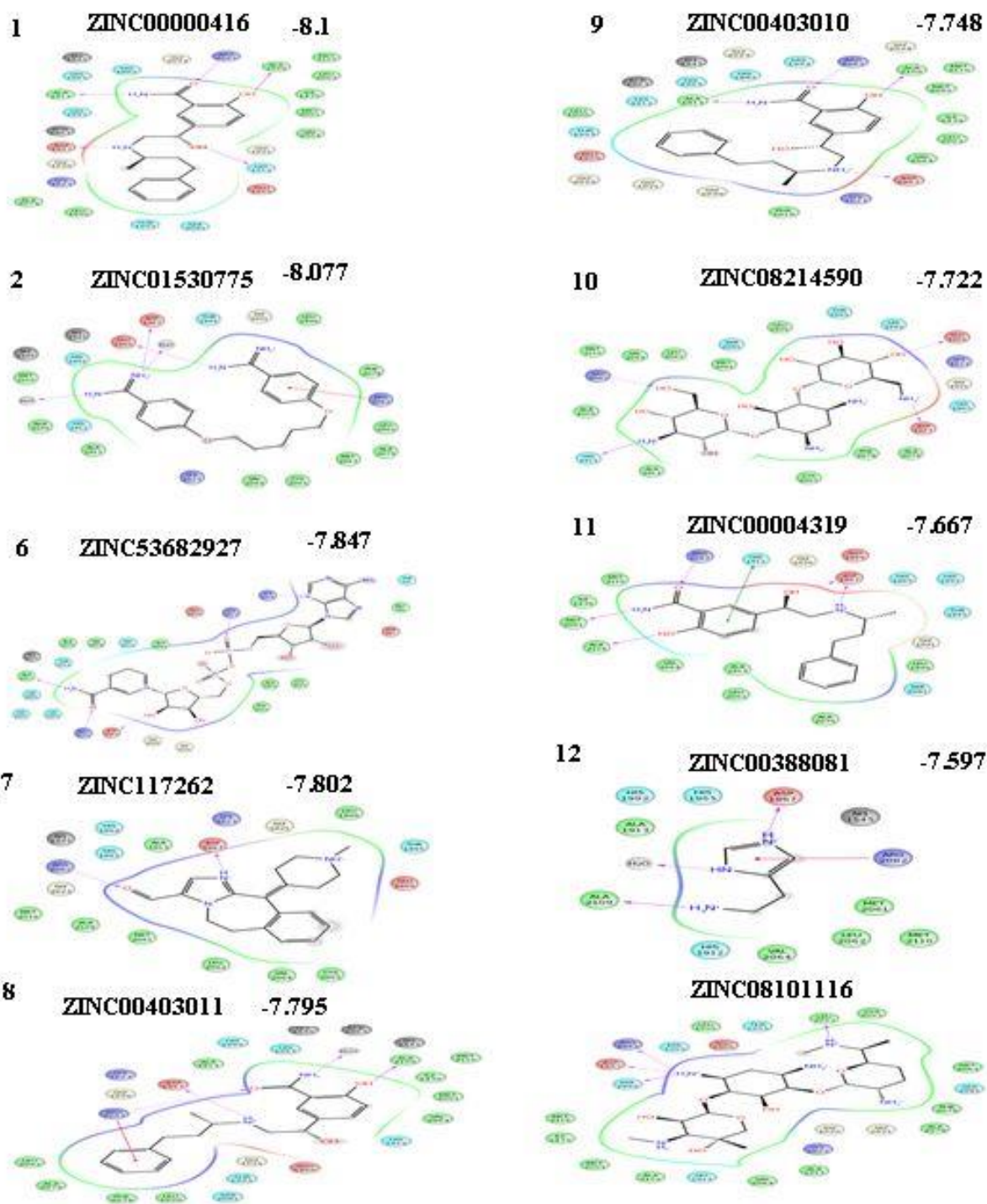
Figure 6-2. Top ZDD ligands interacting with the residues of the active site. Species are colored by type or name: HIS is colored in light blue, GLY in white, hydrophobic residues green, charged residues red (-) or violet (+), and all other species in gray. Interactions with amino acids are shown as solid (backbone) or dashed (sidechain) lines.

6.4 ZINC Natural Products

The ZINC Natural Products (ZNP) subset consists of a collection of natural products and natural product derivatives that are commercially available. This subset is built from 180,313 compounds made available through several vendors. Many of the ligands within this subset received considerably low docking scores, but due to availability and price, only a few could be selected for further analysis (refer to "Post-docking Analysis" further within the chapter) Table (6-3) catalogs the 11 highest-scoring ligands; we chose to include an 11th ligand in this instance due to the fact that the 9th, 10th, and 11th-ranked ligands all very nearly featured the same docking score. Figure (6-3) displays the interactions between these ligands and relevant active site residues and ions. The top-ranked ligand, with a docking score of -10.53, features 2 backbone and 2 sidechain hydrogen bonds, one of each occurring with either an ester or hydroxyl group. The 2nd-ranked ligand, with a docking score of -10.50, features two sidechain hydrogen bonds that share a common carbonyl group, as well as a π-stacking interaction with HIS 521. The 3rd-ranked ligand, with a score of -10.46, features 5 sidechain hydrogen bonds, all of which involve interactions with carboxyl oxygens. A stereoisomer of this ligand, named desmosine, was used in further study (refer to "Post-docking Analysis) despite its lower score due to availability and pricing. The 4th-ranked ligand, with a docking score of -10.32, consists of four backbone and three sidechain interactions; all terminal hydroxyl groups are involved with at least one of these hydrogen bonds. The 5th-ranked ligand, with a score of -10.28, features one backbone and two sidechain hydrogen bond interactions, as well as a unique interaction between a carboxyl group and a nickel ion (Ni1). The 6th-ranked ligand, with a score of -10.19, features 3 sidechain interactions and 2 backbone interactions, largely with amide groups (with the exception of an amine interacting with PRO3303). The 7th ligand on the list, also possessing a docking score of -10.19, featured 3 sidechain hydrogen interactions, while a carboxyl group interacted

simultaneously with ONB, H11, and Ni1. The 8th-ranked ligand, with a score of -10.18, possesses 2 sidechain and 3 backbone hydrogen interactions; only one sidechain interaction does not involve an amide group. The final three ligands all share a docking score of -10.14, though their interaction sets are considerably varied. The first of these three ligands features one backbone and three sidechain hydrogen interactions split between hydroxyl and ether groups. The second features 3 sidechain hydrogen interactions as well as two π-stacking interactions – one on either end of the ligand – with HIS residues. The final compound features 4 sidechain hydrogen bonds involving two carboxyl groups, coupled with interactions with Ni1 and H11. Within the top ligands, there once again does not appear to be any correlation of sorts between docking score and any observable property, including hydrogen bond type and count, molecular weight, or π interactions. However, the frequency of sidechain hydrogen bonds, as well as the lack of π-cation interactions (particularly with ARG2082) is noteworthy across this group.

Table 6-3. Top 10 ZNP Ligands by Docking Score
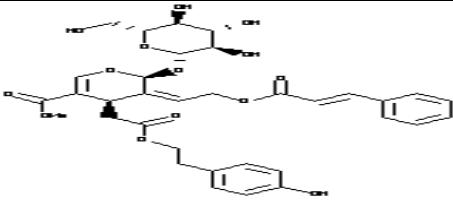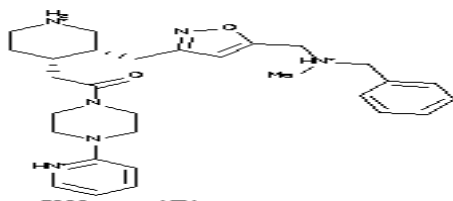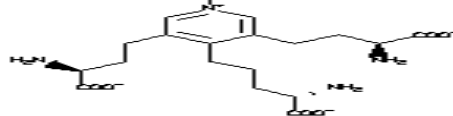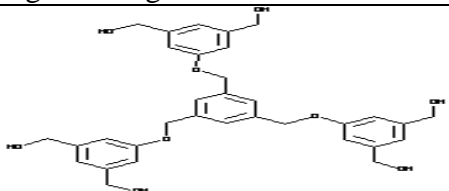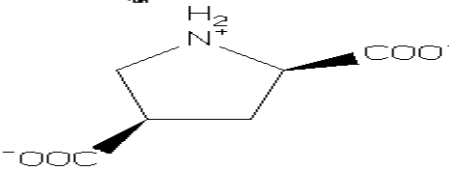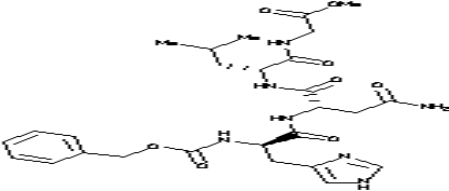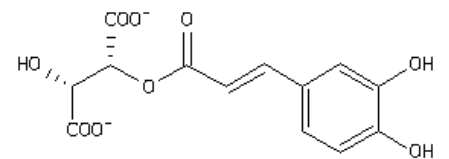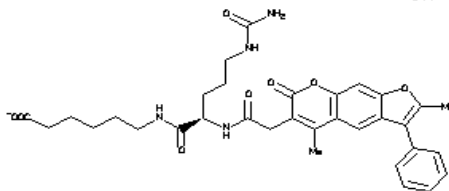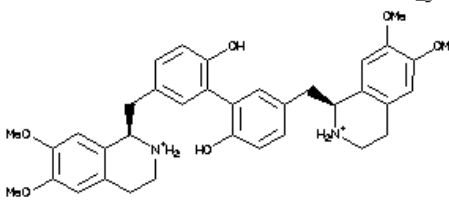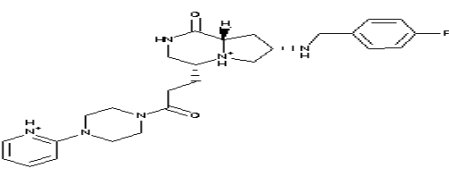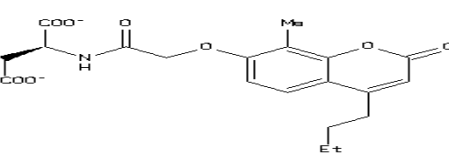
Rank	Name	ZINC ID	MW	Score	Ligand Image
1	methyl (2S,4R,Z)-3-(2-(cinnamoyloxy)ethylidene)-4-(2-(4-hydroxyphenethoxy)-2-oxoethyl)-2-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2…	067911960	670.664	-10.53	
2	2-[(3R,4S)-3-[[5-[(benzyl-methyl-amino)methyl]isoxazol-3-yl]methyl]-4-piperidyl]-1-[4-(2-pyridyl)pip	20463590	505.687	-10.50	
3	Desmosine*	35024530	523.587	-10.46	

Table 6-3. Continued

Rank	Name	ZINC ID	MW	Score	Ligand Image
4	[3-[[3,5-bis[[3,5-bis(hydroxymethyl)phenoxy]methyl]phenyl]methoxy]-5-(hydroxymethyl)phenyl]methanol	04264743	576.642	-10.32	
5	2,4-Pyrrolidinedicarboxylic acid	03873053	158.11	-10.28	
6	methyl 2-[2-[2-[2-benzyloxycarbonylamino-3-(1H-imidazol-4-yl)propanoyl]amino-3-carbamoyl-propanoyl]amino-4-methyl-pentanoyl]aminoacetate	70665145	587.634	-10.19	
7	Caftaric acid	02389524	310.214	-10.19	
8	6-[[(2R)-2-[[2-(2,5-dimethyl-7-oxo-3-phenyl-furo[3,2-g]chromen-6-yl)acetyl]amino]-5-ureido-pentanoyl	70706518	617.679	-10.18	
9	4-[[(1R)-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-1-yl]methyl]-2-[5-[[(1S)-6,7-dimethoxy-1,2,3,4-…	68568485	598.74	-10.14	
10	(4R,7S,8aS)-7-[(4-fluorophenyl)methylamino]-4-[3-oxo-3-[4-(2-pyridyl)piperazin-1-yl]propyl]-3,4,6,7,…	22936876	481.596	-10.14	
11	(2S)-2-[[2-(4-butyl-2-keto-8-methyl-chromen-7-yl)oxyacetyl]amino]succinate	02122226	403.387	-10.14	
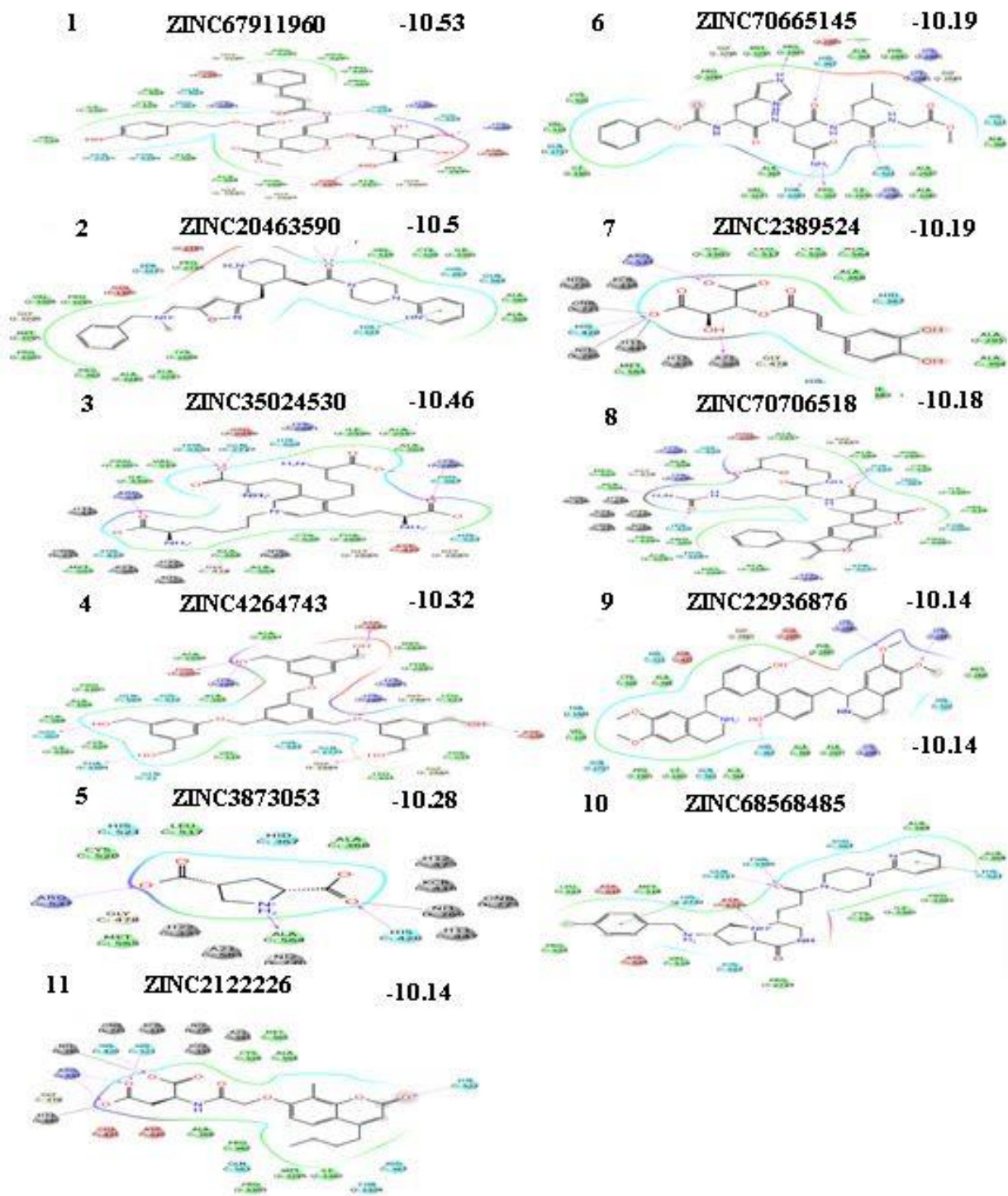
Figure 6-3. Top ZNP ligands interacting with the residues of the active site. Species are colored by type or name: HIS is colored in light blue, GLY in white, hydrophobic residues green, charged residues red (-) or violet (+), and all other species in gray. Interactions with amino acids are shown as solid (backbone) or dashed (sidechain) lines.

