

DEVELOPING A COMPUTATIONAL MODEL OF ASPARAGINE SYNTHETASE-B
TOWARDS RATIONAL INHIBITOR DESIGN

By

ROBERT N. HUMKEY

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

2009

© 2009 Robert N. Humkey

To my grandfather, Charles Robert Ruyle, and my entire family

ACKNOWLEDGMENTS

This work was completed due in large part to the wonderful support I have received from all of my family and friends. I have to first thank my parents, John and Lynne Humkey, for all the love and support they have shown me in whatever I have done. I would like to thank my brothers, Travis and Greg, and sister, Katie, for their continued encouragement along the way. I would also like to thank my advisor, Dr. Richards, and all of the Richards' group, especially Sangbae Lee and Megan Meyer, for their encouragement and discussions that have helped with this work. I would also like to thank Lori Clark for her continued moral support and assistance guiding me through my time here at UF.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	4
LIST OF TABLES	7
LIST OF FIGURES	8
ABSTRACT	11
CHAPTER	
1 INTRODUCTION	13
Acute Lymphoblastic Leukemia.....	13
Asparagine Synthetase.....	14
Glutamine-Dependent Asparagine Synthetase	15
Catalytic mechanism	15
Enzyme structure.....	17
Intramolecular ammonia translocation.....	22
Research Objectives.....	23
2 PARAMETERIZATION OF A REACTION INTERMEDIATE OF ASPARAGINE SYNTHETASE-B FOR USE IN CHARMM.....	26
Introduction.....	26
Methods	28
Parameterization Methodology	28
Parameterization Strategy.....	30
Results and Discussion	31
Model Selection and Initial Parameters.....	31
Atom Types and Geometric Parameters.....	33
Partial Atomic Charges.....	34
Force Constants and Dihedral Rotation.....	36
Vibrational Spectra for the Model Complex	44
Conclusions.....	46
3 DEVELOPMENT OF A COMPUTATIONAL MODEL OF ASPARAGINE SYNTHETASE-B.....	49
Introduction.....	49
Methods	51
Simulated Annealing Methodolgy.....	51
Simulated Annealing Procedure	51
Model Systems	53
AspATP model system.....	53

βAspAMP model system	54
Results and Discussion.....	55
AspATP Model.....	55
Simulated annealing run 1	58
Simulated annealing run 2.....	60
Simulated annealing run 3	65
βAspAMP Model	68
Simulated annealing run 1	71
Simulated annealing run 2.....	75
Simulated annealing run 3	78
Conclusions.....	81
4 ENHANCING THE PMF99 SCORING FUNCTION FOR MOLECULAR DOCKING AND VIRTUAL SCREENING.....	83
Introduction.....	83
Methods	85
Protein/Ligand Structure Preparation.....	85
Docking Algorithm.....	86
Results and Discussion	87
The Enhanced PMF99 Scoring Function	87
Molecular Docking Algorithm	91
Assessing the Performance of the ePMF99 Scoring Function in Molecular Docking.....	93
Comparison to Other Docking/Scoring Algorithms.....	97
Virtual Screening for Thymidine Kinase	98
Correlation of the ePMF99 Score and Biological Activity	102
Conclusions.....	103
5 CONCLUSIONS	109
Concluding Remarks	109
The Future of the AS-B Project from a Computational Perspective	110
Modeling the Inhibitors	111
Modeling the Residue Mutations.....	114
APPENDIX CHARMM SIMULATED ANNEALING EXAMPLE INPUT FILES.....	116
Heating Input File.....	116
Equilibration Input File.....	117
Annealing Input File	118
Final Minimization Input File.....	126
LIST OF REFERENCES.....	128
BIOGRAPHICAL SKETCH	140

LIST OF TABLES

<u>Table</u>		<u>page</u>
2-1	The bonds, angles and dihedrals to be parameterized with initial parameters.....	32
2-2	Atomic charges and atom types for β AspAMP and the model complex.....	33
2-3	Model complex, <i>ab initio</i> and CHARMM geometry data.	35
2-4	Vibrational frequencies for the model complex as calculated by HF and CHARMM.	45
2-5	Final parameters for the model complex.	48
3-1	Final energies for the SA Runs for the AspATP model system	56
3-2	Final energies for the SA Runs for the β AspAMP model system	68
4-1	Comparison of rms deviation (\AA) for flexible docking	95
4-2	Accuracy in cross-docking of thymidine kinase inhibitors to the 1kim active site.	98
4.3	Complete docking test set and calculated rmsd of docked ligand vs original crystal structure ligand.....	105
5-1	Sulfoximine 4 model complex and CHARMM parameters necessary	112
5-2	Initial CHARMM atom types and CHelpG charges assigned for the sulfoximine model complex.....	112
5-3	AS-B synthetase domain mutants with the putative relevance.....	115

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1-1 The three reactions catalyzed by AS.....	15
1-2 The bioconversion of aspartate into asparagine that is catalyzed by glutamine-dependent asparagine synthetase.	16
1-3 The putative reaction mechanism of ASNS that proceeds through the reaction intermediate, β AspAMP.	17
1-4 The AS-B monomer as depicted from the crystal structure.....	18
1-5 The homodimer as depicted from the crystal structure of AS-B.	19
1-6 C1A AS-B model.....	20
1-7 The intramolecular tunnel of AS-B.....	21
2-1 The structures of the intermediate, β AspAMP, and of the tetrahedral transition state for the ASNS catalyzed reaction.....	27
2-2 β AspAMP and the model complex with unique atom labels.	30
2-3 The 2-dimensional <i>ab initio</i> energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex.....	38
2-4 The 2-dimensional CHARMM energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex without energy contributions from the dihedrals in Table2-1.....	39
2-5 Newman projections for the dihedrals O2A-PA-O3A-CG and PA-O3A-CG-OD1.....	40
2-6 The 2-dimensional CHARMM energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex with energy contributions from the dihedrals.....	41
2-7 The 2-dimensional CHARMM energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex using the final set of developed parameters.....	43
3-1 Creation of β AspAMP intermediate.....	55
3-2 Synthetase active site interaction in the initial AspATP model system.....	57
3-3 Structural comparison of the initial structure of the AspATP model with the same structure heated to 600 K.....	58

3-4	AspATP SA Run 1.....	59
3-5	AspATP SA Run 2.....	61
3-6	Structural alignments of the minimized structure at 390 K and the structure equilibrated at 300 K with the initial structure.	62
3-7	A look at ATP binding interactions in SA Run 2.	63
3-8	A look at free aspartate binding interactions in SA Run 2.	64
3-9	AspATP SA Run 3.....	66
3-10	Structural comparison of the initial structure of the AspATP model with the minimized structure at 300 K.....	67
3-11	Synthetase active site interaction in the initial β AspAMP model system.	69
3-12	Structural comparison of the initial structure of the β AspAMP model with the same structure heated to 600 K.....	70
3-13	β AspAMP SA Run 1.	71
3-14	Active site comparison between the initial β AspAMP model and the model minimized at 300 K.....	73
3-15	The shift of Glu-348.....	74
3-16	β AspAMP SA Run 2.	75
3-17	Active site comparison between the initial β AspAMP model and the model minimized at 420 K and the model minimized after equilibration at 300 K.	77
3-18	β AspAMP SA Run 3.	79
3-19	Active site comparison between the initial β AspAMP model and the model minimized at 300 K.....	80
4-1	A plot of the pairwise scores for protein/ligand atom pair of NC-OD.	91
4-2	Graphical visualization of optimal ligand poses for arabinose in the arabinose-binding protein.....	96
4-3	Structures of ligands employed in cross-docking studies on thymidine kinase.....	100
4-4	The effect of water on the binding of hpt and dT with TK.....	102
4-5	Plot showing the correlation of the ePMF99 score vs. $\log K_i$ for the pyrimidine (■) and purine (◆) analogs used in the cross-docking studies.....	103

5-1	Structures of the AS reaction intermediate, β AspAMP 1 , transition-state 2 , and known synthetase site inhibitors.	111
5-2	The 2-dimensional <i>ab initio</i> energy surface for dihedral angle CB-SA-NA-PA versus dihedral angle CA-CB-SA-NA of the sulfoximine model complex.	113

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

DEVELOPING A COMPUTATIONAL MODEL OF ASPARAGINE SYNTHETASE-B
TOWARDS RATIONAL INHIBITOR DESIGN

By

Robert N. Humkey

December 2009

Chair: Nigel G. J. Richards

Major: Chemistry

Although Acute Lymphoblastic Leukemia (ALL) has an initial treatment rate of nearly 95%, it claims the lives of over 1,000 people in the US each year. These patients often develop resistance to L-asparaginase, a main chemotherapeutic agent used in the treatment of ALL. There has been a long withstanding, inverse correlation to patients' resistance to L-asparaginase and up-regulation of the enzyme, asparagine synthetase. Therefore, selective potent inhibitors of asparagine synthetase (AS) may be useful for the clinical treatment of ALL.

This work focuses on the development of a structural model of *E. coli* asparagine synthetase-B (AS-B); the glutamine-dependent bacterial form that is similar to human AS. Novel parameters were developed in CHARMM for incorporation of the AS reaction intermediate, β -aspartyl-AMP, into the AS-B model. This intermediate has been successfully used as a model for inhibitor design. Therefore, modeling this intermediate will provide understanding of the protein structure and active site interactions, which will aid in the development of future inhibitors.

Simulated annealing (SA) was used to arrive at an optimized model of AS-B with glutamine bound in the glutaminase active site and β -aspartyl-AMP, pyrophosphate and Mg^{2+} bound in the synthetase active site. The final model was heated from 0 K – 600 K; cooled back

down to 300 K; then equilibrated at 300 K before a final minimization. The optimized model had a very interesting side chain flip occur for Glu-348. The repositioned side chain would make the translocation of NH₃ from the glutaminase site to the synthetase site more probable than the translocation is in the current AS-B crystal structure.

The availability of a structural model for AS sets the stage for the application of *in silico* screening of virtual libraries to identify novel drugs for the treatment of asparaginase-resistant ALL. The validation of algorithms and potentials for molecular docking is a necessary prelude to such efforts. Therefore, an enhanced Potential of Mean Force (PMF) scoring function was developed for *in silico* screening studies.

The sum of the work presented here sets the stage for *in silico* screening or the rational design of inhibitors for AS.