6.5 ZINC In Man

The next subset, ZINC In Man (ZIM), is an expansion of the ZDD library that contains all compounds that have ever existed within a human being at any point in time. This includes approved drugs, experimental compounds, nutrient supplements, recreational drugs of varying legality, and the like. This subset contains 11,421 entries. As in previous databases, stereoisomers of each molecule are included, up to a maximum of four per molecule. The top 10 ligands for which interaction data is available are presented in Table (6-4), while their interactions within the active site are depicted in Figure (6-4). The top-scoring ligand, with a docking score of -10.07, features 2 backbone and 4 sidechain hydrogen bond interactions, as well as 3 interactions between the phosphate group and H11, ONB, and a nickel ion. The second-ranked ligand, with a score of -10.05, also features 4 sidechain hydrogen bonds, but only one backbone bond. An interaction between a carboxyl group and a nickel ion is also observed. The third-ranked ligand, with a score of -9.96, features 3 sidechain hydrogen interactions, 1 backbone interaction, and a phosphate group that interacts with ONB, H11, and a nickel ion. The fourth-ranked ligand, with a score of -9.75, features only one hydrogen bond with a sidechain and one with the backbone, but maintains interactions with ONB, H11, and a nickel ion via its single carboxyl group. The fifth-ranked ligand, scoring -9.62, features 2 sidechain and 1 backbone hydrogen interaction, as well as an interaction between a nickel ion and sulfonyl group. The sixth-ranked ligand, with a score of -9.54, features 2 sidechain and 1 backbone hydrogen bond, and interacts with ONB, H11, and Ni1 via its phosphate group. The seventh-ranked ligand, with a score of -9.53, features 1 sidechain and 1 backbone hydrogen interaction, and interacts with H11, ONB, and a nickel ion via a single carboxyl group. The 9th-ranked ligand, with a score of -9.36 features two backbone hydrogen interactions and one sidechain hydrogen interaction, but no other notable features. The 11th-ranked ligand, a stereoisomer of the 2nd-ranked ligand that

180

possesses a score of -9.34, features 1 backbone and 2 sidechain hydrogen interactions, and interacts with H22, ONB, and a nickel ion via a carboxyl group. The 12th-ranked ligand features a unique set of interactions for the 4th-ranked ligand; with a score of -9.33, this interaction profile features 1 sidechain and one backbone bond as well, but the interactions between the carboxyl group and the nickel ion, ONB, and H11 species have been shuffled, and included a second interaction between H11 and the carboxyl group (specifically, one with each oxygen). Furthermore, there exists two π-stacking interactions between the chlorobenzyl group and both HIS521 and HIS522. It would appear that molecular weight, number of hydrogen bonds, and number of interactions between the ligand and H11, ONB, and the nickel ion do not correlate well with docking scores; however, comparison between the unique interaction profiles of Bepotastine Besilate might suggest an undesirable effect stemming from either the π-stacking interactions or the distribution of interactions about the carboxyl group.

Table 6-4. Top 10 ZIM Ligands by Docking Score
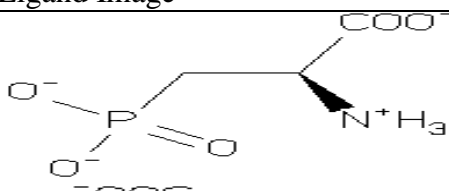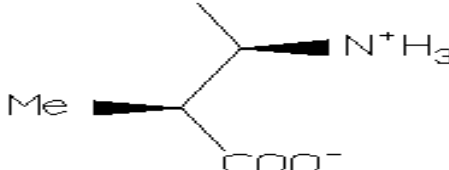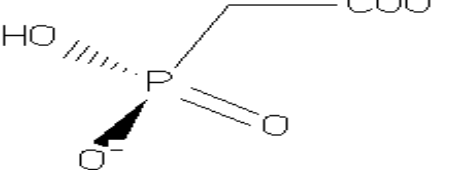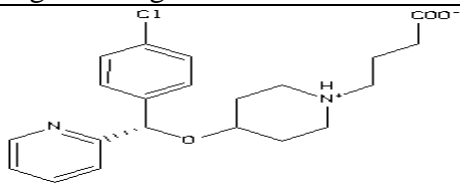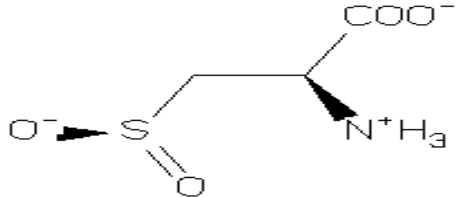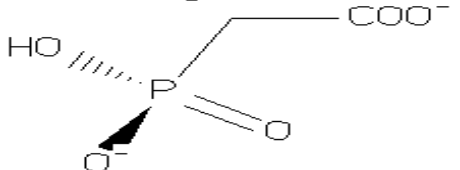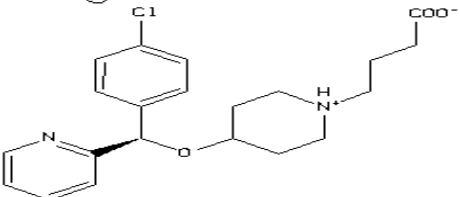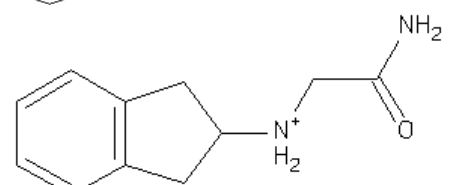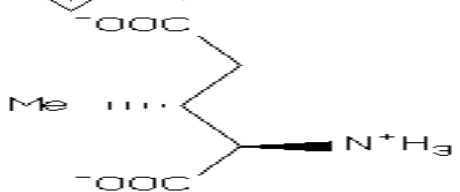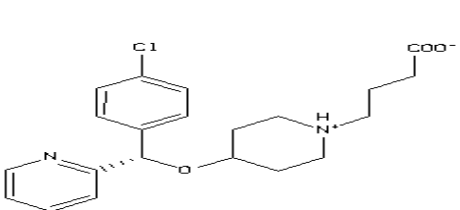
Rank	Name	ZINC ID	MW	Score	Ligand Image
1	Aminophosphonopropionic acid	03873028	167.057	-10.07	
2	2-Amino-3-methylsuccinic acid	00901781	146.122	-10.05	
3	Phosphonoacetic acid	19850691	138.015	-9.96	

Table 6-4. Continued

| Rank	Name	ZINC ID	MW	Score	Ligand Image
4	Bepotastine Besilate	00602128	388.895	-9.75	
5	L-cysteinesulfinic acid	03869282	152.151	-9.62	
6	Phosphonoacetic acid	19850691	138.015	-9.54	
7	TAU-284DS	03871703	388.895	-9.53	
9	CHF-3381	00009005	191.254	-9.36	
11	2-Amino-3-methyl-pentanedioic acid	01672631	160.149	-9.34	
12	Bepotastine Besilate	00602128	388.895	-9.33	
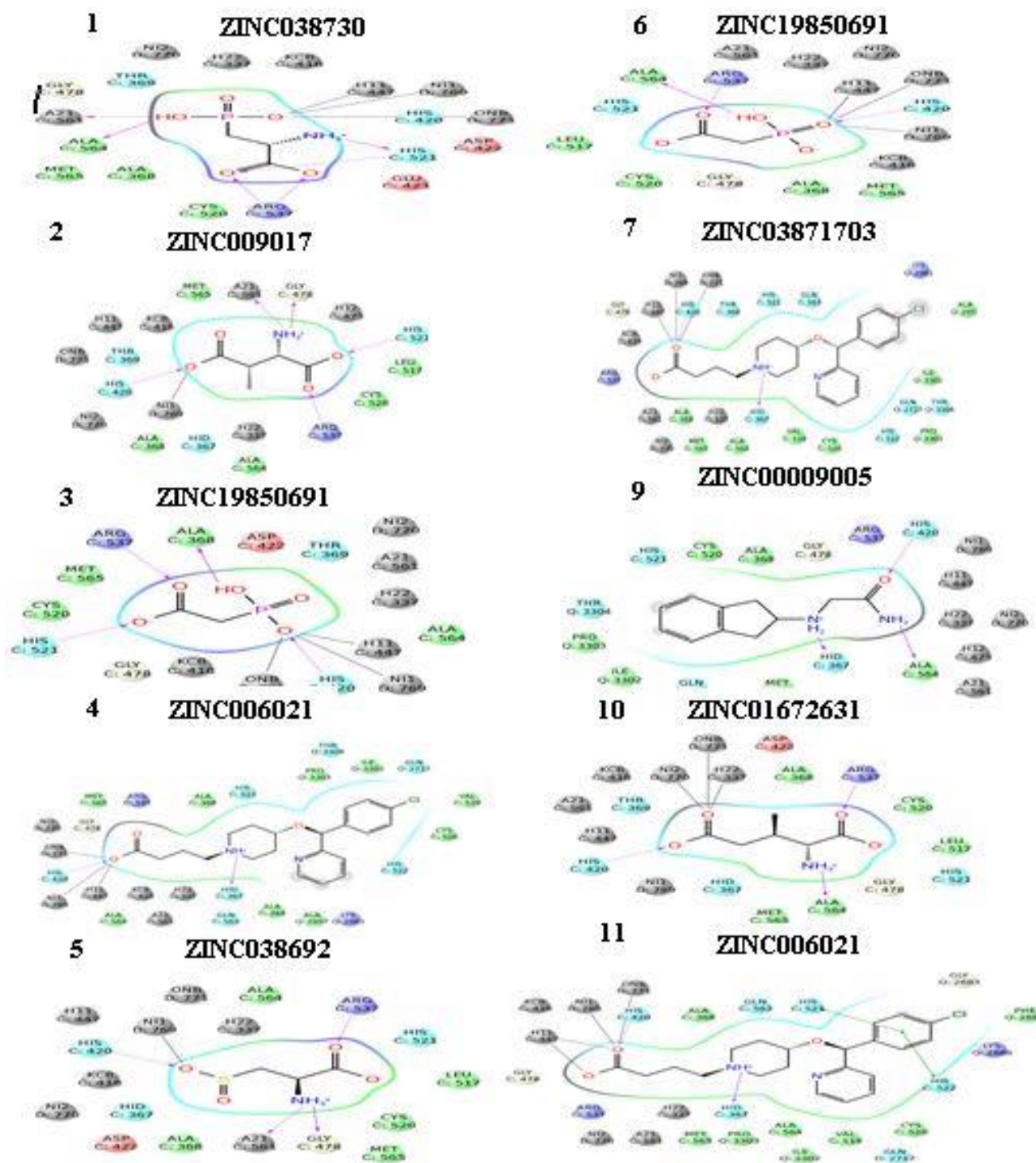
Figure 6-4. Top ZIM ligands interacting with the residues of the active site. Species are colored by type or name: HIS is colored in light blue, GLY in white, hydrophobic residues green, charged residues red (-) or violet (+), and all other species in gray. Interactions with amino acids are shown as solid (backbone) or dashed (sidechain) lines.

6.6 ZINC Biogenic Compound Leads

The ZINC Biogenic Compound Leads (ZBC Leads) subset consists of lead-like compounds within the ZBC subset. In this instance, a lead or lead-like compound is a molecule that demonstrates signs of therapeutic utility, but requires additional modification to be effective as a drug (that is, it is a compound that "leads" one to a novel drug). This library consists of 26,504 entries, with molecules following the same rules for stereoisomer quantities as outlined in past sections. Table (6-5) catalogs the top 10 ligands selected for discussion, while Figure (6-5) displays their interactions within the active site. The top-ranked ligand, with a docking score of -9.68, features only one sidechain and one backbone hydrogen interaction, interestingly. The 2nd-ranked ligand, with a score of -9.41, consists of 2 sidechain hydrogen bonds and 2 hydrogen bonds with water molecules. The 3rd-ranked ligand, with a score of -9.256, features hydrogen interactions with one backbone and one sidechain, while also including a π-cation interaction with ARG2082. The 4th-ranked ligand, with a score of -9.236, consists of 1 backbone and 2 sidechain interactions, as well as a π-stacking interaction with HIS1912 and a hydrogen bond between a water molecule and amide. The 5th-ranked ligand, with a score of -9.14, features 2 sidechain and one backbone hydrogen interaction, as well as a π-cation interaction involving ARG 2082. The 6th-ranked ligand, with a score of -9.13, features one sidechain and one backbone hydrogen bond, as well as a similar interaction with water. The 7th-ranked ligand, with a score of -9.11, consists of 1 backbone and 2 sidechain hydrogen interactions; one of each involved with the same hydroxyl group. The 8th-ranked ligand, with a score of -9.07, has no clear interactions with any residue in the binding pocket, curiously. The 9th-ranked ligand features 1 backbone and two sidechain hydrogen interactions, a hydrogen bond with a water molecule, and a π-stacking interaction with HIS 1912. The 10th-ranked ligand features 2 sidechain interactions, one backbone interaction, and a π-cation interaction with ARG2082. Ultimately, the quantity of

184

observable interactions, the quantity of water molecules involved, and the spread of these

interactions across the ligand did not appear to correlate well with docking score, though it is

interesting to note that the hydrogen bond profile of the top-scoring ligand was relatively

simplistic, with only 2 total bonds depicted.

Table 6-5. Top 10 ZBC Leads Ligands by Docking Score
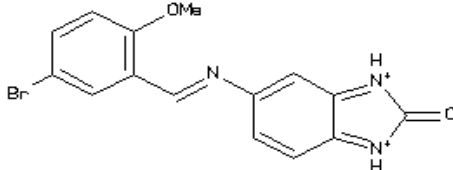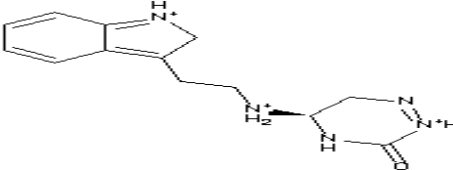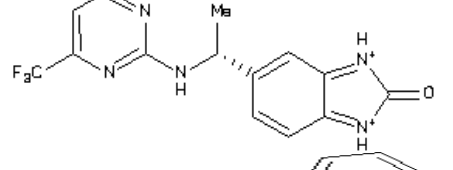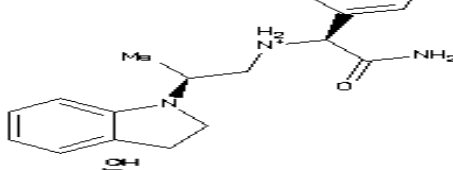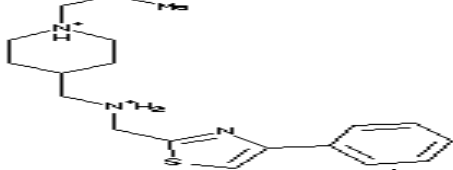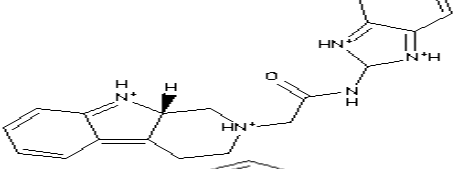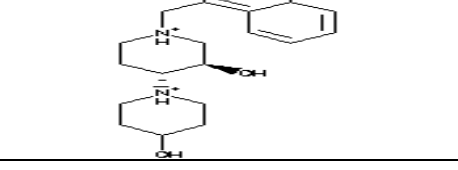
Rank	Name	ZINC ID	MW	Score	Ligand Image
1	5-[(E)-(5-bromo-2-methoxy-phenyl)methyleneamino]benzimidazol-2-one	40009214	346.184	-9.68	
2	(5R)-5-[2-(2H-indol-3-yl)ethylamino]-5,6-dihydro-4H-1,2,4-triazin-3-one	63904478	260.321	-9.41	
3	5-[(1R)-1-[[4-(trifluoromethyl)pyrimidin-2-yl]amino]ethyl]benzimidazol-2-one	63952292	191.254	-9.256	
4	(2S)-2-[[(2S)-2-indolin-1-ylpropyl]amino]-2-phenyl-acetamide	27014473	310.421	-9.236	
5	(2S)-1-[4-[[(4-phenylthiazol-2-yl)methylamino]methyl]-1-piperidyl]propan-2-ol	69347385	347.528	-9.14	
6	2-[(9aS)-1,3,4,9a-tetrahydropyrido[3,4-b]indol-2-yl]-N-(2H-benzimidazol-2-yl)acetamide	65395991	349.438	-9.13	
7	(3R,4R)-4-(4-hydroxy-1-piperidyl)-1-(1-naphthylmethyl)piperidin-3-ol	12757903	342.483	-9.11	

Table 6-5. Continued

Rank	Name	ZINC ID	MW	Score	Ligand Image
8	(2R)-2-[(2S,3S)-2,5-dimethyl-1-(3-pyridylmethyl)-2,3-dihydropyrrol-3-yl]-2H-benzimidazole-5-carbonit	63868913	332.431	-9.07	
9	(2S)-2-(benzothiophen-3-ylmethylamino)-2-phenyl-acetamide	71818570	296.395	-9.04	
10	3-(2H-benzimidazol-2-yl)-2-imino-chromen-7-ol	63393995	280.307	-9.00	
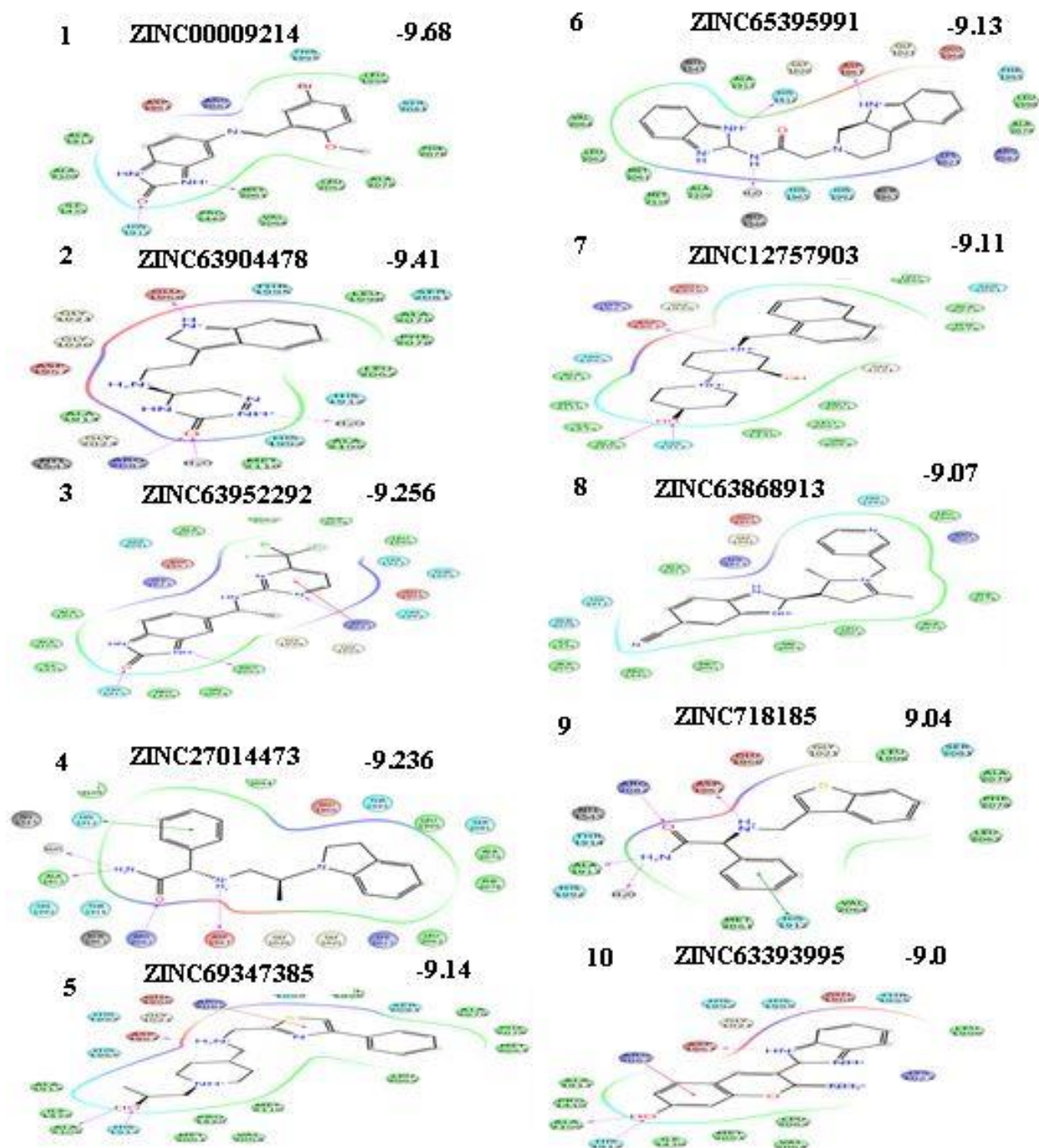
Figure 6-5. Top ZBC leads interacting with the residues of the active site. Species are colored by type or name: HIS is colored in light blue, GLY in white, hydrophobic residues green, charged residues red (-) or violet (+), and all other species in gray. Interactions with amino acids are shown as solid (backbone) or dashed (sidechain) lines.

6.7 Individual Ligands

To construct the final subset of ligands for analysis, we compiled all ligands that were shown to bind well with *H. pylori* urease in the literature via experiment. In doing so, we gathered 23 ligands for analysis. These ligands have been arranged by descending docking score in Table (6-6); note that ligands for which no docking image or data exists have been omitted for the time being. Meanwhile, Figure (6-6) displays these ligands' interactions with the active site. The top-ranked ligand, catechin, with a score of -6.969, features 4 backbone hydrogen interactions, two of which are directed to the same ALA 2309 residue. The 2nd-ranked ligand, quercetin, with a score of -6.384, features a similar interaction profile, with the exception of the missing interaction with LEU 2062 and an additional sidechain bond with ASP 1967. The next ligand, kaemperol, with a score of -6.22, consists of three sidechain and two backbone interactions – many sharing the same hydroxyl groups – as well as a π-cation interaction with ARG2082. The 5th-ranked ligand, epicatechin gallate, features 2 backbone and 2 sidechain hydrogen interactions – again predominantly via hydroxyl groups – as well as a hydrogen bond with water. The next ligand, gallocatechin, features one sidechain and three backbone hydrogen interactions, as well as an interaction with water. Two π-cation interactions also exist; one involved the commonly observed ARG 2082, whereas the other involved LYS 1023. The next ligand, epigallocatechin, with a score of -5.6, features four sidechain interactions shared between two residues (ARG 2082 and ASP 1967), as well as two backbone hydrogen bonds. ClAHA6g, the next-ranking ligand, with a score of -5.94, features two sidechain and two backbone hydrogen interactions. One of these interactions does not involve a hydroxyl or carbonyl group on the ligand; ClAHA6g is the highest-ranking on the list with that particular feature. The next ligand, myricetin, with a score of -5.59, features two hydrogen bonds with the backbone, one with a sidechain, and one with water; no residues or functional groups were shared among any

interaction pairs. The next ligand, epigallocatechin gallate, with a score of -5.545, features one backbone and four sidechain hydrogen interactions, as well as one interaction with water. These interactions are split between two phenol groups, one of which is involved with a π-cation interaction with ARG2082. The 13th ligand, epicatechin, with a score of -5.536, features two sidechain and two backbone interactions, one of which involves an amine. The following ligand, acetohydroxamic acid (AHA), with a score of -5.363, consists of two sidechain interactions and one backbone interaction involving an amine. The next ligand, N-(n-butyl)phosphorictriamide, with a docking score of -4.91, features two backbone interactions, one sidechain interaction, and one interaction with water, three of which involve amine groups (with the exception of the sidechain interaction with ARG2082). The next ligand, ADD355, with a score of -4.80, features 2 sidechain and 2 backbone hydrogen bonds, as well as an interaction with water. The 17th-ranked ligand, caffeic acid, with a score of -4.742, features three sidechain interactions – two of which share ARG 2082 – and one backbone hydrogen bond. Tartaric acid, the 18th-ranked ligand, features a score of -4.46 and is involved in no backbone interactions, consisting of three sidechain hydrogen bonds and one interaction with water. The next ligand, anionic quercetin, scored considerably lower than its uncharged counterpart, receiving -4.289. Anionic quercetin features one sidechain and two backbone interactions, each with a unique residue. Finally, the ligand 8D, with a score of -4.244, consists of only one sidechain and one backbone hydrogen interaction.

Table 6-6. Individually Selected Ligands by Docking Score
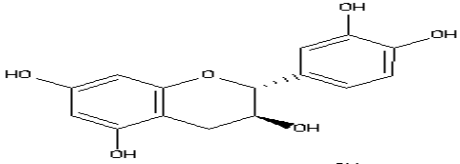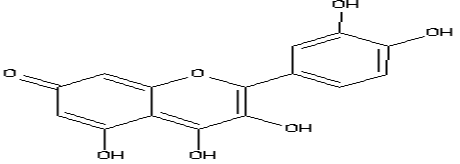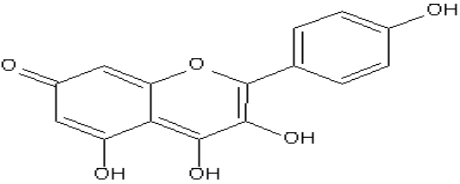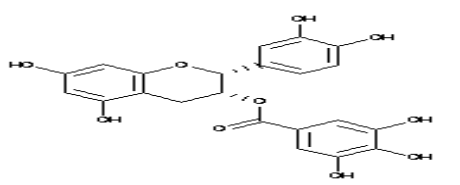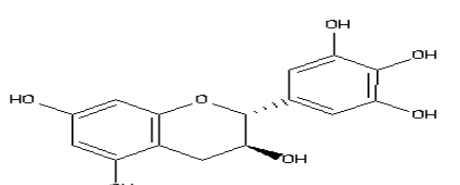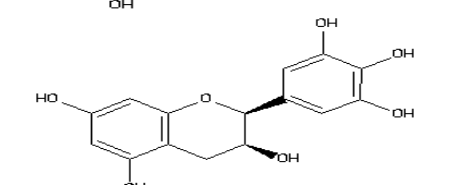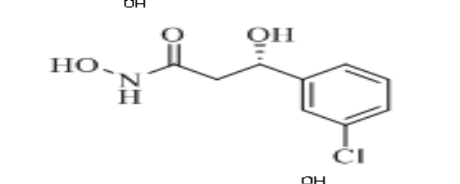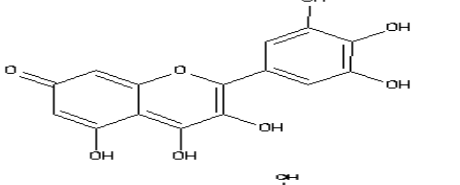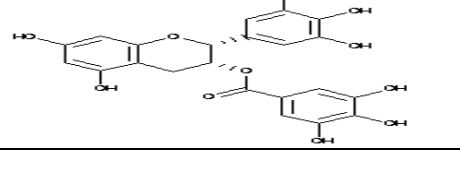
Rank	Name	ZINC ID	MW	Score	Ligand Image
1	Catechin	00119983	290.271	-6.969	
2	Quercitin	03869685	302.238	-6.384	
3	Kaempferol	03869768	286.239	-6.22	
5	(-)-Epicatechin gallate	03978503	442.376	-6.148	
6	(+)-Gallocatechin	03870338	306.27	-6.08	
8	Epigallocatechin	03870339	306.27	-5.96	
9	Claha6g	N/A	215.629	-5.94	
10	Myricetin	03874317	318.237	-5.59	
11	(-)-Epigallocatechin gallate	03870412	458.375	-5.545	