CHAPTER 1 INTRODUCTION

Acute Lymphoblastic Leukemia

Acute lymphoblastic leukemia (ALL) is a cancer of the blood, in which bone marrow produces large quantities of immature white blood cells. This over production of useless white blood cells, in effect, crowds the properly functioning white and red blood cells along with platelets in the blood. ALL is most common in children below the age of 14, affecting three to four children per hundred thousand each year. The American Cancer Society estimates that in 2009, 5,760 new cases of ALL will be diagnosed and 1,400 deaths will be attributed to this form of leukemia.¹

The enzyme L-asparaginase, which catalyzes the hydrolysis of asparagine to aspartate and ammonia,²⁻⁴ is commonly used in conjunction with other chemotherapeutic drugs for the treatment of ALL. This usage is due, in large part, to the substantial evidence supporting an inverse correlation between the levels of intracellular asparagine biosynthesis and the drug susceptibility of T-cell leukemia. L-asparaginase administered alone can result in complete remission for 40 – 60% of ALL cases.^{5,6} When combined with other chemotherapeutic agents, the percentage of untreated ALL patients that experience complete remission jumps to 95%. Despite the apparent benefits, L-asparaginase treatment has limited clinical viability due to three key factors. The first limiting factor is the wide variety of side effects that result from L-asparaginase treatment.⁷ The second factor limiting clinical viability is the fact that many of the patients in remission experience relapses with L-asparaginase resistant tumors.⁷⁻¹⁰ The final limiting factor is the observation that L-asparaginase actually promotes the growth of resistant tumors and raises their metastatic activity.^{5,11} The potential adverse effects of L-asparaginase

treatment has limited its use for the treatment of ALL, despite estimates that 5 – 10% of all solid tumors, not solely ALL tumors, may be susceptible to this treatment.¹⁰

A major problem in the clinical treatment of relapsed ALL patients is the lack of understanding of L-asparaginase resistance on the molecular level.^{12,13} Early studies on L-asparaginase resistant leukemia patients found elevated levels of the enzyme asparagine synthetase (AS), which catalyzed the biosynthesis of asparagine in L-asparaginase resistant patients.¹⁴ This was the first insight into the correlation between the cellular upregulation of AS and L-asparaginase resistance in leukemia cells. Multiple studies since have confirmed this observation and strengthened the theory that AS is linked to L-asparaginase resistance in leukemia cells.^{9,15-18} A human MOLT-4 cell line was established from a 19-year-old male patient, diagnosed with ALL in 1971.¹⁹ This cell line has provided an excellent model for testing leukemia cells response to varying stimuli. It was found that short-term treatment of drug-sensitive MOLT-4 leukemia cells with L-asparaginase resulted in elevated levels of AS, an effect that is not fully reversible.¹⁷ Further, work on the MOLT-4 cell line has provided direct evidence that the overexpression of AS is adequate for development of L-asparaginase resistance. When the drug-sensitive MOLT-4 cell line was transformed using a retrovirus containing the human AS gene under a constitutively active promoter, the resulting overexpression of AS was sufficient to induce L-asparaginase drug resistance in the MOLT-4 cell line.¹⁷ The apparent correlation between AS and L-asparaginase resistance in ALL,^{12,20-22} coupled with the fact that a nanomolar affinity inhibitor of AS has demonstrated an ability to suppress proliferation of a drug-resistant MOLT-4 cell line makes AS a clinically interesting enzyme.²³

Asparagine Synthetase

Asparagine synthetase (AS), a member of the amidotransferase family of enzymes, catalyzes the biosynthesis of L-asparagine from L-aspartate, in an adenosine triphosphate (ATP)

dependent reaction that uses either glutamine or ammonia as a source of nitrogen.²⁴ The bacterium *Escherichia coli* (*E. coli*) contains two separate genes that both code for AS. The first gene, *asnA*, codes for *E. coli* asparagine synthetase-A (AS-A) and catalyzes the ammonia-dependent synthesis of asparagine.²⁵⁻²⁷ The second gene, *asnB*, codes for *E. coli* asparagine synthetase-B (AS-B)²⁸ and like human AS²⁹ catalyzes the glutamine-dependent synthesis of asparagine, although AS-B is not evolutionarily related to the human form of AS. All told, AS-B catalyzes three distinct reactions as shown in Figure 1-1.^{30,31}

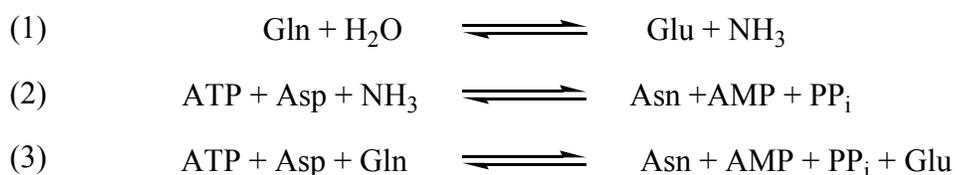


Figure 1-1. The three reactions catalyzed by AS. Reaction (1) is the hydrolysis of glutamine to glutamate and ammonia. Reaction (2) is the ammonia-dependent conversion of aspartate to asparagine, using ATP and producing AMP and pyrophosphate. Reaction (3) is the glutamine-dependent conversion of aspartate to asparagine.

The first residue of glutamine-dependent AS (ASNS), cysteine (Cys-1), mediates the glutaminase activity (reaction (1)) in the N-terminal domain. The sequence of the N-terminal glutamine amide transfer (GAT) domain, along with the distinguishing cysteine residue, places AS in the Class II, or N-terminal nucleophilic (Ntn), family of glutamine-dependent amidotransferases.^{32,33} Also included in the Ntn-family are glutamine 5'-phosphoribosyl-1-pyrophosphate amidotransferase (GPATase),^{34,35} glutamine fructose-6-phosphate amidotransferase (GFAT)^{36,37} and glutamine synthase.³⁸⁻⁴⁰

Glutamine-Dependent Asparagine Synthetase

Catalytic mechanism

As mentioned earlier, ASNS catalyzes the bioconversion of L-aspartate into L-asparagine using ATP and glutamine as seen in Figure 1-2. ASNS has also shown the ability to use

ammonia, hydroxylamine and hydrazine as its source of nitrogen under *in vitro* conditions.⁴¹ While the enzyme has been isolated from mammals,⁴²⁻⁴⁴ yeast,⁴⁵ and bacteria,^{28,46} the only native mammalian enzyme studied in detail has been that which is present in the bovine pancreas.^{44,47-49} Most of the characterization on the eukaryotic form of the enzyme, however, has been done on active, recombinant human ASNS.⁵⁰ Recombinant forms of ASNS can also be obtained from *E. coli*³¹ and *Vibrio cholerae*.⁵¹ ASNS exhibits high glutaminase activity even in the absence of aspartate. This has resulted in several different suggestions for the order of substrate binding and product release.^{49,51-55} Based on an observed dependence of the glutamine:asparagine ratio on the initial glutamine concentration, one kinetic scheme suggested by Tesson, *et al.* predicts that the catalytic activity of synthetase and glutaminase sites are only weakly coupled.⁵² This observation means ASNS is unlike the vast majority of the other amidotransferases, which all exhibit tightly coupled active sites. Recently, asparagine has been shown to bind to the glutaminase site as a competitive inhibitor in both AS-B⁵⁶ and the *Vibrio* enzyme.⁵¹ Therefore, asparagine can act as a means of regulating the glutamine-dependent activity of ASNS.

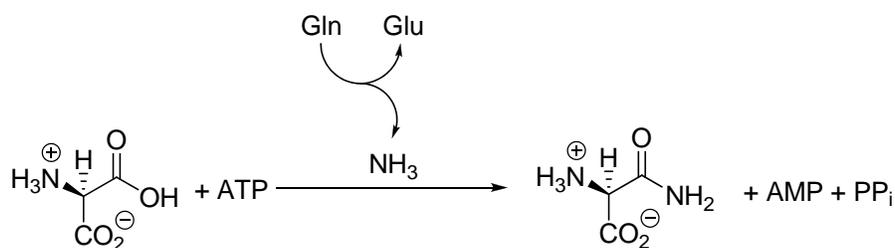


Figure 1-2. The bioconversion of aspartate into asparagine that is catalyzed by glutamine-dependent asparagine synthetase.

The mechanism by which ASNS catalyzes the conversion of aspartate into asparagine is believed to occur via the reaction mechanism depicted in Figure 1-3. Aspartate binds in the C-terminal synthetase active site where the side chain carboxylate is activated by an Mg^{2+} -ATP complex to form the β -aspartyl-AMP intermediate (βAspAMP) and pyrophosphate (PP_i). The

β AspAMP formation has been demonstrated by ^{18}O isotope labeling experiments.^{44,53} While this is occurring in the synthetase site, glutamine is hydrolyzed to glutamate and ammonia in the N-terminal glutaminase site. The third and final reaction involves the attack of ammonia, which is shuttled from the glutaminase site to the synthetase site via the proposed intramolecular tunnel, on β AspAMP yielding AMP and aspartate.

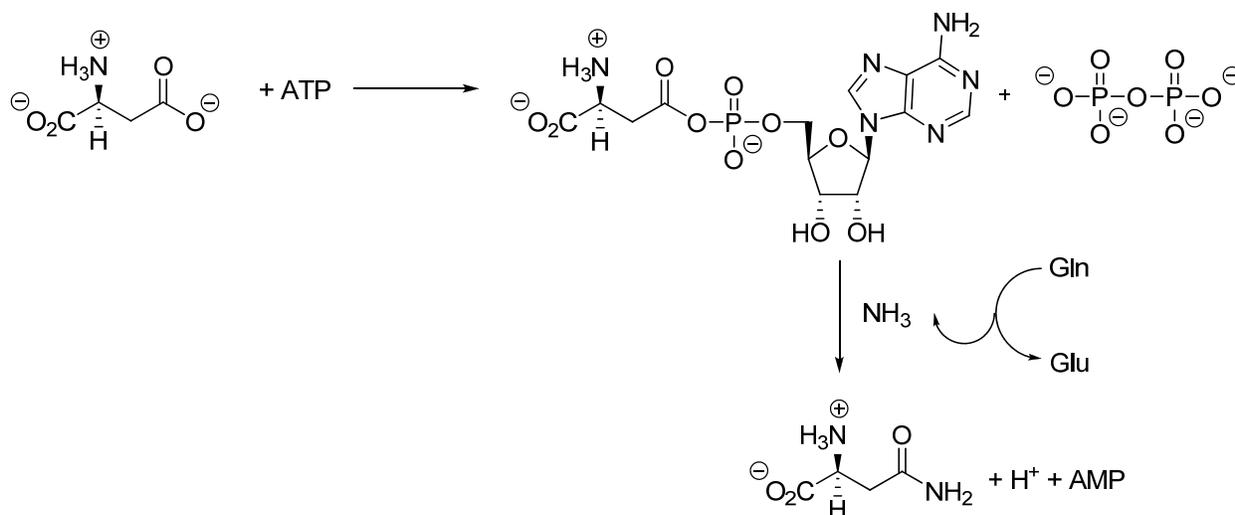


Figure 1-3. The putative reaction mechanism of ASNS that proceeds through the reaction intermediate, β AspAMP.