Table 6-6. Continued

| Rank	Name	ZINC ID	MW	Score	Ligand Image
12	Epicatechin	00119988	290.271	-5.536	
13	Acetohydroxamic acid	04658603	75.067	-5.536	
14	N-(n-butyl)phosphoric triamide	02562350	151.149	-4.91	
15	ADD335	N/A	299.250	-4.805	
16	Caffeic acid	00058172	179.151	-4.742	
17	Choline bitartrate	00895297	148.07	-4.46	
19	Anionic quercitin	03869685	302.238	-4.289	
20	8D	N/A	358.321	-4.244	
21	(-)-Catechin Gallate	08681494	442.37	-6.02	
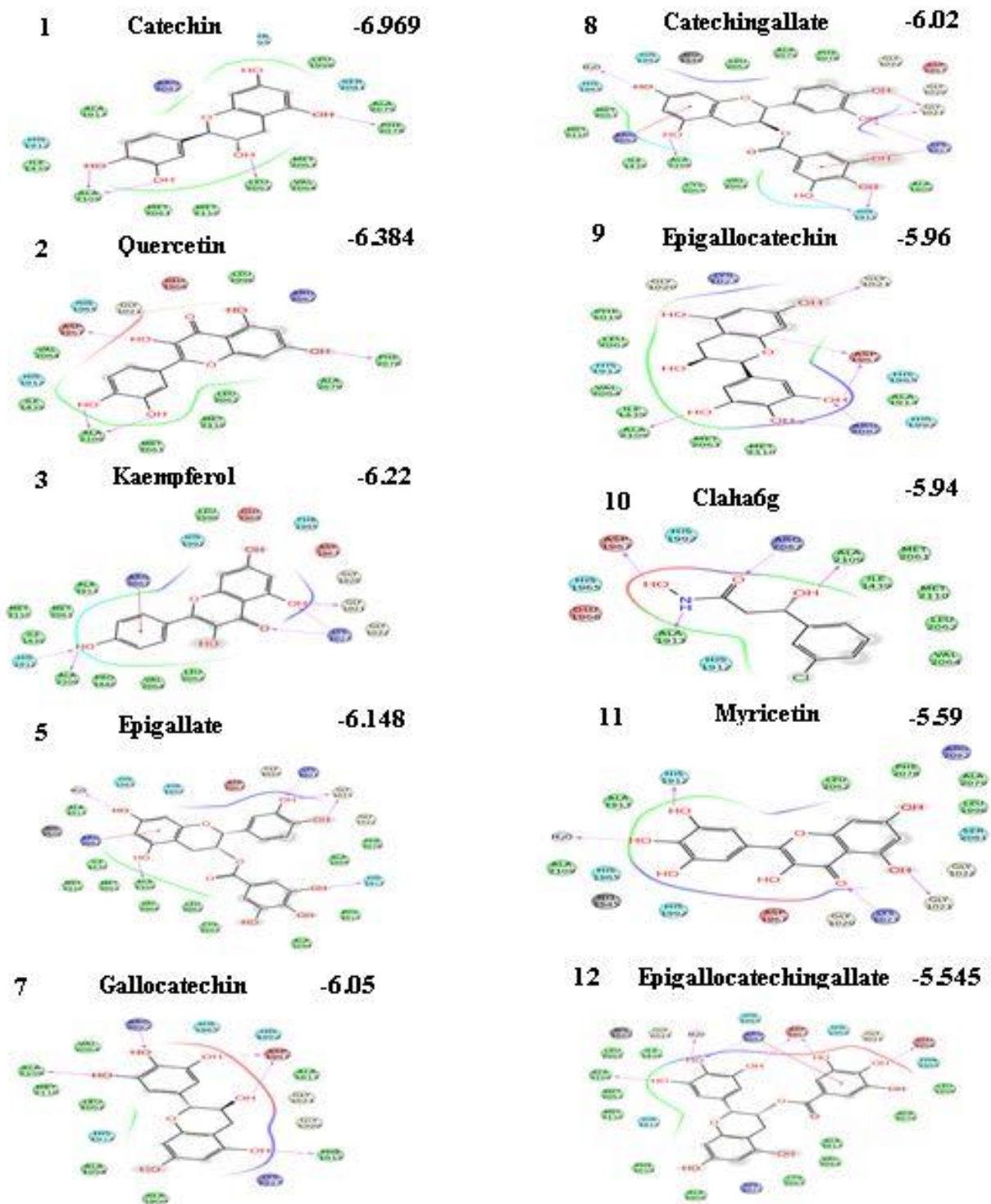
Figure 6-6. The set of individually chosen ligands interacting with the residues of the active site. Species are colored by type or name: HIS is colored in light blue, GLY in white, hydrophobic residues green, charged residues red (-) or violet (+), and all other species in gray. Interactions with amino acids are shown as solid (backbone) or dashed (sidechain) lines.
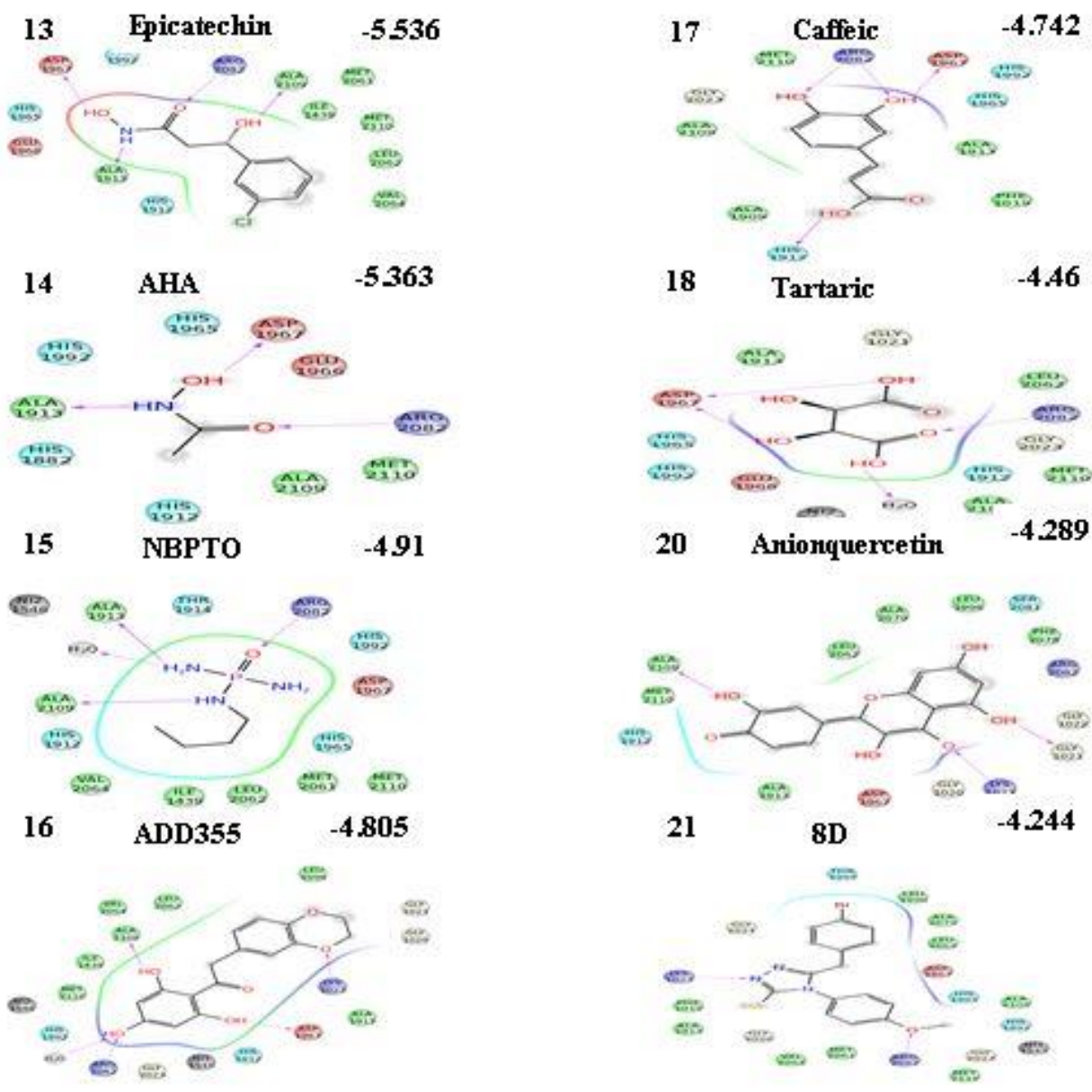
Figure 6-6. Continued

6.8 Post-Docking Analysis

After performing the docking analysis on the aforementioned ZINC subsets, the ligands with the best balance between docking score and practicality were selected for further experimental analysis. This analysis was conducted by collaborative researchers, using both wild-type and C319A mutants of *K. aerogenes* urease.[81] As alluded to previously, *K. aerogenes* urease was selected due to the well-established understanding of the enzyme's structure and behavior in the literature, a result of the extensive work conducted to study it. Furthermore, the enzyme itself is smaller and easier to work with in an experimental environment, while retaining the common properties of *H. pylori* urease.[6, 12, 79-80] This is especially true of the active site of *K. aerogenes* urease, which features two nickel ions bridged by water and a carbamylated LYS residue. These ions are further coordinated by either two HIS and an additional water molecule (in the case of the first ion) or two HIS, one ASP, and a water molecule (in the case of the second ion).[6, 79-80] The use of C319A mutants alongside wild-type urease served to confirm the increase in inhibitor resistance predicted by computational models.

Among the ligands selected were kanamycin, labetalol, desmosine, and caftaric acid, while maintaining the catechols quercetin[11, 82] and epigallocatechin[12, 83] as control (Table (6-7)); the docking poses associated with their respective scores are depicted in Figure (6-7). Tanimoto coefficient scores were calculated between each possible pair of molecules using MACCS keys (Table (6-8)) to evaluate chemical similarity between the ligands; with the exception of the pair of catechols, each pair of molecules was determined to be likely chemically dissimilar, using a similarity cut-off of 0.85. Tanimoto coefficients were also calculated between each ligand and the libraries from which they were derived (Table (6-9)) to check for potentially similar entries; few were found, though the catechol pair again featured comparatively more hits than the four

chosen ligands. Basic statistical analysis for the hits found in each library was performed as well (Table (6-10)).

Unfortunately, experimental analysis performed by our collaborators revealed little to no inhibitory activity related to the four chosen inhibitors. Quercetin and epigallocatechin, meanwhile, reduced urease activity in a time-dependent fashion, suggesting that either the molecules work as slow-binding inhibitors or that complete inhibition requires a secondary step. Furthermore, the C319A mutant enzyme featured a reduced loss of activity when coupled with epigallocatechin and quercetin, though inhibitory activity was still observed; this result suggests a peculiar combination of possibilities – namely, that the catechols covalently interact with CYS 319, but that other sites of interaction may also be involved in the activity of the enzyme. Regardless, the theory that a covalent adduct of some form is involved with nonreversible inhibitory activity was further strengthened upon the irreversibility of reaction via dialysis or thiol exposure, particularly when considering that other catechol inhibitors typically displayed thiol-dependent reversibility when bound to urease.[84]
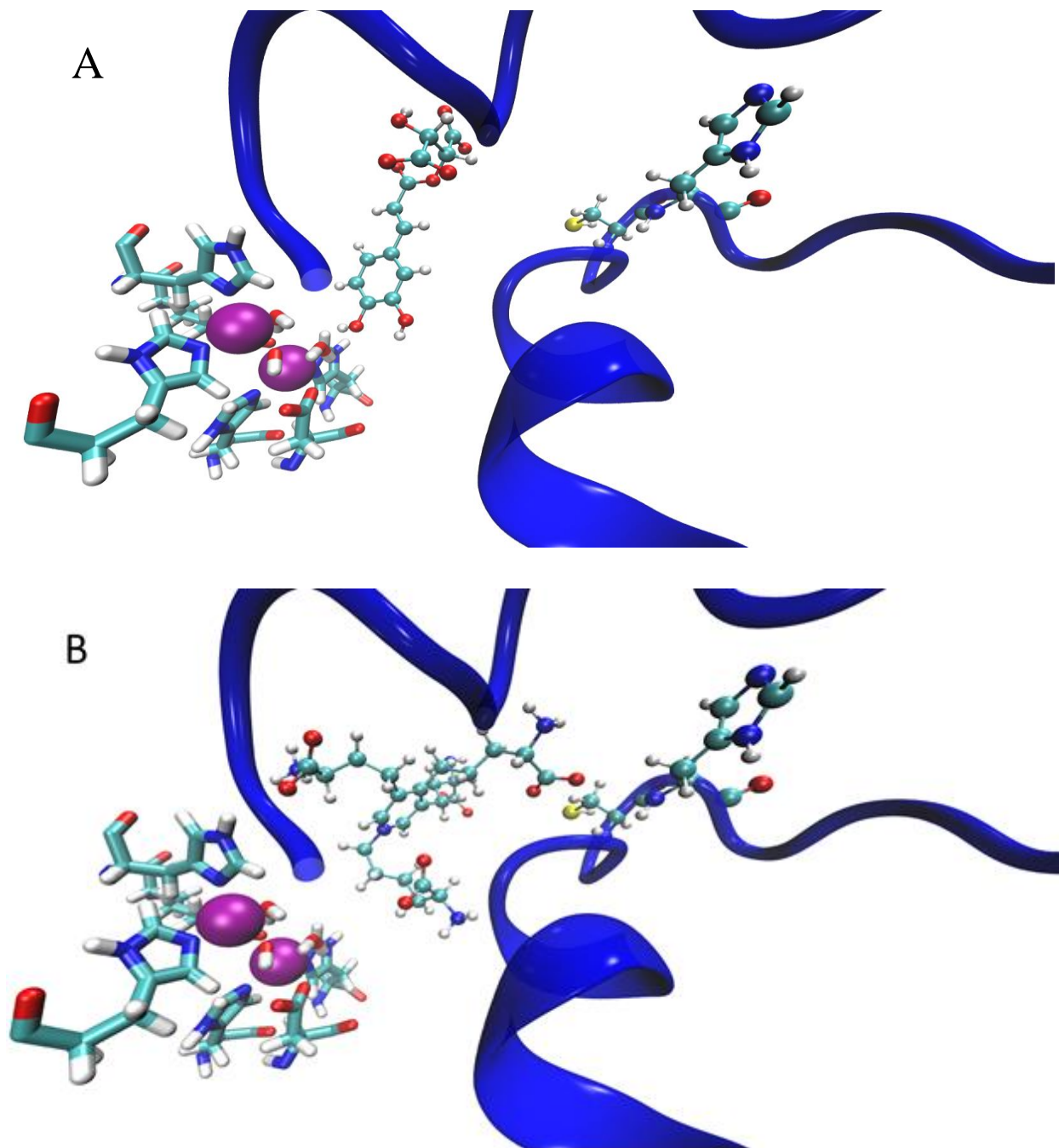
Figure 6-7. Depictions of the best-scoring docking poses for each of 4 potential urease ligands (panels A-D) and the 2 control ligands (panels E-F). In each figure, the active site flap is shown in a blue cartoon representation, the Ni pair is shown as magenta-colored Van der Waals spheres, the coordination sphere is shown in licorice representation colored by atom type, and the docked ligand – as well as the Cys319 and His320 residues – are shown as CPK representations colored by atom type. Shown in order are docked images of caftaric acid (A), desmosine (B), labetalol (C), kanamycin (D), quercetin (E), and eipgallocatechin (F).

Figure 6-7. Continued

E
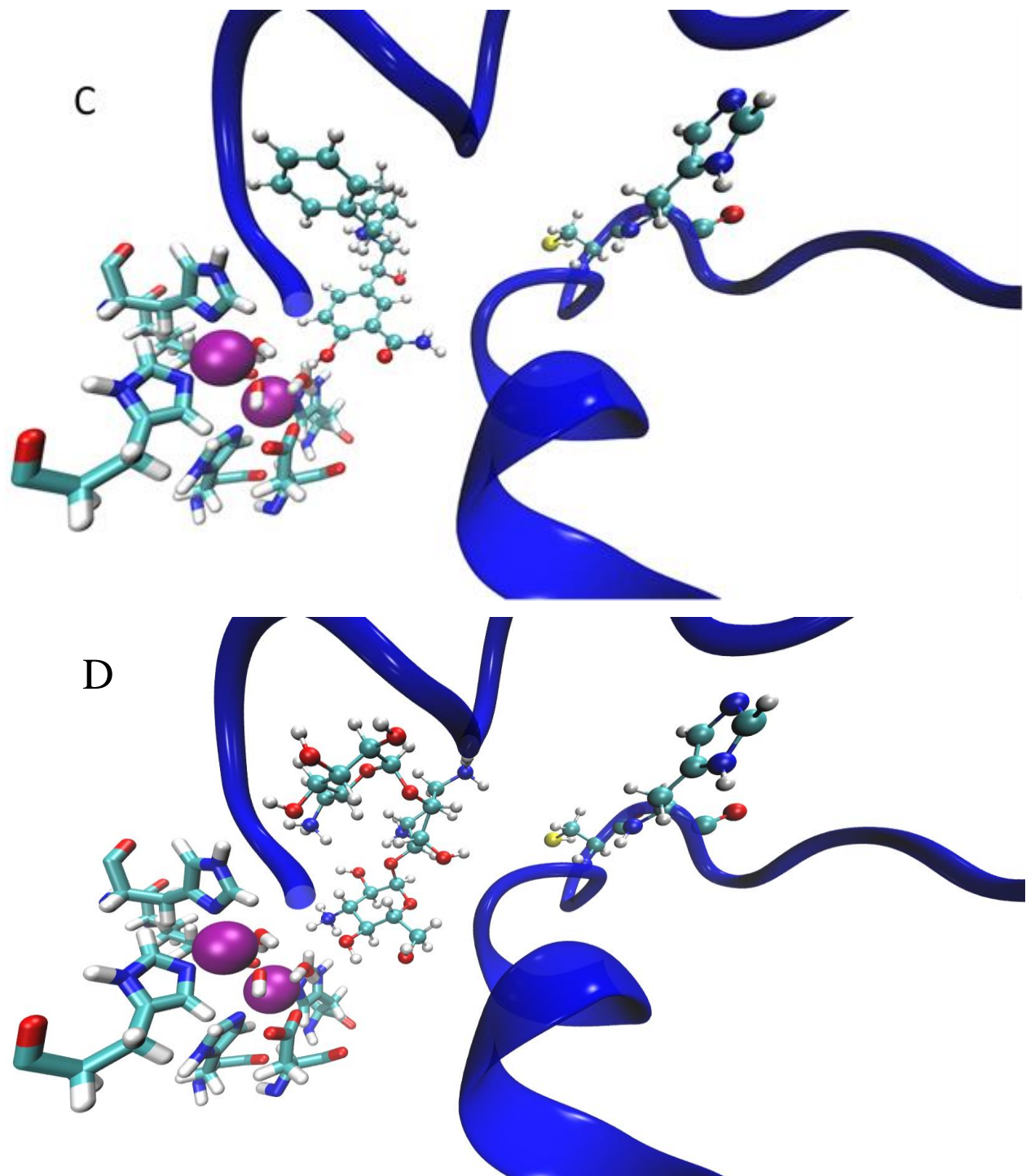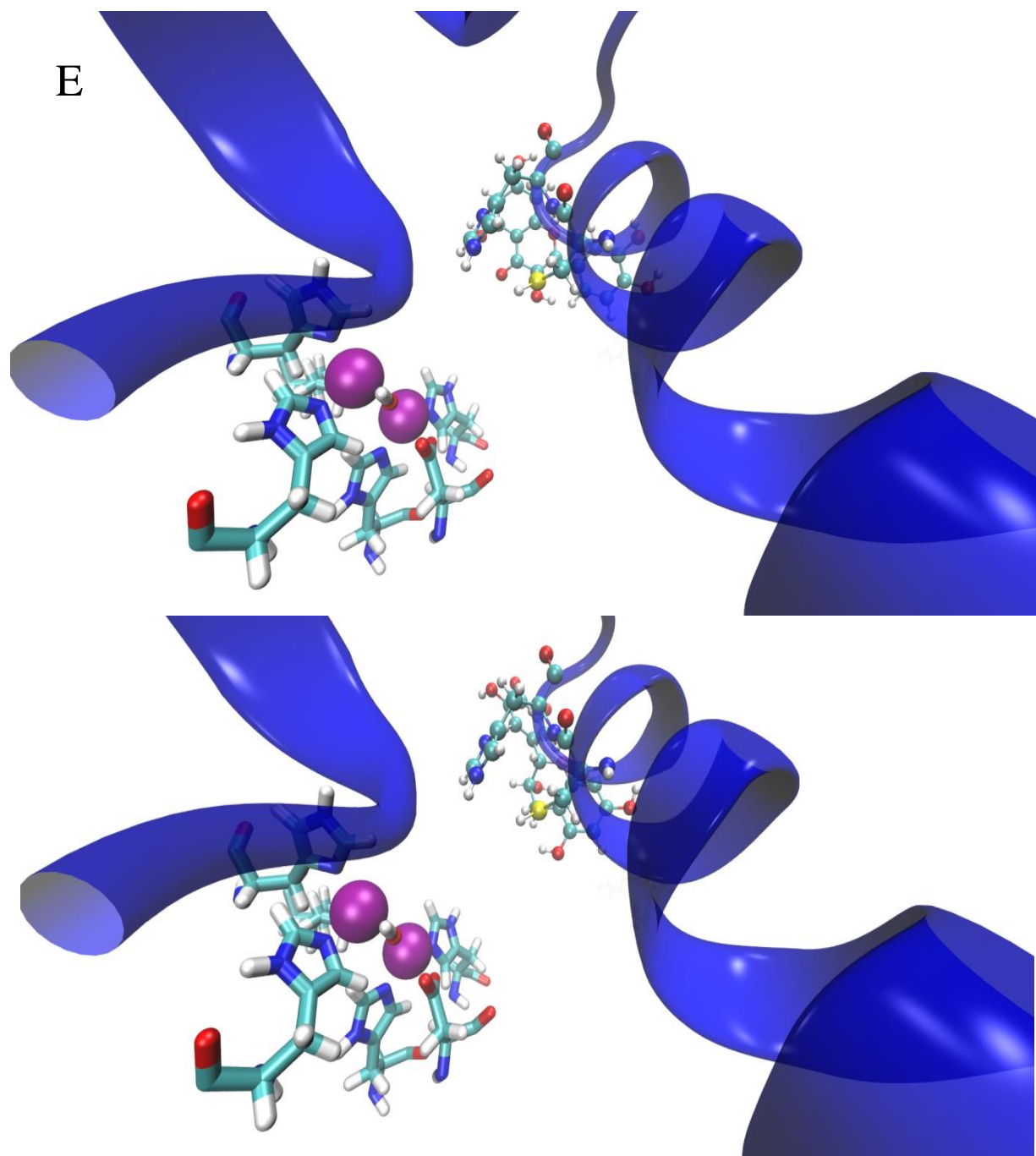
Figure 6-7. Continued

Table 6-7. Docking scores, masses, and structures of four potential urease ligands obtained from the ZINC natural products and ZINC drug databases along with two control ligands.

Entry (ZN Number) [Sublibrary]	Docking score	Molecular weight	Structure
Caftaric Acid 02389524 Drug Database	-10.2	310.21	
Desmosine 35024527 Drug Database	-10.1	523.59	
Labetalol 00000416 Natural Products	-8.1	329.42	
Kanamycin 08214590 Natural Products	-7.7	488.54	
Quercetin 03869685 Control	-6.4	302.24	
Epigallocatechin 03870339 Control	-6.0	306.27	

Table 6-8. Tanimoto similarity scores for each pair of molecules.

Tanimoto Scores (MACCS)	Caftaric Acid	Desmosine	Labetalol	Kanamycin	Quercetin	Epigallocatechin
Caftaric Acid	1.00	0.29	0.36	0.37	0.61	0.53
Desmosine	0.29	1.00	0.49	0.43	0.25	0.27
Labetalol	0.36	0.49	1.00	0.53	0.36	0.40
Kanamycin	0.37	0.43	0.53	1.00	0.40	0.47
Quercetin	0.61	0.25	0.36	0.40	1.00	0.87
Epigallocatechin	0.53	0.27	0.40	0.47	0.87	1.00

Table 6-9. Number of compounds with Tanimoto similarity scores > 0.85 in ZINC Natural Products (ZNP) and ZINC Drug Database (ZDD) libraries.

Entry	ZNP	ZDD
Caftaric Acid	16	0
Desmosine	5	0
Labetalol	4	4
Kanamycin	69	7
Querctin	285	4
Epigallocatechin	83	4

Table 6-10. Mean, median and standard deviation of docking score for all compounds with Tanimoto similarity scores greater than 0.85.

Entry	Mean	Median	St Dev	Docking Score
Caftaric Acid	-5.7	-5.3	1.8	-10.2
Desmosine	-9.3	-9.9	1.2	-10.1
Labetalol	-7.4	-7.7	0.5	-8.1
Kanamycin	-7.5	-7.5	0.7	-7.7

CHAPTER 7
EXPERIENCES OF A BLIND CHEMISTRY GRADUATE STUDENT

The following chapter has been adapted from *Implementation of Protocols To Enable Doctoral Training in Physical and Computational Chemistry of a Blind Graduate Student,* published in the Journal of Chemical Education in 2015 (http://pubs.acs.org/doi/abs/10.1021/ed5009552)[85], with the permission of the American Chemical Society. Authored by Micheal N. Weaver, Jim Gorske, Clifford R. Bowers, and Kenneth M. Merz Jr..

7.1 Introduction

The World Health Organization estimates 285 million persons have visual impairments in the world, with 39 million completely blind.[86] Until recently, primary educators have been apprehensive in providing a complete science, technology, engineering and mathematics (STEM) education to blind students, whether that was because of a perceived difficulty in teaching or learning, or from a safety perspective.[87] Effort to adapt laboratory exercises for blind college chemistry students, such as titration experiments utilizing audio and tactile (Braille) cues, is present in the literature from four decades past, but this early development appears to have been limited to an as-needed basis.[88] More recent efforts have been made to adapt general chemistry lectures and laboratories for undergraduate students.[89] Tactile diagrams convey scientific imagery such as chemical structures to blind users.[90] Tactile learning aids such as molecular models precisely relate structural information to introductory organic chemistry students.[91] Other efforts have focused on vibration to convey information from chemical instruments.[92] Recently, Pereira and coworkers introduced a web portal allowing the construction of simple organic structures for blind users.[93] Still, accessible resources are limited compared to the information readily available to sighted students and researchers.

The National Center for Blind Youth in Science (NCBYS) initiative has sought to alter this paradigm since 2004.[94] Programs are available for elementary through high school students and basic science resources are becoming more readily available. This underscores a need for advances at both the collegiate and graduate school levels. As more individuals with visual impairments gain a high quality basic science education, the desire to pursue advanced degrees and career paths requiring STEM training will continue to increase. While any influx of blind and low-vision students into graduate programs is likely several years down the road, there is a stark need for documentation of tools and techniques useful for these future students.

Herein we document the first four years of graduate work in computational chemistry at the University of Florida completed by Mona Minkara. At age seven, Mona was diagnosed with macular degeneration and cone-rod dystrophy. The progression of her condition has rendered her with no vision in her right eye and two percent vision in her left. She is able to read very large print, albeit at a slow pace, and does not know Braille, thus limiting options for gaining access to critical materials for graduate school success. The purposes of this article are threefold: (1) to document the methods utilized to deliver large volumes of audio and visual information to Mona, (2) to point out considerations and plans that have or should have been implemented to facilitate her work, and (3) to offer suggestions to future blind graduate students and their research advisors. The recent development of resources available to blind and low-vision youths in the sciences will surely result in a larger number of visually impaired students seeking a post-secondary education in chemistry and we hope this account will serve as a guide for future graduate students, advisors, and disability offices.

7.2 Student Perspective

7.2.1 Preliminary Considerations

Before starting my education at the University of Florida (UF), I attempted to start the hiring process for readers before my arrival. This was difficult to accomplish and failed. Therefore, upon my arrival at UF, I was desperate not to fall behind or fail in my first semester of classes. Frantically, with the aid of UF's Disability Resource Center (DRC), a job description was posted in a jobs listing forum. The initial applicants were not suitable due in large part to (1) the lack of time to reach a large number of appropriate applicants and (2) the lack of computer skills of those who did apply. Only two applications were received from the initial posting. One applicant was decent, but informed us that due to his age the job seemed too energetically demanding. The second applicant stated that she had previous lab experience and had done research in the past. However, once hired, it was clear on the very first day that she had no computer skills whatsoever. This made the simple task of even filling out a form online completely impossible. I required a capable reader and in order to locate one, I needed a properly constructed advertisement placed in the right locations with an appropriate amount of time before the selection process began. The hiring process cannot be an afterthought.

In subsequent years, word of mouth helped me, since people knew of a blind chemistry graduate student who would require paid assistance. Tailored job flyers assisted in obtaining a larger applicant pool of approximately 30 people. One part of identifying good candidates was asking them to describe images that I often deal with in my research. Figure (7-1) depicts a ligand docked in the active site of *Klebsiella aerogenes* (KA) urease. I asked the applicants to describe that image to the best of their ability. To me, it was more important for the interviewers to describe everything accurately than for them to be able to tell me the specific name of the chemical compound by recognition. For example, with this image, it is important to read the title

and state that there exists a compound centered on the page, surrounded by multicolored circles

with three letter codes and a four-digit number. It is also very important to state that there are

different line styles: arrowed, solid and dashed. Below is a typical example of the descriptive

information required from readers:

> Title of image: 1finalkawideopen-ZINC00388081. Holding the page horizontally,
> there is a compound in the middle surrounded by multicolored circles. Some of
> the components of the rings have arrows pointing to some colored circles. The
> compound seems to be made of 2 rings in the middle, the ring on the bottom is
> open on 2 sides on the left. From the first ring there is a red point leading to
> ARG2082, a purple arrow pointing from HN^+ to ASP1967 and a purple arrow
> pointing from HN to H20. From the bottom ring, there is a purple arrow pointing
> from H_3N^+ to ALA2109. The two aromatic rings are surrounded on the left side
> by a greenish/blue line, but it starts to fade as it goes to the right side of the image.
> Also surrounding the aromatic rings and the line are labels HIS, ALA, VAL,
> LEU, MET.

Most people do not even notice the title at the top of the page, or mention that there are

multi colored circles! Other typical omissions include differences, like the fact that some are

dashed and some are solid, or that the arrows exist at all (i.e. information on directionality is not

conveyed). The ability of someone to verbally convey descriptions such as these is more

important to me than the ability of someone to understand the biochemistry.

Figure 7-1. Representative 2-Dimensional diagram of ligand interacting with *K. aerogenes* urease.

Additional considerations required attention before addressing my academic situation, namely the problem of living in a new city with no contacts. I was assigned a mobility trainer the first few months I was in Florida, who was amazing. When I initially moved into my apartment, simply leaving and returning home were immensely difficult due to the extremely complicated setup of the complex. I remember feeling completely trapped, not even knowing where the bus stop was located. My mobility trainer was phenomenal on the first day of classes. She took me to my first class, stayed with me and showed me where my different classrooms were located, and took the time to drive me around the normal bus route. She showed me every bus stop around my

complex and she showed me that I lived across from a shopping center. However, I was unable to get to the shopping center on my own, as it required me to cross a major intersection. She suggested I contact the city council and request the placement of a talking crosswalk and after several months this was installed. Now, with the addition of the talking cross walk, I can get my groceries by myself, which is huge.

Another individual that played a crucial role in my success, especially in my home life, was my occupational therapist. I started preparations for the transition to graduate school and learned different living techniques in Massachusetts. I was assigned an occupational therapist who gave me hints regarding home organization. She was extraordinary and even flew down to Florida on her personal vacation to visit me and help me set up. She aided me by placing tactical bumps and the necessary buttons on a microwave, placing tactile bumps on my dryer and washing machine and giving me tips on other house hold items.