Enzyme structure

In an effort to elucidate the mechanism of ASNS, the structure of the Cysteine-1 to alanine (C1A) mutant of *E. coli* AS-B was crystallized in a ternary complex with glutamine in the N-terminal glutaminase site and adenosine monophosphate (AMP) in the C-terminal synthetase site. The crystal structure (PDB code: 1ct9) is shown in Figure 1-4. The mutant lacks all glutaminase activity due to the replacement of the first cysteine with alanine. This crystal structure was resolved to 2.0 \AA .⁵⁷ AS-B is comprised of 553 amino acids, with a molecular weight of 62.5 kDa, and is believed to function as a homodimer, i.e. two identical monomer units, with each monomer consisting of two domains, as represented in Figure 1-5.⁵⁸ AS-B

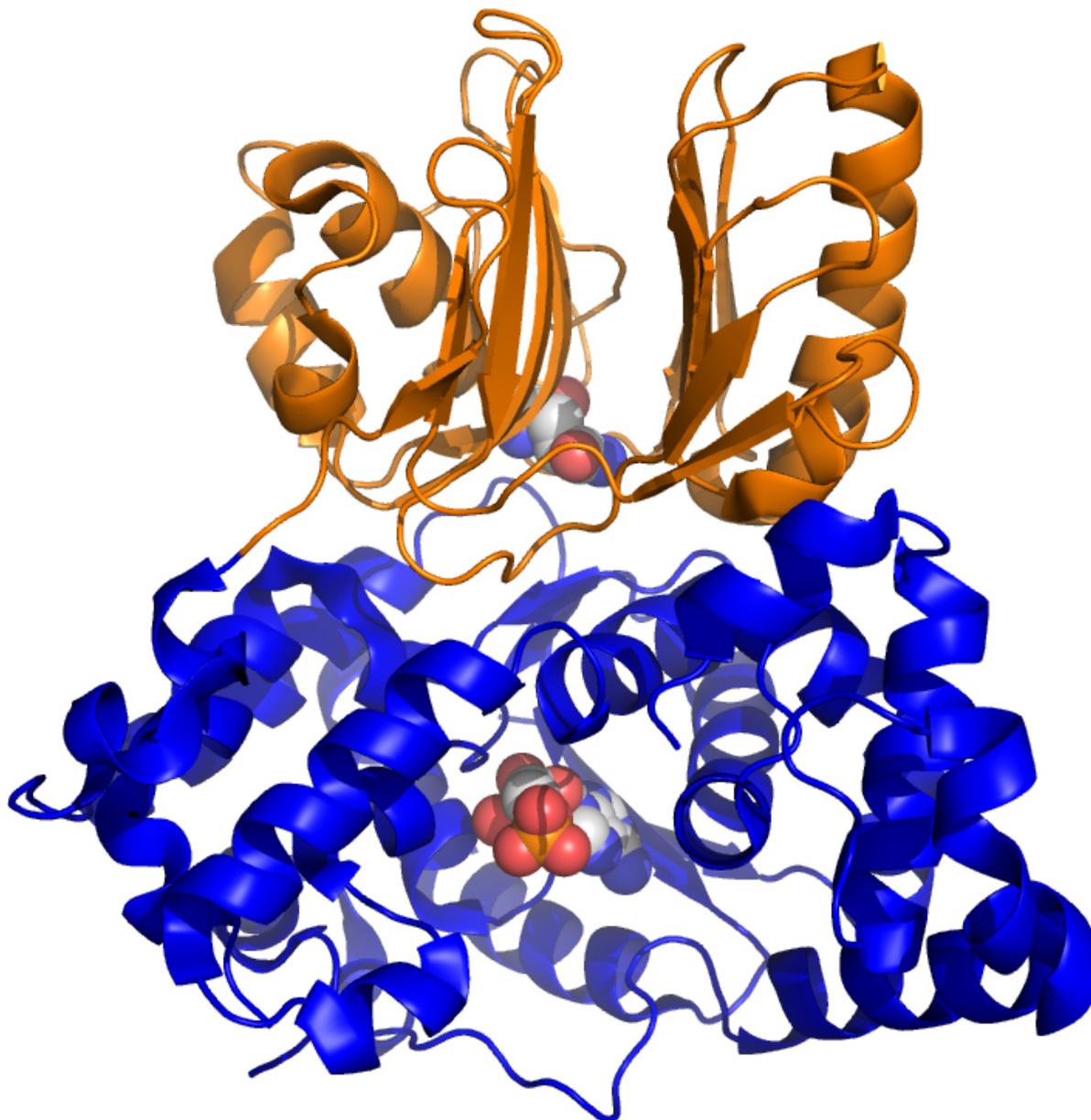


Figure 1-4. The AS-B monomer as depicted from the crystal structure (PDB Code: 1ct9). The glutaminase domain is colored in orange and the synthetase domain is colored in blue. Glutamine is present in the glutaminase active site and AMP is present in the synthetase active site. Both bound ligands are depicted in sphere representation and colored by element.

possesses the N-terminal catalytic domain consistent of all Ntn amidotransferases, constructed from two layers of six-stranded anti-parallel β -sheets that regulate the hydrolysis of glutamine (Gln) to glutamate (Glu) and ammonia.⁵⁷ The C-terminal domain is the larger domain and

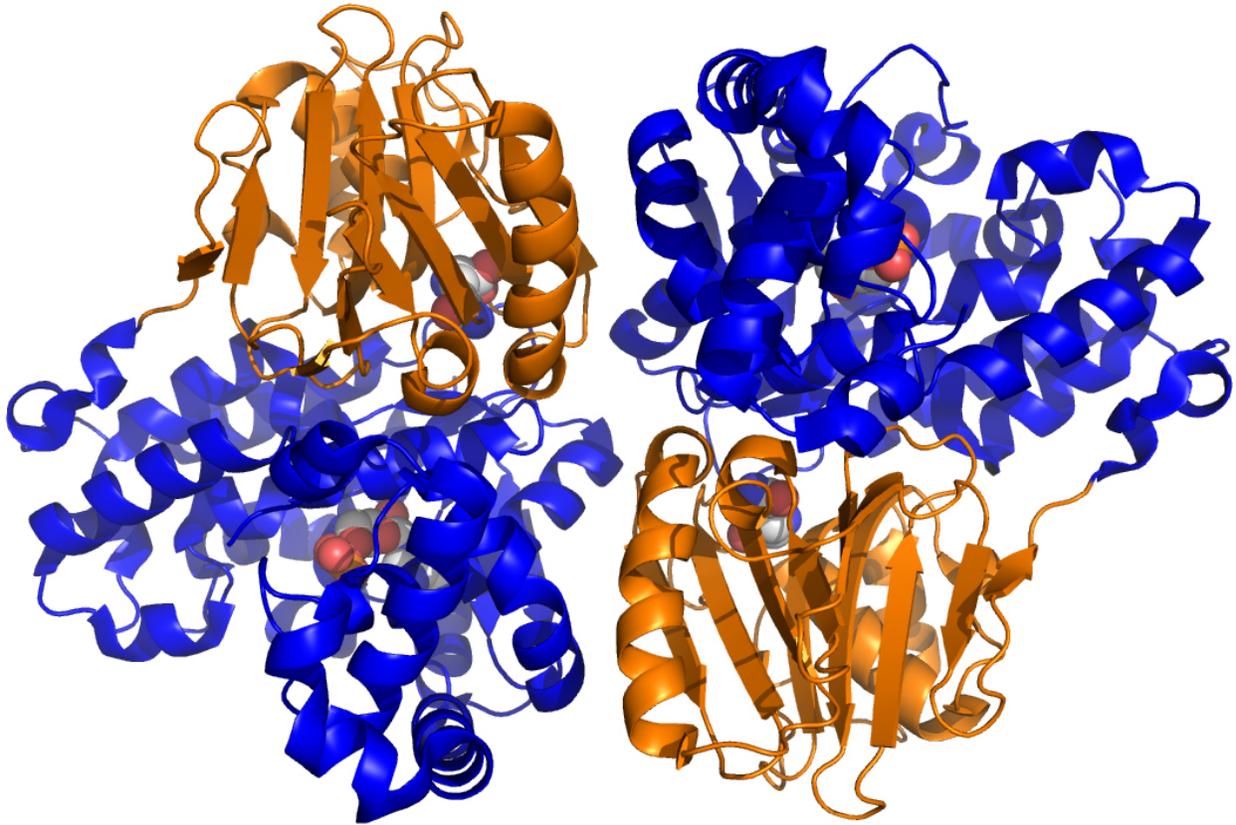


Figure 1-5. The homodimer as depicted from the crystal structure of AS-B (PDB Code 1ct9). The N-terminal glutaminase domains are shown in orange and the C-terminal synthetase domains are shown in blue. The bound ligands are colored by element and shown in sphere representation.

houses the synthetase site where aspartate (Asp) is converted to asparagine (Asn), while consuming one molecule of ATP, yielding AMP and pyrophosphate (PP_i).⁵⁹ Interestingly, the C-terminal domain of AS-B does not resemble that of the ammonia-dependent AS-A^{25-27,60,61} or GPATase,⁶²⁻⁶⁶ but rather the C-terminal domain of AS-B was found to be structurally homologous to guanosine-5'-monophosphate synthetase (GMPS),⁶⁷ argininosuccinate synthetase,^{68,69} ATP sulfurylase,^{70,71} β-lactam synthetase,⁷²⁻⁷⁴ and ThiI (4-thiouridine synthetase).^{75,76} Catalysis in all of these enzymes leads to the formation of PP_i from ATP.⁵⁹ The formation of PP_i as a by-product, along with the conserved amino acid motif SGGXDS, which is known as a pyrophosphatase loop, places these enzymes in the ATP pyrophosphatase family.^{59,77}

Disorder in the C-terminal domain of the AS-B crystal structure did not allow two particular loop regions (Ala-250 to Leu-267 and Cys-422 to Ala-426) or the final 37 residues to be resolved, thus resulting in their absence in the crystal structure. The two unresolved loop regions were modeled onto the crystal structure using computational methods. The structure is depicted in Figure 1-6, as part of Dr. Yun Ding's dissertation,⁷⁸ leaving only the final 37 residues absent.