Finally, much early consideration should have been given to which classes to register for, who my professors were and where my classes were located during my first semester of graduate school. Getting these details in order ahead of time would have been a tremendous help in obtaining recorded reading material, large print notes and other resources that were not easily accessible until about a month of the semester had elapsed.

7.2.2 Coursework

For any graduate student, a significant amount of coursework is required and success in these classes is also expected in order to matriculate. It is essential that materials be prepared ahead of time in order for a blind student to have any chance at success in these high level courses. Audio resources are lacking in these advanced subjects and a significant failure of PDF readers at this level is the equation density of the texts, particularly in physical and quantum chemistry. For example, it was crucial for me to have class notes, slides and other lecture

materials before the scheduled class time. When such materials were available, I was able to review them before the professor lectured on the topic. For example, during my first semester at UF, instructors provided lecture notes typeset in 28-point font, including equations and figured enlarged to fill an entire page. This would enable me to mentally visualize exactly what was conveyed to the students on the board or through slides during the lecture and become more of an active participant.

An additional necessity was having all of the reading material, such as textbooks, recorded ahead of time in order to keep up with the required reading. This requires a lot of work on whoever is helping due to the extent of preparation that goes into getting the material recorded and placed in an easily accessible format. During class, good note taking by the reader was essential, and these notes were then typed in large font and recorded so that they could be reviewed with or without a reader present.

7.2.3 Cumulative Exams

An integral part of receiving a PhD in many graduate programs is to successfully accumulate enough points to pass a set of cumulative exams designed to test knowledge in the major field as well as evaluate a student's ability to process information in a short window of time. In the Physical Chemistry division at the University of Florida, these exams were set up in a particular manner. In your first year, they are administered once a month, with a requirement that four points be obtained out of six possible exams, where points are earned for either a pass (one) or half pass (one-half). These exams did not have set dates, rather a signup notice was administered and reading material would be issued approximately one week in advance of the test date.

One of the difficulties faced by me as a blind student was to have these materials ready for me to study just like everyone else in the class. Dr. Michael Weaver devised a system in

which he obtained the material ahead of time in order to record the paper, including data tables and descriptions of the figures. With this system, when students were given the topic and relevant literature, I could commence study immediately, regardless of whether a reader was available. Again, recorded materials are critical in allowing studying to occur in the absence of readers, who generally are unavailable from 10pm until 7am. The first time I took one of these cumulative exams, the materials were unavailable ahead of time for recording and I did not pass the exam due to the lag in having access to study materials after hours. Having ready-recorded materials for the exam at the moment the topic was distributed to the rest of the students allowed me the opportunity to succeed. I was able to earn my four passes in the first year, never again failing to earn a full or half-pass on any subsequent examination.

7.2.4 Research

Conducting computational chemistry research as a blind student is extremely difficult, particularly in the realm of molecular dynamics when the research targets are large proteins. One of the primary barriers associated with my research is that another individual must first visualize everything. As a blind chemist, I require the assistance of readers constantly as they represent the main conduit between research results and myself. I cannot have a direct look at the results or data in a reasonable amount of time, which makes research analysis a very frustrating and time-consuming process. My first research project as a graduate student entailed molecular dynamics simulations on the *Helicobacter pylori* (HP) urease[3, 6, 95] enzyme and the subsequent analysis of these results.

HP urease is an enzyme with approximately 148,000 atoms (9744 residues) and the simulation was run in explicit water, adding another 150,000 atoms to the system (Figure (7-2)). This massive system is difficult for a sighted person to analyze, let alone a blind individual. One of the techniques I used was to detail motion in terms distance between set residue pairs as

shown in Figure (7-3). It was very important for the reader to be able to inform me that there are three lines at varying heights along the y-axis. Seemingly trivial pieces of information such as titles, axis labels and units are also of critical importance in the reader's description. I would ask the readers to describe any general trends: for example, the green line has a steep upward climb from 100 to 200 ns. This is a mathematical approach to understanding molecular motion, utilizing a reader's detailing of minima, maxima, plateaus and other features on a Cartesian plot. This is much easier for a reader to describe than the often-complex motions observed in an MD trajectory.

Another technique I utilized was to continually develop hypotheses and subsequently utilize precise questions posed to a sighted person in order to test these. I additionally asked readers to visually describe to the best of their abilities what they were seeing, starting with a general picture and then honing in with my questions to obtain a more detailed understanding. A relatable analogy is drawing an outline of what is going on and then slowly filling in details. Other useful techniques involved the use of play-dough to build a pseudo-structure of the protein so I could get an understanding of what was visually going on. It took time and effort, but eventually I understood the structure so well that I am now able to determine the location of most amino acids based on the sequence numbering. This particular protein has twelve identical parts, so coloring the different parts distinctly using Visual Molecular Dynamics[96] allowed me to gain a better understanding of how the pieces fit with one another without the need to identify individual residues. When I had a general idea of how this protein looked, I was able to slowly fill in finer detail. I became adept at knowing which amino acids were present based on the sequence name and where they might be on the protein. The patience exercised and techniques learned throughout my first two years of research on this enzyme culminated in the publication

of a study on the motion of active site-covering flaps of HP urease in the *Journal of Chemical Theory and Computation*.[46]
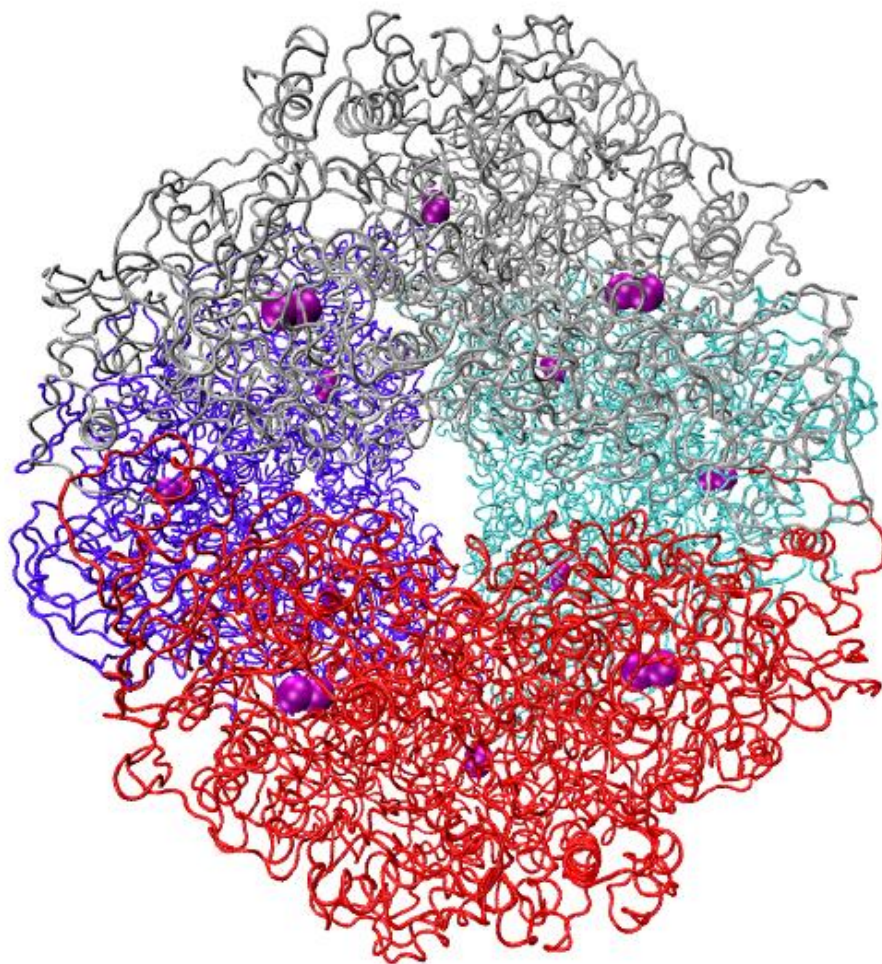
Figure 7-2. Representation of the *H. pylori* urease enzyme.
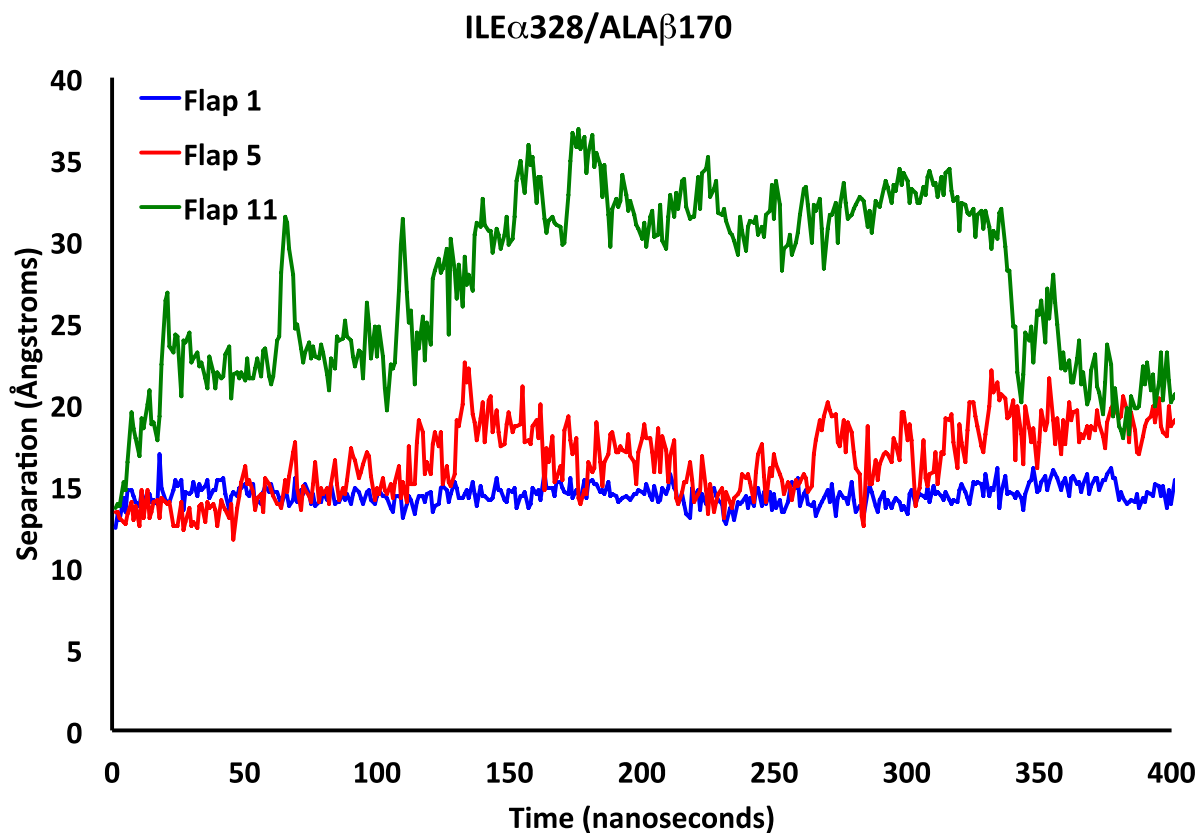
Figure 7-3. Residue separations in HP urease used to quantify extent of flap opening.

7.2.5 Teaching

As a blind graduate student, I had a very interesting time teaching for the general chemistry class provided here at the University of Florida (CHM2045). Initially, some individuals were apprehensive about how I would be able to teach discussion sections of the course. In the end, the students loved me and the chemistry department asked me to teach again. I had decent attendance throughout the semester even though it was not required. The way I approached teaching was with the advice of Dr. Michael Weaver. I prepared all of the problems and solutions for the different topics on PowerPoint slides so that for every discussion class, I had a set of slides to teach from. I studied these slides very well to prepare for possible questions asked by students. I also had a list of topics covered in the slide show at the beginning of each

PowerPoint and I would ask the class which topics they wanted to cover. I was very interactive; I not only lectured, but also included the students in the discussion. I informed the class I was blind and could not call on people in the traditional manner, they had to speak out and be very vocal about their questions. They told me that they really enjoyed this new way of learning due to the interactive quality. I had students go the board to work out problems and I would have the rest of the class watch the person on the board and we would work through the problem audibly. I was very clear about what should be on the board, and later provided slides containing clear reproductions of the board's content97.. I also had students read the questions aloud. The students highly enjoyed this process; one student told me, "This class always goes by so fast," a sign of student interest. This experience opened me up to the idea of teaching and also made me a better chemist.

7.2.6 Disability Resource Center

Choosing a school with a great DRC is paramount for the success of a blind graduate student in chemistry. The DRC is truly amazing here at the University of Florida in terms of the equipment provided and hiring of readers in my support. I feel in order to ensure the success of a blind student, there needs to be a disability office affiliated with the institution that ultimately has the student's best interest at heart and supports the student in whatever the student requires.

7.3 Postdoctoral Perspective

Numerous aspects of assisting in the training and providing day-to-day support render the task both difficult and extremely time consuming. Especially when coursework was in progress, the volume of work to be typed, formatted, recorded and uploaded was constantly accumulating. Due to the finite time length associated with a semester-long graduate course, preparation of these study materials often entailed 12 to 16 hour workdays, despite numerous graduate and undergraduate students who aided in recording and daily work. Furthermore, when research was

added to the mix, the recording of background materials, current literature related to the research project(s) and technical manuals were added to the list of items to prepare. Cumulative exams added yet another time-sensitive element. Exam materials were generally made available for recording 2-3 days prior to the topic announcement and distribution of resources to the graduate student population, requiring an extremely fast turn-around without allowing preparation of class materials to slip.

A principle component of ensuring resources were available for Mona's success was the recording and distribution of research, course, cumulative exam and other review material. All recorded materials were placed on a secure website, accessible anywhere at any time. Reading for coursework was obtained as early as possible in order to have it prepared for the upcoming semester, with assignments obtained from the professor (and not disclosed to the student) well ahead of time. Material for cumulative exams was typically obtained a few days prior to release of the information to the graduate student population in order to have files recorded and a webpage prepared. In this manner, at the instant the topic was announced, Mona would have the same immediate access to the readings as the other graduate students. The link was kept dead and the directory password protected until the professor in charge of the exam released the topic.

A postdoctoral associate working in the capacity described herein has a bifurcated role. The traditional role as a mentor to graduate students is still applicable and there is the added role of serving as eyes for the student in order to convey information. A clear delineation must be made between which role is being served at which time. Too little mentoring results in the blind or low-vision student being deprived the same learning opportunities that other graduate students receive. Too much mentoring could result in less of the work of the blind student being his or her own intellectual product. Both cases are detrimental to the educational process, and in the

postdoctoral role diligent attention must always be paid to which hat is required and/or appropriate for the task at hand.

7.4 Advisor Perspective

The process by which a graduate student selects a Ph.D. advisor is a task not to be taken lightly. Issues ranging from research area to the manner in which the advisor runs his or her group are critical in making the right choice. The same goes for the potential advisor: he/she has to assess whether or not the prospective graduate student is a good fit for their research group. In Mona's case, she desired a Computational/Chemistry/Biology group and an advisor that was hands-on rather than more aloof, which fit my overall profile quite well.