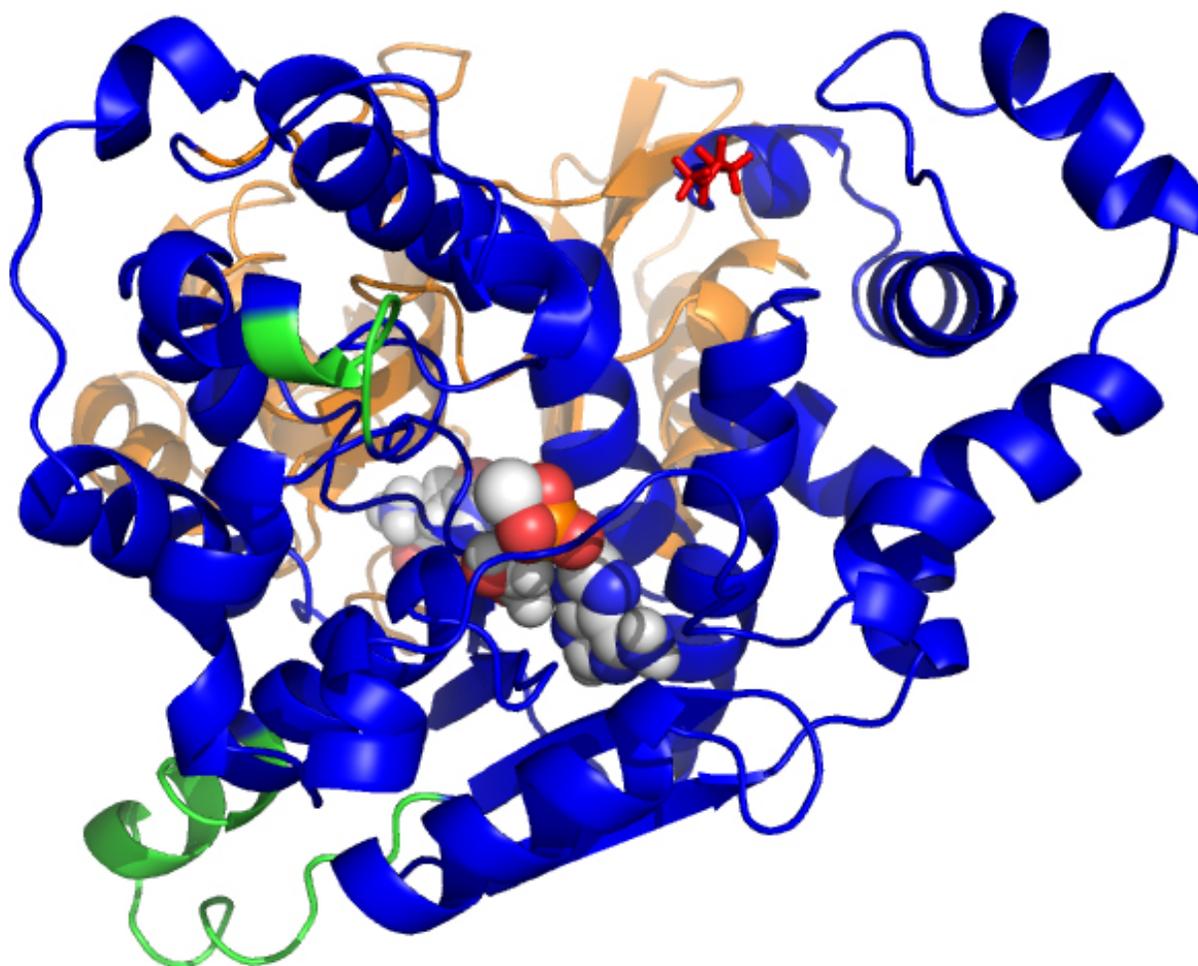


Figure 1-6. C1A AS-B model. The glutaminase domain is depicted in orange and the synthetase domain is depicted in blue. The two loop regions absent in the 1ct9 crystal structure have been modeled in (shown in green) and the last residue of the crystal structure (Gly-516) is shown in red in stick representation. The final 37 residues that remain absent would extend from Gly-516. Glutamine (in the glutaminase site) and AMP (in the synthetase site) are shown in sphere representation.

The N-terminal glutaminase active site and the C-terminal synthetase active site are separated in the AS-B crystal structure by approximately 20 Å. A solvent-inaccessible intramolecular tunnel, shown in Figure 1-7, is believed to connect the two active sites for ammonia translocation. The putative tunnel is lined almost exclusively with hydrophobic residues (Met-120, Ile-142, Ile-143, Leu-232, Met-329, Ala-399 and Val-401), with only a few highly conserved polar residues found near the tunnel openings at the synthetase site (Ser-346 and Glu-348) and at the glutaminase site (Arg-30). Because the majority of the identified tunnel residues are located in the interior of the protein and associated with secondary structural

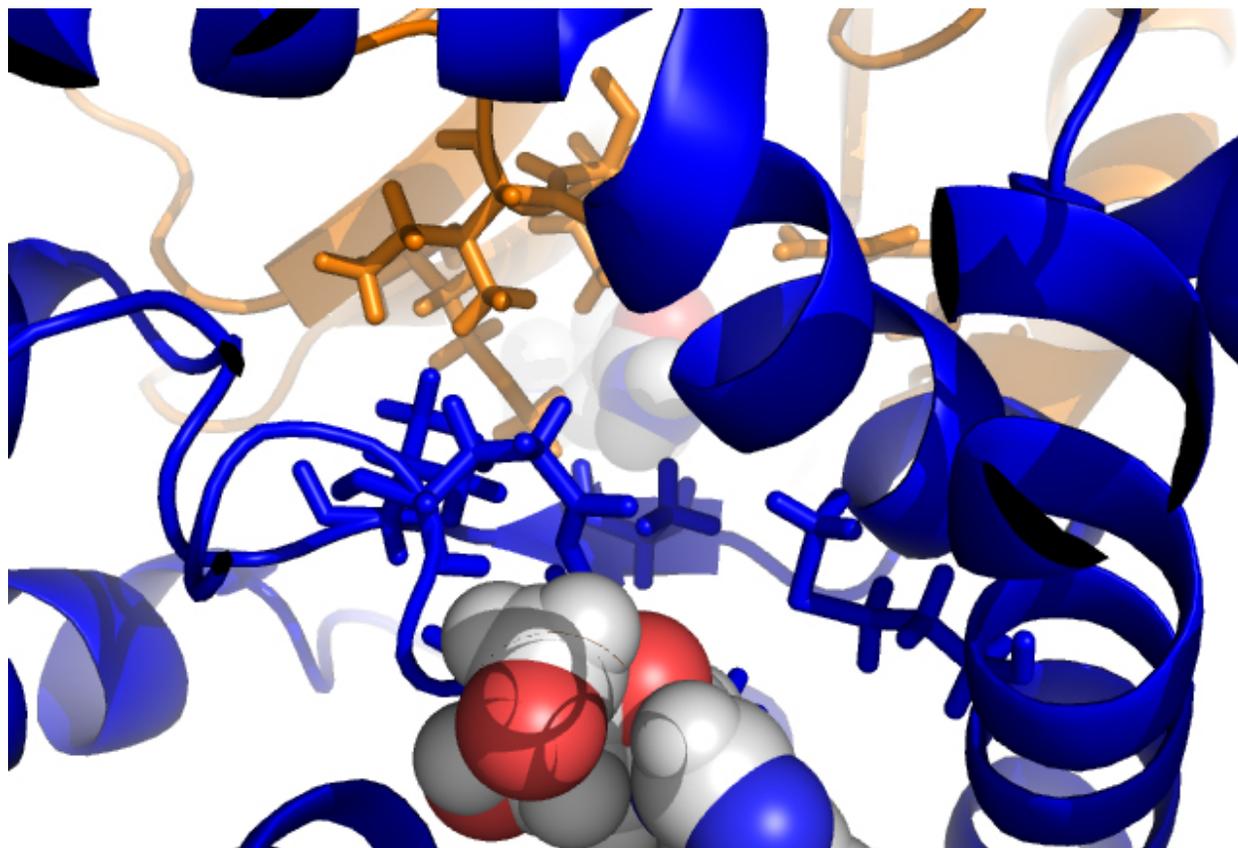


Figure 1-7. The intramolecular tunnel of AS-B. This figure is a view looking down the tunnel from the synthetase site (bottom of picture in blue) with AMP present to the glutaminase site (top of picture in orange) with glutamine present. The residues believed to line the tunnel are shown in stick representation.

elements, the current belief is that tunnel is always present and does not open and close through the catalytic cycle as in other amidotransferases. The presence, in the crystal structure, of water molecules at the interface of the two domains gives credence to the idea that free ammonia could potentially access the tunnel as the protein undergoes conformational changes during catalysis.

Intramolecular ammonia translocation

Ideally, it is assumed that the catalytic mechanism of AS-B would exhibit a 1:1 ratio of Glu:Asn. The presence of an intramolecular tunnel should make it possible for every molecule of ammonia released from the hydrolysis of Gln to Glu to be efficiently directed to the synthetase site for use in the conversion of Asp to Asn. However, an experimental and simulated study of the steady-state kinetics of AS-B found the catalytic ratio of Glu:Asn to be 1.89⁵². This demonstrates an inefficient coupling of the glutaminase and synthetase sites in ASNS that could be explained by flaws in translocation of ammonia via the intramolecular tunnel. To test the ability of exogenous ammonia to suppress the incorporation of nitrogen from glutamine hydrolysis, a competition experiment using *E. coli* AS-B was undertaken.⁵⁶ This experiment utilized a new isotope-edited ¹H NMR-based assay⁷⁹ that has enhanced sensitivity for ¹⁵N NMR measurements. It is important to note that in these experiments only ammonium chloride was ¹⁵N labeled. The results of these experiments demonstrated that, in spite of glutamine concentrations upwards of 40mM, ¹⁵N incorporation into asparagine from ¹⁵NH₃ still took place. A previous experiment had shown that ¹⁵N ammonia from solution could also attack the thioester intermediate in the glutaminase active site,^{80,81} even with saturating levels of ATP and aspartate. This observation led to the belief that exogenous ammonia can access the tunnel through conformational changes to the protein that render the tunnel accessible to solvent when ATP, aspartate and glutamine are all present. The mechanism by which this would occur is still unknown, but would hypothetically be the underlying cause for the uncoupling of the

glutaminase and synthetase activities.⁵² The inefficiency of ASNS in coupling its separate active site activities makes it unique among the amidotransferases,^{82,83} which all demonstrate tightly-coupled glutaminase and synthetase activities.

Interestingly, when aspartate is not present, the glutaminase activity of AS-B is enhanced almost 2-fold when ATP, AMP-PNP (a non-hydrolysable ATP analog⁸⁴), or AMP and PP_i are bound in the synthetase site.³¹ As a control, it was shown that the glutaminase activity was not stimulated if only one of AMP or PP_i was present in the assay.³¹ It was also proven in the same set of kinetic experiments that the highly conserved Arg-30 plays an important role in the glutaminase stimulation. When Arg-30 was mutated to either alanine or lysine, the glutaminase activity was no longer stimulated.³¹ These results suggest that similar to GPATase,⁶² GFAT^{85,86} and GltS,^{40,87} AS-B possess inter-domain interactions that lead to conformational changes for the signaling of active site occupancy. What is still unclear is whether these “signals” are transferred within an AS-B monomer or between separate monomers (glutaminase with synthetase) of the heterodimer unit as was depicted in Figure 1-5. The Arg-30 likely plays a prominent role in a hydrogen bonding network⁵⁷ that orients the side chain of Asn-74, which is necessary for stabilization of the tetrahedral intermediate created during the glutaminase reaction via the oxyanion hole.⁸⁸ This low-level active site communication may have evolved due to the ability of asparagine to compete for the glutaminase site with glutamine, hence lowering the detrimental cellular effects of the uniquely elevated glutaminase activity of ASNS.

Research Objectives

The ultimate goal of the work on asparagine synthetase is to develop potent viable inhibitors of human form of the enzyme. The development of viable inhibitors for ASNS could aid in the treatment of ALL, as well as be utilized in the treatment of other solid tumors. As was discussed previously, inhibition of ASNS in drug-resistant ALL could provide those patients

with a second opportunity for a successful treatment that is currently unavailable. Inhibitors of ASNS could also potentially have applications in the treatment of solid tumors in general.¹⁰ However, to date the structure of human asparagine synthetase remains unresolved. The structure of the C1A mutant of *E. coli* AS-B⁵⁷ is still the best available structure of ASNS and is believed to be similar to the human form of the enzyme. This structure still possess certain flaws, unresolved regions, and does not have the synthetase reaction intermediate or transition state bound. Knowing the structure-function relationships of the intermediate, transition state or any of the known inhibitors will aid the progression of better inhibitor design. A better AS-B or human ASNS structure or model would be very useful in better understanding the specific interactions that are involved in the transition state stabilization and could be mimicked more directly in the rational design of inhibitors.

The over-riding goal of this work is to further develop a rough model of AS-B that was originally undertaken by Dr. Yun Ding as part of her dissertation work.⁷⁸ In order to refine an AS-B model, with the β AspAMP synthetase intermediate bound to the synthetase active site, in the CHARMM modeling package,^{89,90} a new set of parameters must be developed. Another goal of this work was the testing and refinement of parameters for a molecular docking package. The availability of a refined docking algorithm would allow for the virtual screening of entire libraries of compounds with a model of AS-B. Having the capability to virtually screen would aid in the identification, design and direction for new or untested inhibitors of ASNS.

The specific goals for this research were to (i) develop the necessary parameters for the synthetase reaction intermediate, β AspAMP, for its inclusion in CHARMM; (ii) develop and refine a model of AS-B with glutamine in the glutaminase site and β AspAMP, PP_i and Mg^{2+} in the synthetase site to understand the structure-function relationship between AS-B and

β AspAMP; (iii) to explore the development of a new docking potential energy function that can be used in future virtual screening experiments.

CHAPTER 2 PARAMETERIZATION OF A REACTION INTERMEDIATE OF ASPARAGINE SYNTHETASE-B FOR USE IN CHARMM

Introduction

The computational modeling of large-scale biological systems continues to grow in popularity, as the methods used to gain insight into these systems have developed. The ability of computer simulations to offer a greater understanding of the fundamental properties governing the molecules within these large systems has brought computational biochemistry wider acceptance. High-level quantum mechanical methods will always be the gold standard for understanding chemical properties. Unfortunately, even with the dramatic increases in computer speeds and parallel programming algorithms, the current technologies still restrict *ab initio* and density functional theory calculations to well under 100 heavy atoms. This limitation makes using pure quantum mechanical methods impossible when studying the likes of enzymatic reactions and protein folding. Instead, molecular dynamics (MD)⁹¹⁻⁹³ and Monte Carlo^{94,95} simulations, which are governed by empirical molecular mechanics force fields, are used. These methods have the distinct advantage of computational efficiency, but rigorous parameterization of the potentials is crucial for realistic results to be expected. The parameterization of the empirical potential is accomplished via small model compounds so that the parameters can then be transferred to larger molecules or proteins.

In the past few decades, a number of empirical potential energy force fields have been extensively parameterized and tested for accuracy.^{90,96-103} The CHARMM force field^{89,90,104} was developed for use with large biomolecules, namely proteins, lipids and nucleic acids. The continued parameterization of CHARMM is always done with the aim of improving accuracy without increasing the complexity of the potential energy function (that can be seen in Equation (2-1) below). By keeping the potential energy function static and optimizing the parameters or

adding to the parameters, CHARMM has improved accuracy without major increases to the computational time required to perform the calculations. CHARMM currently employs an all-atom parameter approach for proteins,¹⁰⁴ lipids,¹⁰⁵ nucleic acids,¹⁰⁶ carbohydrates,¹⁰⁷ and more that has been shown to produce quality results in a variety of different simulations.^{108,109} One issue that arises in any parameter set is the lack of transferability of parameters to any unique atom type, bond, angle, or dihedral. This is due to the basic philosophy of parameterization that seeks to maximize the reproducibility of target data for a particular set of compounds. In order to overcome this deficiency, substrate-specific parameters are continually developed by users for the direct application of their system, as is the case here.

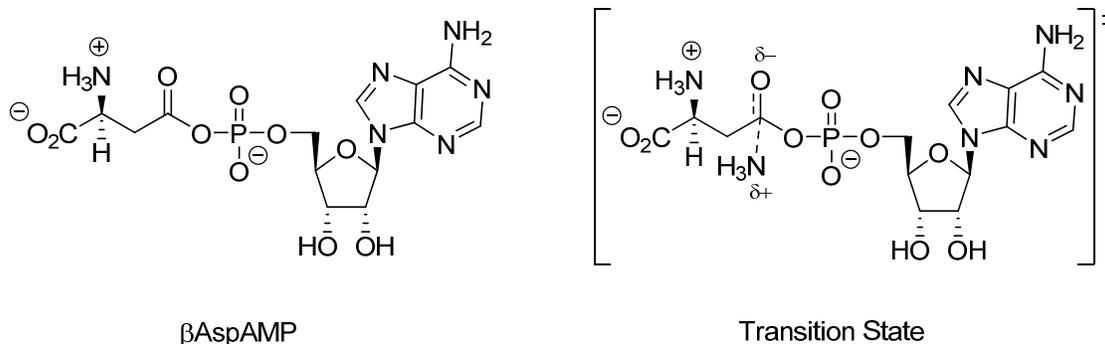


Figure 2-1. The structures of the intermediate, β AspAMP, and of the tetrahedral transition state for the ASNS catalyzed reaction.

The intermediate and transition state have been identified in the bioconversion of aspartate into asparagine catalyzed by ASNS. The intermediate, β AspAMP, and the tetrahedral transition state, shown in Figure 2-1, have been successfully used as models for inhibitor design.²³ Unfortunately, no structural information is available, crystallographic or otherwise, that would provide much needed information on the structure-function relationship between either β AspAMP with ASNS or the transition state with ASNS. In order to develop a reliable model of the enzyme that contains either of these species, molecular dynamics must be used in the

investigation. However, for this type of simulation to be performed in CHARMM, the uncommon bonds, angles and dihedrals within the β AspAMP must first have parameters developed that can be integrated into the existing set of parameters that CHARMM currently employs. This chapter will describe the methods utilized to develop the parameters for β AspAMP that heretofore were absent in the parameter set used by CHARMM. Without the development of the aforementioned parameters, the investigation of the structure-function relationship of the β AspAMP intermediate with ASNS would not be possible using the molecular mechanical methods within the CHARMM package.