Mona started in August of 2010 and we had done some preparative work through the summer months to ensure that she got off to a fast start. Even with our best-laid plans the fall semester did not go as smoothly as we had hoped. The need to obtain readers, get Mona's hardware/software needs addressed and identifying her point support person all proved to take a bit longer than expected. The best advice I can provide to another advisor taking on a visually challenged graduate student is to have them arrive at the start of summer before the fall semester to get the student set up appropriately. That said, Mona overall had an excellent fall semester given the challenges faced.

As a graduate student, Mona was largely no different than any of the graduate students I have had. She made mistakes, which we corrected; she has had numerous successes and she will have more; she has had trials and tribulations, for example, her laptop was stolen, but we have overcome these. However, as you might expect there are several major differences with Mona's graduate experience relative to a sighted individual. Probably the biggest difference is the managerial skills a visually challenged individual has to rapidly develop as a graduate student. Mona runs a small staff and dealing with her staff and training them adds additional burden to

214

the graduate experience. In Mona's case she rapidly developed excellent management skills and has developed into a skillful manager of her staff. The second important difference for the advisor is the need to spend time visually reviewing the student's research. In computational biology this means visualizing the protein molecule being simulated and confirming what the student has observed using the aid of both experienced and inexperienced visual support staff. Overall, Mona has an intense drive to succeed and is simply a joy to work with. I have personally learned a lot from Mona and look forward to her receiving the Ph.D. in the coming months.

7.6 Disability Resource Center Perspective

The support of the Disability Resource Center is important for any student with a visual impairment. Support is *critical* for a student with a visual impairment in an academic program where technology does not provide the requisite level of access to course and other materials (e.g. textbooks, hard copy resources, homework sets, exams). From the perspective of the DRC, the following three items were or have been important in helping our office ensure that Mona's accommodations and services are implemented and maintained over time.

As an applicant, Mona visited the graduate program at each of the schools she was applying. She also visited the DRC, in order to get a sense of the level of support each was willing to offer and provide. While all post-secondary institutions provide a mechanism to support students with disabilities, there is variability in the manner in which these supports may be coordinated, funded, and provided. When she met with our office she shared her insight on what accommodations and supports have been required and/or provided in the past and what she felt would be needed as a student in a graduate program. This initial connection was critical and gave our office a heads-up should Mona choose the University of Florida for her graduate studies. After she accepted, she visited our office again to review and discuss her support needs. This visit was important, as significant advance coordination would be necessary (e.g. recruiting

215

and hiring support staff) to ensure that staffing support could potentially be available when she began her program.

The task of recruiting and hiring qualified support staff was daunting. Our initial approach was to have a smaller pool of staff, each providing a larger number of hours. To that end we looked to hire non-student community members. While we had a decent response as a result of community connections (e.g. Vocational Rehabilitation, Division of Blind Services), the staff we hired, although qualified on paper, did not have the requisite technical skills and/or sense of connection to the intensity and demands of the program's coursework. Recruiting through the Human Resources job board and departmental listservs yielded better results. However, it was Mona's handmade job postings within her department that over time yielded the best results. We also realized that it was important to recruit for specific tasks (e.g. reader/recorder, note-taker, assistance with homework sets, test scribe/proctor) rather than attempting to find one or two individuals who could provide broader support.

Support coordination meetings were held each semester during Mona's first two years in the program, when the focus of her program was primarily academic. The meetings were typically held at the end of the semester so that the instructors for the following semester could attend. Additionally, Mona's graduate coordinator, her advisor, her postdoctoral staff member, chemistry departmental support staff, the director of the High-Performance Computing Center as well as the DRC attended the meetings. The meetings provided an excellent opportunity to review what went well and what did not during the current semester, review current and future staffing needs, and afford an opportunity for current instructors to engage future instructors.

7.7 Concluding Remarks

A successful graduate education for Mona has required an intense, coordinated effort on the part of many different groups as well as a strong commitment by Mona herself. Critical in

Mona's success was planning ahead to ensure that resources were developed and ready in advance for exams, research and teaching activities. With these resources in place, Mona was able to perform at or above the level of the average graduate student in all facets of her PhD training. While certain speed bumps were encountered along the way, this is to be expected of any student; it was simply the nature of said impediments that differed slightly from the norm. A critical component of Mona's graduate education was not lowering expectations due to her disability. With proper foresight and planning, she was able to complete all requisite components of the program to date according to standard departmental schedule. We hope that this account will aid in the future education of blind and low-vision students who aspire to higher education in chemistry and other STEM related fields.

REFERENCES

1.   (a) Marshall, B., Unidentified Curved Bacilli on Gastric Epithelium in Active Chronic Gastritis. *Lancet* **1983,** *1* (8336), 1273-1275; (b) Marshall, B. J., The Pathogenesis of Non-Ulcer Dyspepsia. *Med J Australia* **1985,** *143* (7), 319-319; (c) Marshall, B. J.; Warren, J. R., Unidentified Curved Bacilli in the Stomach of Patients with Gastritis and Peptic-Ulceration. *Lancet* **1984,** *1* (8390), 1311-1315.

2.   Buck, G. E., Campylobacter-Pylori and Gastroduodenal Disease. *Clin Microbiol Rev* **1990,** *3* (1), 1-12.

3.   Karplus, P. A.; Pearson, M. A.; Hausinger, R. P., 70 years of crystalline urease: What have we learned? *Accounts Chem Res* **1997,** *30* (8), 330-337.

4.   (a) Mobley, H. L. T.; Island, M. D.; Hausinger, R. P., Molecular-Biology of Microbial Ureases. *Microbiol Rev* **1995,** *59* (3), 451-480; (b) Hu, L. T.; Mobley, H. L. T., Purification and N-Terminal Analysis of Urease from Helicobacter-Pylori. *Infect Immun* **1990,** *58* (4), 992-998.

5.   Sumner, J. B., The isolation and crystallization of the enzyme urease. Preliminary paper. *J Biol Chem* **1926,** *69* (2), 435-441.

6.   Ha, N. C.; Oh, S. T.; Sung, J. Y.; Cha, K. A.; Lee, M. H.; Oh, B. H., Supramolecular assembly and acid resistance of Helicobacter pylori urease. *Nat Struct Biol* **2001,** *8* (6), 505-509.

7.   (a) Zerner, B., Recent Advances in the Chemistry of an Old Enzyme, Urease. *Bioorg Chem* **1991,** *19* (1), 116-131; (b) Dixon, N. E.; Gazzola, C.; Blakeley, R. L.; Zerner, B., Jack-Bean Urease (Ec 3.5.1.5) - Metalloenzyme - Simple Biological Role for Nickel. *J Am Chem Soc* **1975,** *97* (14), 4131-4133.

8.   (a) Carter, E. L.; Proshlyakov, D. A.; Hausinger, R. P., Apoprotein isolation and activation, and vibrational structure of the Helicobacter mustelae iron urease. *J Inorg Biochem* **2012,** *111*, 195-202; (b) Carter, E. L.; Tronrud, D. E.; Taber, S. R.; Karplus, P. A.; Hausinger, R. P., Iron-containing urease in a pathogenic bacterium. *P Natl Acad Sci USA* **2011,** *108* (32), 13095-13099.

9.   Pinkse, M. W. H.; Maier, C. S.; Kim, J. I.; Oh, B. H.; Heck, A. J. R., Macromolecular asserably of Helicobacter pylori urease investigated by mass spectrometry. *J Mass Spectrom* **2003,** *38* (3), 315-320.

10.   Dixon, N. E.; Gazzola, C.; Watters, J. J.; Blakeley, R. L.; Zerner, B., Inhibition of Jack-Bean Urease (Ec 3.5.1.5) by Acetohydroxamic Acid and by Phosphoramidate - Equivalent Weight for Urease. *J Am Chem Soc* **1975,** *97* (14), 4130-4131.

11.   Xiao, Z. P.; Wang, X. D.; Peng, Z. Y.; Huang, S.; Yang, P.; Li, Q. S.; Zhou, L. H.; Hu, X. J.; Wu, L. J.; Zhou, Y.; Zhu, H. L., Molecular Docking, Kinetics Study, and Structure-Activity Relationship Analysis of Quercetin and Its Analogous as Helicobacter pylori Urease Inhibitors. *J Agr Food Chem* **2012,** *60* (42), 10572-10577.

12.   Matsubara, S.; Shibata, H.; Ishikawa, F.; Yokokura, T.; Takahashi, M.; Sugimura, T.; Wakabayashi, K., Suppression of Helicobacter pylori-induced gastritis by green tea extract in Mongolian gerbils. *Biochem Bioph Res Co* **2003,** *310* (3), 715-719.

13.   Dominguez, M. J.; Sanmartin, C.; Font, M.; Palop, J. A.; Francisco, S. S.; Urrutia, O.; Houdusse, F.; Garcia-Mina, J. M., Design, synthesis, and biological evaluation of phosphoramide derivatives as urease inhibitors. *J Agr Food Chem* **2008,** *56* (10), 3721-3731.

14.   Abid, O. U. R.; Babar, T. M.; Ali, F. I.; Ahmed, S.; Wadood, A.; Rama, N. H.; Uddin, R.; Zaheer-ul-Haq; Khan, A.; Choudhary, M. I., Identification of Novel Urease Inhibitors by High-Throughput Virtual and in Vitro Screening. *Acs Med Chem Lett* **2010,** *1* (4), 145-149.

15.   Benini, S.; Kosikowska, P.; Cianci, M.; Mazzei, L.; Vara, A. G.; Berlicki, L.; Ciurli, S., The crystal structure of Sporosarcina pasteurii urease in a complex with citrate provides new hints for inhibitor design. *J Biol Inorg Chem* **2013,** *18* (3), 391-399.

16.   Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *J Chem Phys* **1953,** *21* (6), 1087-1092.

17.   Alder, B. J.; Wainwright, T. E., Studies in Molecular Dynamics .1. General Method. *J Chem Phys* **1959,** *31* (2), 459-466.

18.   Rahman, A., Correlations in Motion of Atoms in Liquid Argon. *Phys Rev a-Gen Phys* **1964,** *136* (2A), A405-&.

19.   Cramer, C. J., *Essentials of computational chemistry : theories and models*. 2nd ed.; Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2004; p xx, 596 p.

20.   (a) Verlet, L., Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* **1967,** *159* (1), 98-&; (b) Verlet, L., Computer Experiments on Classical Fluids .2. Equilibrium Correlation Functions. *Phys Rev* **1968,** *165* (1), 201-&.

21.   (a) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvai, I.; Wong, K. F.; Paisani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hseih, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 12. *AMBER 12* **2012**, University of California, San Francisco; (b) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvai, I.; Wong, K. F.; Paisani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hseih, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 11. *AMBER 11* **2010**, University of California, San Francisco.

22.   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J Am Chem Soc* **1995,** *117* (19), 5179-5197.

23.   (a) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J Am Chem Soc* **1984,** *106* (3), 765-784; (b) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *J Comput Chem* **1986,** *7* (2), 230-252.

24.   (a) Singh, U. C.; Kollman, P. A., An Approach to Computing Electrostatic Charges for Molecules. *J Comput Chem* **1984,** *5* (2), 129-145; (b) Besler, B. H.; Merz, K. M.; Kollman, P. A., Atomic Charges Derived from Semiempirical Methods. *J Comput Chem* **1990,** *11* (4), 431-439.

25.   Momany, F. A.; Mcguire, R. F.; Burgess, A. W.; Scheraga, H. A., Energy Parameters in Polypeptides .7. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen-Bond Interactions, and Intrinsic Torsional Potentials for Naturally Occurring Amino-Acids. *J Phys Chem-Us* **1975,** *79* (22), 2361-2381.

26.   Lavery, R.; Zakrzewska, K.; Sklenar, H., Jumna (Junction Minimization of Nucleic-Acids). *Comput Phys Commun* **1995,** *91* (1-3), 135-158.

27.   Clark, M.; Cramer, R. D.; Vanopdenbosch, N., Validation of the General-Purpose Tripos 5.2 Force-Field. *J Comput Chem* **1989,** *10* (8), 982-1012.

28.   Vedani, A.; Huhta, D. W., A New Force-Field for Modeling Metalloproteins. *J Am Chem Soc* **1990,** *112* (12), 4759-4767.

29.   (a) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996,** *118* (45), 11225-11236; (b) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* **2001,** *105* (28), 6474-6487.

30.   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **1983,** *4* (2), 187-217.

31.   (a) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F., The GROMOS biomolecular simulation program package. *J Phys Chem A* **1999,** *103* (19), 3596-3607; (b) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J Biophy* **2011,** *40* (7), 843-856; (c) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F., Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput Phys Commun* **2012,** *183* (4), 890-903.

32.   (a) Allinger, N. L., Conformational-Analysis .130. Mm2 - Hydrocarbon Force-Field Utilizing V1 and V2 Torsional Terms. *J Am Chem Soc* **1977,** *99* (25), 8127-8134; (b) Allinger, N. L.; Profeta, S., An Mm2 Force-Field Treatment of Aliphatic-Amines - Structures and Conformational Energies of Simple Cyclic and Acyclic Systems. *Abstr Pap Am Chem S* **1980,** *180* (Aug), 49-ORGN.

33.   (a) Allinger, N. L.; Yuh, Y. H.; Lii, J. H., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .1. *J Am Chem Soc* **1989,** *111* (23), 8551-8566; (b) Lii, J. H.; Allinger, N. L., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .3. The Vanderwaals Potentials and Crystal Data for Aliphatic and Aromatic-Hydrocarbons. *J Am Chem Soc* **1989,** *111* (23), 8576-8582; (c) Lii, J. H.; Allinger, N. L., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .2. Vibrational Frequencies and Thermodynamics. *J Am Chem Soc* **1989,** *111* (23), 8566-8575.

34.   Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations. *J Am Chem Soc* **1992,** *114* (25), 10024-10035.

35.   (a) Halgren, T. A., Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* **1996,** *17* (5-6), 490-519; (b) Halgren, T. A., Merck molecular force field .2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J Comput Chem* **1996,** *17* (5-6), 520-552; (c) Halgren, T. A., Merck molecular force field .3. Molecular geometries and vibrational frequencies for MMFF94. *J Comput Chem* **1996,** *17* (5-6), 553-586; (d) Halgren, T. A.; Nachbar, R. B., Merck molecular force field .4. Conformational energies and geometries for MMFF94. *J Comput Chem* **1996,** *17* (5-6), 587-615; (e) Halgren, T. A., Merck molecular force field .5. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J Comput Chem* **1996,** *17* (5-6), 616-641.

36.   Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-725.

37.   Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* **1977,** *23* (3), 327-341.

38.   Roothaan, C. C. J., New Developments in Molecular Orbital Theory. *Rev Mod Phys* **1951,** *23* (2), 69-89.

39.   (a) Becke, A. D., Density-Functional Thermochemistry .3. The Role of Exact Exchange. *J Chem Phys* **1993,** *98* (7), 5648-5652; (b) Lee, C. T.; Yang, W. T.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys Rev B* **1988,** *37* (2), 785-789; (c) Vosko, S. H.; Wilk, L.; Nusair, M., Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin-Density Calculations - a Critical Analysis. *Can J Phys* **1980,** *58* (8), 1200-1211; (d) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *J Phys Chem-Us* **1994,** *98* (45), 11623-11627.