Methods

Parameterization Methodology

The parameters for the β AspAMP intermediate were developed using the methods and philosophies adopted in the development of the CHARMM simulation program,⁹⁰ wherever possible. The empirical energy function called in CHARMM is comprised of both intramolecular and intermolecular terms of the following form:

$$\begin{aligned}
 U(\vec{R}) = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{angle} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) \\
 & + \sum_{impropers} K_{imp} (\varphi - \varphi_0)^2 + \sum_{nonbonded} 4\mathcal{E} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}}
 \end{aligned} \tag{2-1}$$

where K_b , K_{UB} , K_θ , K_χ , and K_{imp} are the force constants for the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle, respectively. The bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle are represented by b , S , θ , χ , and φ , respectively, with the subscript zero representing the equilibrium values for the individual terms. The Urey-Bradley term is included to improve non-bonded 1,3-interactions that tend to be repulsive. The last two terms strictly cover the bulk of the nonbonded interactions. In the

Lennard-Jones 6 – 12 terms, σ is the radius when the Lennard-Jones energy is zero, $-\epsilon$ is the minimum Lennard-Jones well depth, $2^{1/6} \sigma$ is distance at the Lennard-Jones minimum, and r_{ij} is just the distance between atoms i and j . The last term is the Coulomb contribution to the total energy. Here, q_i and q_j are the partial atomic charges on atoms i and j , respectively, and ϵ_1 is the effective dielectric constant. The vector of the coordinates of the atoms, \vec{R} , when given, makes the distances and angles that are necessary to evaluate $U(\vec{R})$ simple to determine. The Lorentz-Berthelodt combination rules are used in CHARMM to determine the Lennard-Jones parameters between pairs of different atoms. This is done for the calculation of ϵ_{ij} values that are based on the geometric mean of ϵ_i and ϵ_j . The Lennard-Jones minimum, $R_{min,ij}$, is calculated similarly as it is based on the arithmetic mean between $R_{min,i}$ and $R_{min,j}$. For the current version of the CHARMM protein parameter set, CHARMM22,¹⁰⁴ the effective dielectric constant, ϵ_1 , is set at 1 in order to attain a balanced parameterization with regard to the electrostatic contributions to the energetics. If a high dielectric constant solvent is necessary, neutralized charged groups can be added to the system to add some of the effects of shielding.¹¹⁰ The nonbonded interactions are determined for all atoms that are separated by at least three bonds and no scaling is done on these interactions in CHARMM except in specific cases where there is scaling of the 1 – 4 Lennard-Jones term (such as oxygen atoms, aliphatic carbons, and amide nitrogens). There is no term included for hydrogen bonding, as it has been demonstrated that the Coulomb and Lennard-Jones terms can accurately model those interactions.^{111,112} CHARMM uses the TIP3P model¹¹³ for the representation of water in all calculations for consistency of the protein and solvent interactions.

Parameterization Strategy

The work presented here was focused on the development of parameters for the ASNS reaction intermediate, β AspAMP. A model of *E. coli* AS-B with β AspAMP bound in the synthetase site would be a great asset in exploring the structure-function relationship of the enzyme with the intermediate. Determining what enzyme/intermediate interactions are important and resolving what residues play key roles in binding the intermediate and catalyzing the conversion of aspartate to asparagine could help not only to guide future mutagenesis experiments, but also would help with the rational design of future inhibitors. The β AspAMP intermediate is composed of one molecule of AMP and one free aspartate. Parameters for both of those molecules exist in CHARMM. However, as can be seen in Figure 2-2, the area where these two molecules are joined is where CHARMM parameters must be developed for the unique bonds, angles, and dihedrals of β AspAMP. Rather than using the entire β AspAMP (with 49 atoms) in *ab initio* energy calculations to determine a two dihedral potential energy surface (PES)

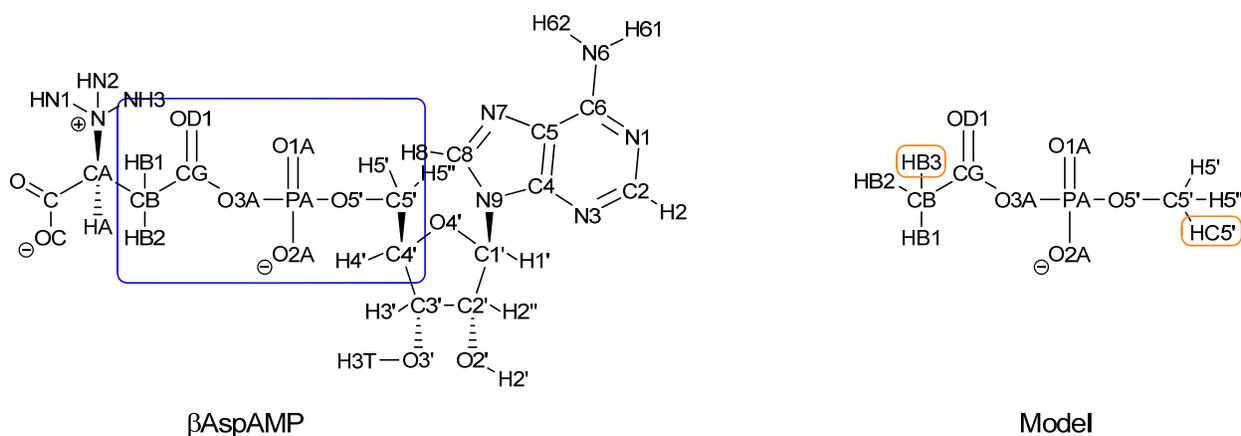


Figure 2-2. β AspAMP and the model complex with unique atom labels. This figure shows the β AspAMP reaction intermediate and the model complex used to develop the parameters necessary to integrate β AspAMP into CHARMM simulations. The model complex is based on the core region within the blue box on β AspAMP. The hydrogen atoms in orange boxes on the model complex were added to the model and are not a part β AspAMP.

and in other necessary *ab initio* calculations, a model complex (with only 15 atoms) that contains all of the critical atoms and connections needed to represent β AspAMP accurately was used. The parameterization of β AspAMP was particularly challenging because there is limited crystallographic data on which β AspAMP can be parameterized. Ideally, the necessary parameters would be developed with the goal of reproducing the crystallographic geometries for the molecule of interest.

Results and Discussion

Model Selection and Initial Parameters

The majority of the parameters for the β AspAMP intermediate were transferred from the available parameters in the CHARMM package for AMP in the nucleic acid parameter set¹⁰⁶ and the parameters for aspartate in the protein parameter set.¹⁰⁴ The small set of bonds, angles and dihedrals that do not have existing parameters in CHARMM are related to the unique connection made between the carboxylate group on the side chain of aspartate and the phosphate group of AMP. A list of the bonds, angles and dihedrals that required parameters is summarized in Table 2-1. The model complex shown in Figure 2-2 was chosen to represent the unparameterized region within the β AspAMP for computational efficiency when using *ab initio* methods for calculating the PES that was used to describe the critical dihedrals. The model was also used in the *ab initio* calculations that determined the vibrational spectra and partial atomic charges. This comprised the remainder of the target data for the parameterization. The C_B and C_{5'} carbons of β AspAMP were chosen as the end “caps” for the model and additional hydrogens were added to each carbon to create methyl groups. This common practice of enclosing the region of interest between methyl groups is done to describe more accurately the partial charges on the critical

atoms, as partial charges on the methyl groups are constant and prevent long-range charge dispersion.

Table 2-1. The bonds, angles and dihedrals to be parameterized with initial parameters.

Structure	Bonds	K_b	b_0		
	CC-ON2	650.0 ^a	1.316 Å		
Structure	Angles	K_θ	θ_0	K_{UB}	S_0
	P-ON2-CC	20.0 ^b	128.2°	35.0 ^b	2.33 ^b
	ON2-CC-CT2	70.0 ^c	117.6°		
	ON2-CC-OC	98.0 ^d	120.3°		
Structure	Dihedrals	K_ϕ	N	δ	
	ON2-P-ON2-CC	4.1 ^e	1	0.0°	
	P-ON2-CC-CT2	3.5 ^e	3	0.0°	
	P-ON2-CC-OC	11.2 ^e	2	180.0°	
	P-ON2-CC-OC	11.2 ^f	2	180.0°	
	ON3-P-ON2-CC	0.1 ^g	3	180.0°	

^a Value taken from O-CC in acetamide; ^b Value taken from P-ON2-CN9 in methyl diphosphate; ^c Value taken from ON2-CC-CT2 in CoA; ^d Value taken from ON2-CC-OC in CoA; ^e Calculated from HF/6-31+G(*d*) potential energy surface (O5'-PA-O3A-CG vs PA-O3A-CG-CB); ^f Estimated to be same as P-ON2-CC-CT2; ^g Taken from ON3-P-ON2-CN7 in dimethylphosphate. The b_0 and θ_0 were set to the distance and angles obtain from the *ab initio* minimized structure.

Atom Types and Geometric Parameters

The atom types were defined so that no new atom types would have to be introduced, without compromising any expected accuracy in reproducing the *ab initio* results. The only atom that would have greatly benefited from introducing a new atom type would have been the

Table 2-2. Atomic charges and atom types for β AspAMP and the model complex.

β AspAMP			β AspAMP			Model		
Atom	Type	Charge	Atom	Type	Charge	Atom	Type	Charge
N1	NN3A	-0.74	H4'	HN7	0.09	C5'	CN8B	0.24
C2	CN4	0.50	O4'	ON6B	-0.50	H5'	HN8	0.02
H2	HN3	0.13	C5'	CN8B	0.23	H5''	HN8	-0.01
N3	NN3A	-0.75	H5'	HN8	0.02	HC5'	HN8	-0.01
C4	CN5	0.43	H5''	HN8	-0.01	O5'	ON2	-0.55
C5	CN5	0.28	O5'	ON2	-0.55	PA	P	1.52
C6	CN2	0.46	PA	P	1.52	O1A	ON3	-0.88
N6	NN1	-0.77	O1A	ON3	-0.88	O2A	ON3	-0.89
H61	HN1	0.38	O2A	ON3	-0.89	O3A	ON2	-0.61
H62	HN1	0.38	O3A	ON2	-0.61	CG	CC	1.01
N7	NN4	-0.71	CG	CC	1.01	OD1	OC	-0.74
C8	CN4	0.34	OD1	OC	-0.74	CB	CT2	-0.40
H8	HN3	0.12	CB	CT2	-0.30	HB1	HA	0.10
N9	NN2	-0.05	HB1	HA	0.10	HB2	HA	0.10
C1'	CN7B	0.16	HB2	HA	0.10	HB3	HA	0.10
H1'	HN7	0.09	CA	CT1	0.07			
C2'	CN7B	0.14	HA	HB	0.09			
H2''	HN7	0.09	N	NH3	-0.47			
O2'	ON5	-0.66	HN1	HC	0.31			
H2'	HN5	0.43	HN2	HC	0.25			
C3'	CN7	0.14	HN3	HC	0.26			
H3'	HN7	0.09	C	CC	0.51			
O3'	ON5	-0.66	O	OC	-0.51			
H3T	HN5	0.43	OC	OC	-0.51			
C4'	CN7	0.16						

The “atom” is the unique name given to each atom, while the “type” is the atom type recognized by CHARMM.

bridging oxygen between the aspartate and AMP. However, for simplicity and consistency with the existing CHARMM parameters, the most similar atom type, a bridging oxygen in a phosphate group, was used. A full list of the atom types selected for the model complex as well as the β AspAMP is summarized in Table 2-2.