40.   (a) Foster, J. P.; Weinhold, F., Natural Hybrid Orbitals. *J Am Chem Soc* **1980,** *102* (24), 7211-7218; (b) Reed, A. E.; Weinhold, F., Natural Bond Orbital Analysis of near-Hartree-Fock Water Dimer. *J Chem Phys* **1983,** *78* (6), 4066-4073; (c) Reed, A. E.; Weinhold, F.; Curtiss, L. A.; Pochatko, D. J., Natural Bond Orbital Analysis of Molecular-Interactions - Theoretical-Studies of Binary Complexes of Hf, H2o, Nh3, N-2, O-2, F2, Co, and Co2 with Hf, H2o, and Nh3. *J Chem Phys* **1986,** *84* (10), 5687-5705.

41.   Bohm, H. J., The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *J Comput Aid Mol Des* **1994,** *8* (3), 243-256.

42.   Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R., VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J Am Chem Soc* **1996,** *118* (16), 3959-3969.

43.   Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aid Mol Des* **1997,** *11* (5), 425-445.

44.   (a) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** *47* (7), 1739-1749; (b) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **2004,** *47* (7), 1750-1759.

45.   Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006,** *49* (21), 6177-6196.

46.   Minkara, M. S.; Ucisik, M. N.; Weaver, M. N.; Merz, K. M., Molecular Dynamics Study of Helicobacter pylori Urease. *J Chem Theory Comput* **2014,** *10* (5), 1852-1862.

47.   Musiani, F.; Arnofi, E.; Casadio, R.; Ciurli, S., Structure-based computational study of the catalytic and inhibition mechanisms of urease. *J Biol Inorg Chem* **2001,** *6* (3), 300-314.

48.   (a) Estiu, G.; Merz, K. M., Enzymatic catalysis of urea decomposition: Elimination or hydrolysis? *J Am Chem Soc* **2004,** *126* (38), 11832-11842; (b) Estiu, G.; Merz, K. M., Catalyzed decomposition of urea. Molecular dynamics simulations of the binding of urea to urease. *Biochemistry-Us* **2006,** *45* (14), 4429-4443; (c) Estiu, G.; Suarez, D.; Merz, K. M., Quantum mechanical and molecular dynamics simulations of ureases and Zn beta-lactamases. *J Comput Chem* **2006,** *27* (12), 1240-1262.

49.   Carlsson, H.; Nordlander, E., Computational Modeling of the Mechanism of Urease. *Bioinorg Chem Appl* **2010**.

50.   Valdez, C. E.; Alexandrova, A. N., Why Urease Is a Di-Nickel Enzyme whereas the CcrA beta-Lactamase Is a Di-Zinc Enzyme. *J Phys Chem B* **2012,** *116* (35), 10649-10656.

51.   Barros, P. R.; Stassen, H.; Freitas, M. S.; Carlini, C. R.; Nascimento, M. A. C.; Follmer, C., Membrane-disruptive properties of the bioinsecticide Jaburetox-2Ec: Implications to the mechanism of the action of insecticidal peptides derived from ureases. *Bba-Proteins Proteom* **2009,** *1794* (12), 1848-1854.

52.   Manunza, B.; Deiana, S.; Pintore, M.; Gessa, C., The binding mechanism of urea, hydroxamic acid and N-(N-butyl)-phosphoric triamide to the urease active site. A comparative molecular dynamics study. *Soil Biol Biochem* **1999,** *31* (5), 789-796.

53.   Roberts, B. P.; Miller, B. R.; Roitberg, A. E.; Merz, K. M., Wide-Open Flaps Are Key to Urease Activity. *J Am Chem Soc* **2012,** *134* (24), 9934-9937.

54.   Goetz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* **2012,** *8* (5), 1542-1555.

55.   The PyMOL Molecular Graphics System, V. r., Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

56.   (a) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **2012,** *40* (W1), W537-W541; (b) Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* **2006,** *63* (4), 928-938; (c) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **2005,** *33*, W368-W371.

57.   Roberts, B. P.; Miller, B. E.; Roitberg, A. E.; Merz, K. M., Wide-Open Flaps are Key to Urease Activity. *J Am Chem Soc* **2012,** *Submitted for publication*.

58.   Merz, K. M., Co2 Binding to Human Carbonic Anhydrase-Ii. *J Am Chem Soc* **1991,** *113* (2), 406-411.

59.   Toba, S.; Colombo, G.; Merz, K. M., Solvent dynamics and mechanism of proton transfer in human carbonic anhydrase II. *J Am Chem Soc* **1999,** *121* (10), 2290-2302.

60.   (a) Pace, N. C.; Tanford, C., Thermodynamics of Unfolding of Beta-Lactoglobulin a in Aqueous Urea Solutions between 5 and 55 Degrees. *Biochemistry-Us* **1968,** *7* (1), 198-&; (b) Creighton, T. E., Effects of Urea and Guanidine.Hcl on Folding and Unfolding of Pancreatic Trypsin-Inhibitor. *J Mol Biol* **1977,** *113* (2), 313-328.

61.   Muthuselvi, L.; Miller, R.; Dhathathreyan, A., How does urea really denature myoglobin? *Chem Phys Lett* **2008,** *465* (1-3), 126-130.

62.   TiradoRives, J.; Orozco, M.; Jorgensen, W. L., Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry-Us* **1997,** *36* (24), 7313-7329.

63.   Caflisch, A.; Karplus, M., Structural details of urea binding to barnase: a molecular dynamics analysis. *Struct Fold Des* **1999,** *7* (5), 477-488.

64.   Camilloni, C.; Rocco, A. G.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G., Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophys J* **2008,** *94* (12), 4654-4661.

65.   Cai, Z.; Li, J.; Yin, C.; Yang, Z.; Wu, J.; Zhou, R., Effect of Urea Concentration on Aggregation of Amyloidogenic Hexapeptides (NFGAIL). *J Phys Chem B* **2014,** *118* (1), 48-57.

66.	Moeser, B.; Horinek, D., Unified Description of Urea Denaturation: Backbone and Side Chains Contribute Equally in the Transfer Model. *J Phys Chem B* **2014,** *118* (1), 107-114.

67.	H++ Protonation Server, h. b. c. v. e. H. H++ Protonation Server. http://biophysics.cs.vt.edu/H++.

68.	Price, D. J.; Brooks, C. L., A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys* **2004,** *121* (20), 10096-10103.

69.	Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78* (8), 1950-1958.

70.	Hubbard, S.; Thornton, J. Naccess. http://www.bioinf.manchester.ac.uk/naccess/ (accessed January 10).

71.	Lee, B.; Richards, F. M., Interpretation of Protein Structures - Estimation of Static Accessibility. *J Mol Biol* **1971,** *55* (3), 379-&.

72.	(a) Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G., A series of PDB related databases for everyday needs. *Nucleic Acids Res* **2011,** *39* (Database issue), D411-9; (b) Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-2637.

73.	Candotti, M.; Perez, A.; Ferrer-Costa, C.; Rueda, M.; Meyer, T.; Gelpi, J. L.; Orozco, M., Exploring Early Stages of the Chemical Unfolding of Proteins at the Proteome Scale. *Plos Comput Biol* **2013,** *9* (12).

74.	Wolfenden, R.; Snider, M. J., The depth of chemical time and the power of enzymes as catalysts. *Accounts Chem Res* **2001,** *34* (12), 938-945.

75.	Estiu, L. G.; Merz, K. M., Molecular dynamics simulations of urease catalysis. *Abstr Pap Am Chem S* **2004,** *227*, U1006-U1006.

76.	Roberts, B. P.; Miller, B. R.; Roitberg, A. E.; Merz, K. M., Wide-Open Flaps Are Key to Urease Activity. *J Am Chem Soc* **2012,** *134* (24), 9934-9937.

77.	Macomber, L.; Minkara, M. S.; Hausinger, R. P.; Merz, K. M., Reduction of Urease Activity by Interaction with the Flap Covering the Active Site. *J Chem Inf Model* **2015,** *55* (2), 354-361.

78.	Irwin, J. J.; Shoichet, B. K., ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005,** *45* (1), 177-82.

79.   (a) Jabri, E.; Carr, M. B.; Hausinger, R. P.; Karplus, P. A., The Crystal-Structure of Urease from Klebsiella-Aerogenes. *Science* **1995,** *268* (5213), 998-1004; (b) Pearson, M. A.; Michel, L. O.; Hausinger, R. P.; Karplus, P. A., Structures of Cys319 variants and acetohydroxamate-inhibited Klebsiella aerogenes urease. *Biochemistry-Us* **1997,** *36* (26), 8164-8172.

80.   (a) Benini, S.; Rypniewski, W. R.; Wilson, K. S.; Miletti, S.; Ciurli, S.; Mangani, S., A new proposal for urease mechanism based on the crystal structures of the native and inhibited enzyme from Bacillus pasteurii: why urea hydrolysis casts two nickels. *Struct Fold Des* **1999,** *7* (2), 205-216; (b) Balasubramanian, A.; Ponnuraj, K., Crystal Structure of the First Plant Urease from Jack Bean: 83 Years of Journey from Its First Crystal to Molecular Structure. *J Mol Biol* **2010,** *400* (3), 274-283.

81.   (a) Martin, P. R.; Hausinger, R. P., Site-Directed Mutagenesis of the Active-Site Cysteine in Klebsiella-Aerogenes Urease. *J Biol Chem* **1992,** *267* (28), 20024-20027; (b) Weatherb.Mw, Phenol-Hypochlorite Reaction for Determination of Ammonia. *Anal Chem* **1967,** *39* (8), 971-&; (c) Carter, E. L.; Hausinger, R. P., Characterization of the Klebsiella aerogenes Urease Accessory Protein UreD in Fusion with the Maltose Binding Protein. *J Bacteriol* **2010,** *192* (9), 2294-2304.

82.   Shabana, S.; Kawai, A.; Kai, K.; Akiyama, K.; Hayashi, H., Inhibitory Activity against Urease of Quercetin Glycosides Isolated from Allium cepa and Psidium guajava. *Biosci Biotech Bioch* **2010,** *74* (4), 878-880.

83.   Loes, A. N.; Ruyle, L.; Arvizu, M.; Gresko, K. E.; Wilson, A. L.; Deutch, C. E., Inhibition of urease activity in the urinary tract pathogen Staphylococcus saprophyticus. *Lett Appl Microbiol* **2014,** *58* (1), 31-41.

84.   (a) Tan, L. R.; Su, J. Y.; Wu, D. W.; Yu, X. D.; Su, Z. Q.; He, J. J.; Wu, X. L.; Kong, S. Z.; Lai, X. P.; Lin, J.; Su, Z. R., Kinetics and Mechanism Study of Competitive Inhibition of Jack-Bean Urease by Baicalin. *Sci World J* **2013**; (b) Wu, D. W.; Yu, X. D.; Xie, J. H.; Su, Z. Q.; Su, J. Y.; Tan, L. R.; Huang, X. Q.; Chen, J. N.; Su, Z. R., Inactivation of jack bean urease by scutellarin: Elucidation of inhibitory efficacy, kinetics and mechanism. *Fitoterapia* **2013,** *91*, 60-67.

85.   Minkara, M. S.; Weaver, M. N.; Gorske, J.; Bowers, C. R.; Merz, K. M., Implementation of Protocols to Enable Doctoral Training in Physical and Computational Chemistry of a Blind Gradaute Student. *J. Chem. Educ.* **2015**, ASAP.

86.   http://www.emro.who.int/cpb/facts.htm Control and prevention of blindness and deafness, Accessed March 17 2015.

87.   (a) Crosby, G. A., Chemistry and the Visually Handicapped. *J Chem Educ* **1981,** *58* (3), 206-208; (b) Tombaugh, D., Chemistry and the Visually Impaired - Available Teaching Aids. *J Chem Educ* **1981,** *58* (3), 222-226.

88.   (a) Hiemenz, P. C.; Pfeiffer, E., General Chemistry Experiment for Blind. *J Chem Educ* **1972,** *49* (4), 263-&; (b) Tallman, D. E., Ph Titration Apparatus for Blind Student. *J Chem Educ* **1978,** *55* (9), 605-606.

89.   (a) Supalo, C., Techniques to enhance instructors' teaching effectiveness with chemistry students who are blind or visually impaired. *J Chem Educ* **2005,** *82* (10), 1513-1518; (b) Supalo, C. A.; Mallouk, T. E.; Amorosi, C.; Lanouette, J.; Wohlers, H. D.; McEnnis, K., Using Adaptive Tools and Techniques To Teach a Class of Students Who Are Blind or Low-Vision. *J Chem Educ* **2009,** *86* (5), 587-591; (c) Supalo, C. A.; Mallouk, T. E.; Rankel, L.; Amorosi, C.; Graybill, C. M., Low-cost laboratory adaptations for precollege students who are blind or visually impaired. *J Chem Educ* **2008,** *85* (2), 243-247.

90.   Rawls, R., Tactile diagrams bring chemical structures to blind students. *Chem Eng News* **1999,** *77* (33), 39-41.

91.   Poon, T.; Ovadia, R., Using tactile learning aids for students with visual impairments in a first-semester organic chemistry course. *J Chem Educ* **2008,** *85* (2), 240-242.

92.   Anderson, J. L., Chemical Instrumentation for the Visually Handicapped. *J Chem Educ* **1982,** *59* (10), 871-872.

93.   Pereira, F.; Aires-de-Sousa, J.; Bonifacio, V. D. B.; Mata, P.; Lobo, A. M., MOLinsight: A Web Portal for the Processing of Molecular Structures by Blind Students. *J Chem Educ* **2011,** *88* (3), 361-362.

94.   http://www.blindscience.org National Center for Blind Youth in Science, Accessed March 17 2015.

95.   Dunn, B. E.; Cohen, H.; Blaser, M. J., Helicobacter pylori. *Clin Microbiol Rev* **1997,** *10* (4), 720-&.

96.   Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J Mol Graph Model* **1996,** *14* (1), 33-38.

BIOGRAPHICAL SKETCH

Mona Minkara was born in August of 1987. She was diagnosed with macular degeneration and cone rod dystrophy at the age of 7. Mona grew up in the greater Boston area and those surrounding her, including her teachers, were skeptical of how far her education would go due to the degenerating form of blindness she was born with. She attended Hingham High School, taking standard level classes initially, but eventually working her way up to advanced classes, facing many obstacles from administration along the way. She was involved in the honor society and graduated in the top fifth percent of her class. After high school, she attended the prestigious Wellesley College. During her time at Wellesley College, she double majored in Chemistry and Middle Eastern Studies. Mona conducted research as an undergraduate, and obtained a Research Experiences for Undergraduates (REU) from the National Science Foundation (NSF) each summer of her undergraduate career. Her research experience at Wellesley College culminated in her receiving the post-baccalaureate fellowship from the Howard Hughes Medical Institute (HHMI). This fellowship allowed her to conduct research in computational chemistry and thoroughly consider graduate programs before entering a PhD program. She chose to attend the University of Florida in Gainesville, Florida and completed five years of graduate study in their Chemistry department. Her doctoral research focused on the solution structure of *Helicobacter pylori* urease (HPU) and efforts aimed at identifying potential inhibitors of the enzyme. The HPU enzyme facilitates the survival of *Helicobacter pylori* bacteria in the stomach lining of approximately 2/3rd of the world's population. This bacterium has been implicated in an array of gastrointestinal disorders including gastric and duodenal ulcers and adenocarcinoma. With the support of her family, friends, and the disability resource centers at the institutions she has attended, she has been more than successful in her endeavors. She graduated with her PhD in Chemistry in August of 2015.