As mentioned previously, because there is no crystallographic data available for β AspAMP, the equilibrium bond lengths and angles were set in accordance with the optimized geometry from an *ab initio* calculation. Multiple geometry optimizations were performed on both the model and the full β AspAMP intermediate at the Hartree-Fock (HF) level with a 6-31+G(*d*) basis set (i.e. HF/6-31+G(*d*) level). This and all other *ab initio* calculations were performed using the GAUSSIAN 03 program.¹¹⁴ HF was chosen as the model chemistry in an effort to develop parameters consistent with the parameterization approach taken with CHARMM and in the development of parameters for other similar molecules.¹¹⁵ For the basis set, a diffuse function was added because of the negative charge carried by β AspAMP and the model complex, and the single polarization function was added to help account for the presence of phosphorus. The geometric data is presented in Table 2-3.

Partial Atomic Charges

The partial charges calculated for the model complex were then transferred to the corresponding atoms in β AspAMP. This procedure represents the largest deviation from the parameterization methodology generally employed in the development of CHARMM. Ordinarily, the partial charges in CHARMM are calculated by reproducing *ab initio* interaction energies and geometries between model compounds and water.^{111,112} The approach taken instead was the calculation of electrostatic potential charges according to the CHelpG scheme¹¹⁶ at the HF/6-31+G(*d*) level for the model complex in GAUSSIAN 03. The CHelpG scheme was

Table 2-3. Model complex, *ab initio* and CHARMM geometry data.

Bond	Internal Coordinate	CHARMM	<i>Ab Initio</i>
b1	PA-O1A	1.480	1.472
b2	PA-O3A	1.620	1.687
b3	CG-O3A	1.325	1.315
b4	OD1-CG	1.259	1.198
b5	CB-CG	1.529	1.509
b6	HB1-CB	1.111	1.083
b7	HB2-CB	1.110	1.082
b8	HB3-CB	1.110	1.082
b9	O2A-PA	1.481	1.468
b10	O5'-PA	1.586	1.606
b11	C5'-O5	1.431	1.401
b12	H5'-C5'	1.113	1.084
b13	H5''-C5'	1.113	1.084
b14	HC5'-C5'	1.111	1.082
b1b9	b1+b9/2.0	1.481	1.470
b2b10	b2+b10/2.0	1.603	1.647
Angle	Internal Coordinate	CHARMM	<i>Ab Initio</i>
a1	O3A-PA-O1A	109.6	106.3
a2	CG-O3A-PA	175.7	128.2
a3	OD1-CG-O3A	122.6	120.4
a4	CB-CG-O3A	114.8	117.6
a5	HB1-CB-CG	110.0	109.7
a6	HB2-CB-CG	109.5	109.9
a7	HB3-CB-CG	109.6	108.8
a8	O3A-PA-O2A	107.6	107.8
a9	O3A-PA-O5'	103.5	96.0
a10	PA-O5'-C5'	116.6	120.5
a11	O5'-C5'-H5'	111.7	110.8
a12	O5'-C5'-H5''	111.2	107.1
a13	O5'-C5'-HC5'	108.7	111.0
a1a8	a1+a8/2.0	108.6	107.0
Dihedral	Internal Coordinate	CHARMM	<i>Ab Initio</i>
d1	CG-O3A-PA-O1A	-49.3	65.4
d2	OD1-CG-O3A-PA	133.1	-175.7
d3	CB-CG-O3A-PA	-18.5	4.7
d4	HB1-CB-CG-O3A	177.7	-60.4
d5	HB2-CB-CG-O3A	-62.3	57.3
d6	HB3-CB-CG-O3A	57.6	178.6
d7	O2A-PA-O3A-CG	78.3	-69.2
d8	O5'-PA-O3A-CG	-168.7	178.7
d9	C5'-O5'-PA-O1A	-58.1	37.9
d10	H5'-C5'-O5'-PA	55.7	66.4
d11	H5''-C5'-O5'-PA	-67.4	-174.8
d12	HC5'-C5'-O5'-PA	174.4	-55.2

Bonds are measured in angstroms, while angles and dihedrals are measured in degrees.

selected to calculate the partial charges on the model complex because of its invariance to internal and external rotation. This makes CHelpG a good fit for charge calculation, as the overall focus of this work is essentially a conformational-based study of the β AspAMP intermediate within *E. coli* AS-B. The calculated partial charges were rounded to the nearest hundredth and yielded the appropriate total complex charge of -1. These charges were then integrated into the β AspAMP and the charges on the extra hydrogens in the model were added to their respective carbons (HC5' \rightarrow C5' and HB3 \rightarrow CB). A full list of the charges on both the model complex and the β AspAMP can be seen in Table 2-2.

Force Constants and Dihedral Rotation

The force constants for the bonds and angles were initially taken from already existing parameters in CHARMM. The idea was to use parameters for similar bonds and angles that had previously been rigorously parameterized. A full list of those parameters with the corresponding origins for each parameter is listed above in Table 2-1. The (initial) dihedral parameters were derived from a variety of data, including the PES that was calculated using GAUSSIAN 03 at the HF/6-31+G(*d*) level, already existing dihedral parameters in CHARMM, and from a dihedral model developed in Microsoft Excel. The periodicities, *n*, were the first parameters assigned for the dihedrals. The periodicities were assigned to each dihedral based on the *ab initio* PES, the geometries of the atom at the center of the dihedrals, and the Excel model. The dihedral force constants, K_{ϕ} , were obtained by both calculations (from the PES) and through borrowing parameters from similar dihedrals already parameterized in CHARMM. The phase angles, δ , were assigned from the PES, the Excel model, and Newman projections about the dihedrals.

The parameterization of the dihedral began with the calculation of a 2-dimensional potential energy surface. A surface was compiled by a series of HF/6-31+G(*d*) level *ab initio*

calculations, performed in GAUSSIAN 03, in which the two primary dihedrals (CB-CG-O3A-PA(ϕ_1) vs. CG-O3A-PA-O5'(ϕ_2)) were constrained while the rest of the structure was allowed to relax to the lowest energy conformation allowed by the specific set of dihedrals. To ensure adequate coverage, both dihedrals were increased in 10° increments from 0° to 360°. This created a 2-dimensional energy surface comprised of 1,296 minimized model complex structures. The resulting *ab initio* PES is shown in Figure 2-3. As can be seen in the figure, the *ab initio* surface is symmetrical and has three minima. The global minimum seen at a ϕ_1 _ ϕ_2 of 0°_180° was set as zero. Two additional minima are seen at a ϕ_1 _ ϕ_2 of 0°_70° and at a ϕ_1 _ ϕ_2 of 170°_60° with energies of 2.99 kcal/mol and 3.42 kcal/mol above the zero, respectively. The overall maxima for the surface is seen at a ϕ_1 _ ϕ_2 of 120°_150° and has an energy of 10.40 kcal/mol above the zero. The *ab initio* surface was used to estimate the initial periodicity and force constants for the two primary dihedrals shown in Table 2-1. One-dimensional sections were originally taken from the surface to estimate the periodicity of each dihedral. This approach yielded an initial guess of $n = 2$ for ϕ_1 (CB-CG-O3A-PA) and $n = 1$ and 3 for ϕ_2 (CG-O3A-PA-O5'). Using the equation

$$E = \frac{k_{11}}{2} (1 + \cos(2\phi_1 - 180)) + \frac{k_{21}}{2} (1 + \cos(\phi_2)) + \frac{k_{22}}{2} (1 + \cos(3\phi_2)) \quad (2-2)$$

and three different combinations of energies with the associated ϕ_1 _ ϕ_2 (E, ϕ_1, ϕ_2) from the *ab initio* surface, the force constants k_{11} , k_{21} , and k_{22} were estimated at 11.2, 4.1, and 3.5, respectively. As seen in Equation 2-2, the phase angles, δ , for ϕ_1 and ϕ_2 were set at 180° and 0°, respectively, from the *ab initio* surface. The initial parameters for the two supplementary dihedrals also required by CHARMM (PA-O3A-CG-OD1 and O2A-PA-O3A-CG) were estimated. The initial parameters for the supplementary dihedral PA-O3A-CG-OD1 was

estimated to be the same as the CB-CG-O3A-PA dihedral parameters, while the initial parameters for the supplementary dihedral O2A-PA-O3A-CG were taken from ON3-P-ON2-CN7 in dimethylphosphate.

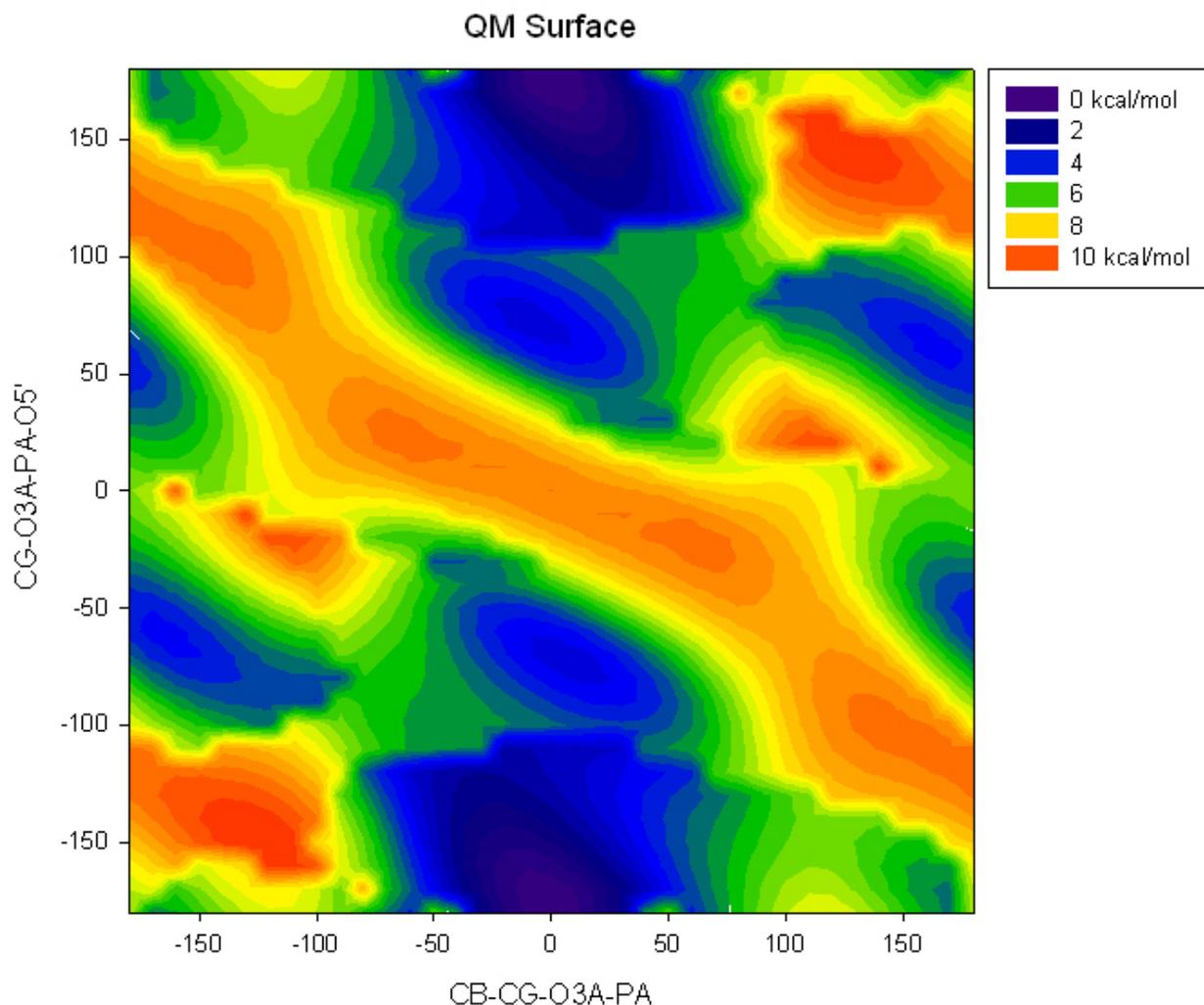


Figure 2-3. The 2-dimensional *ab initio* energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex. The global minimum energy (kcal/mol) was set to zero with all other energies offset equivalently.

The dihedral parameterization began by determining a baseline of all the forces contributing to the energy surface in CHARMM minus the contribution to the energy made by the dihedrals. The resulting surface can be seen in Figure 2-4. The initial dihedral parameters

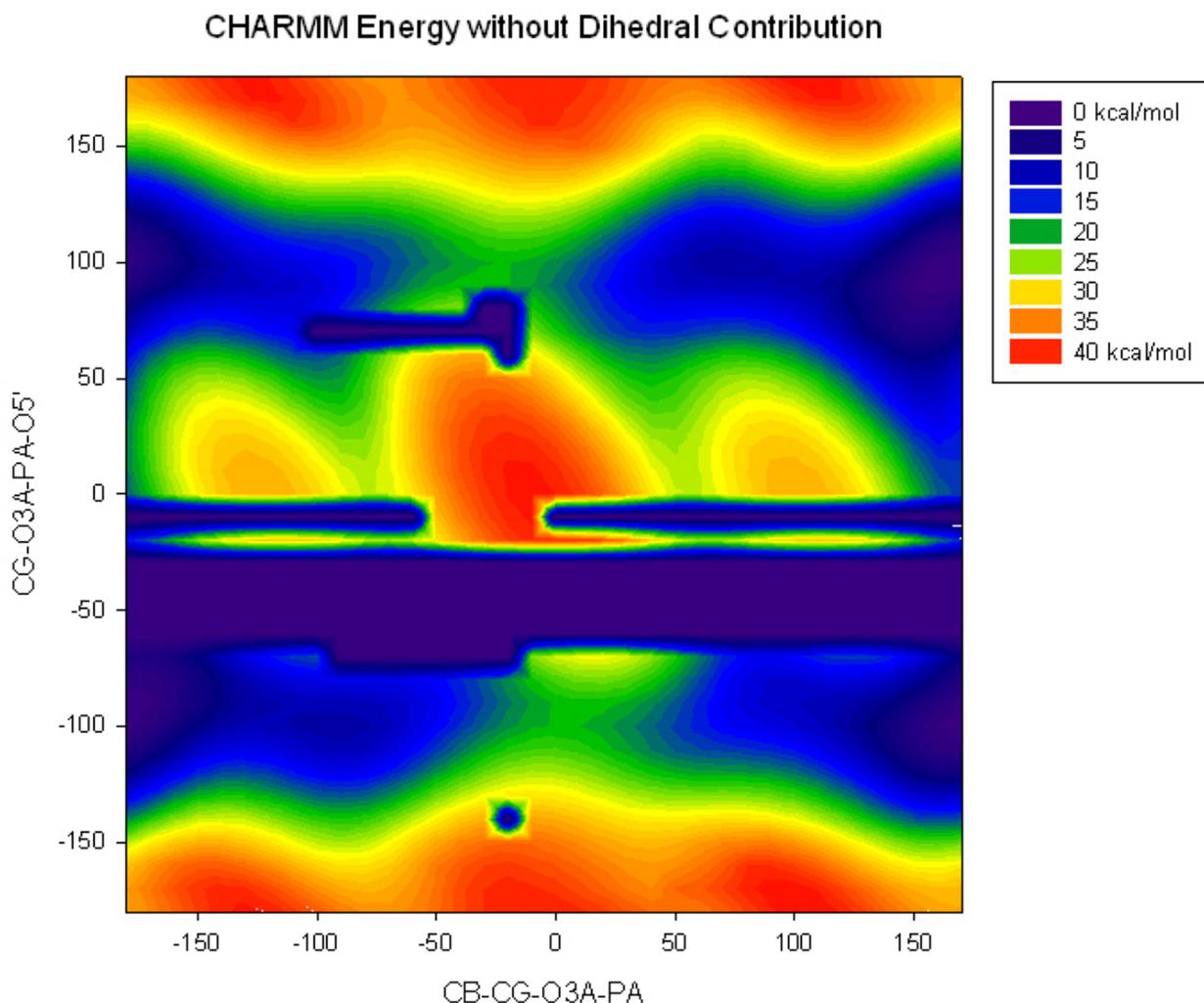


Figure 2-4. The 2-dimensional CHARMM energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex without energy contributions from the dihedrals in Table 2-1. In this figure and in the subsequent dihedral surfaces generated by CHARMM (Figure 2-4, Figure 2-6 and Figure 2-7), there are regions of solid blue where CHARMM was unable to minimize the structure with the specific combination of dihedrals resulting in incorrect energies that had to be set to zero for the remaining surface to be appropriately depicted.

were then added to the energy calculation. Upon the addition of the initial parameters, it was clear that adjustments would be necessary to achieve better agreement of the 2-dimensional energy surfaces calculated by CHARMM with the *ab initio* surface. The first step was adding a second term to help describe the PA-O3A-CG-CB dihedral. Periodicities of both 1 and 3 were

tested and only the term with $n = 1$ seemed to improve the empirical surface calculated by CHARMM. The adjustments of the force constant for this term eventually yielded a K_ϕ value of 6.0 as the best fit. The two supplementary dihedrals were also very closely inspected. Newman projections looking down the two separate dihedrals of the lowest energy conformer of the model complex led to the alteration of the initial phase angles for the dihedrals in question. The force constants for the two supplementary dihedrals depicted in Figure 2-5 were also changed. As was done with the term added to the PA-O3A-CG-CB dihedral, the force constants for these two dihedrals were adjusted to obtain the best agreement. This process produced a K_ϕ for the O2A-PA-O3A-CG dihedral of 1.0 and a K_ϕ for the PA-O3A-CG-OD1 dihedral of 0.3. The periodicities for these two dihedrals were estimated in the belief that the periodicity of the PA-O3A-CG-OD1 dihedral would mimic the primary periodicity of the main dihedral, PA-O3A-CG-CB, and that the periodicity for the O2A-PA-O3A-CG dihedral would be similar to the ON3-P-ON2-CN7 dihedral in dimethylphosphate. The resulting PES calculated by CHARMM with all

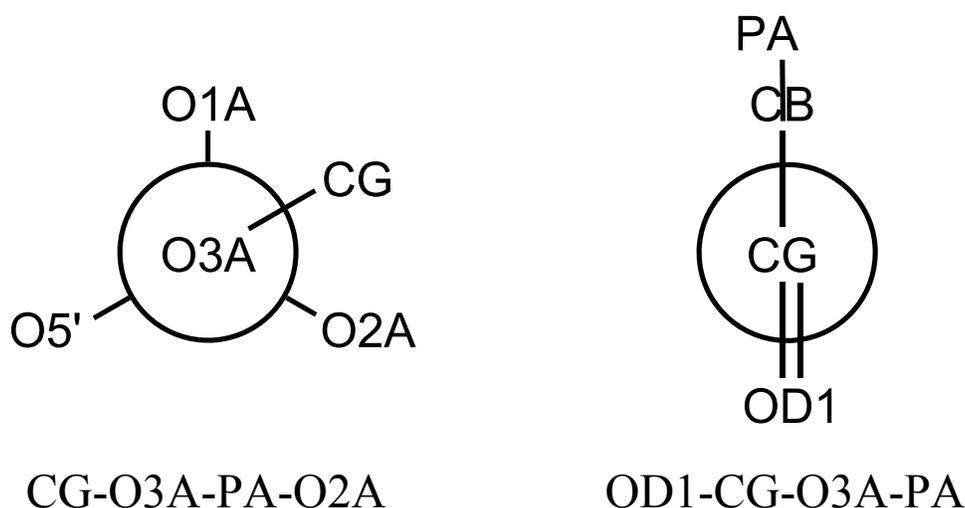


Figure 2-5. Newman projections for the dihedrals O2A-PA-O3A-CG and PA-O3A-CG-OD1. The first dihedral is looking down the O2A-PA-O3A-CG dihedral with truncations made at O5' and CG for clarity. The projection shows a 60° dihedral. The second dihedral is looking down the PA-O3A-CG-OD1 dihedral with truncations made at PA and CB for clarity. The projection shows a 0° dihedral.

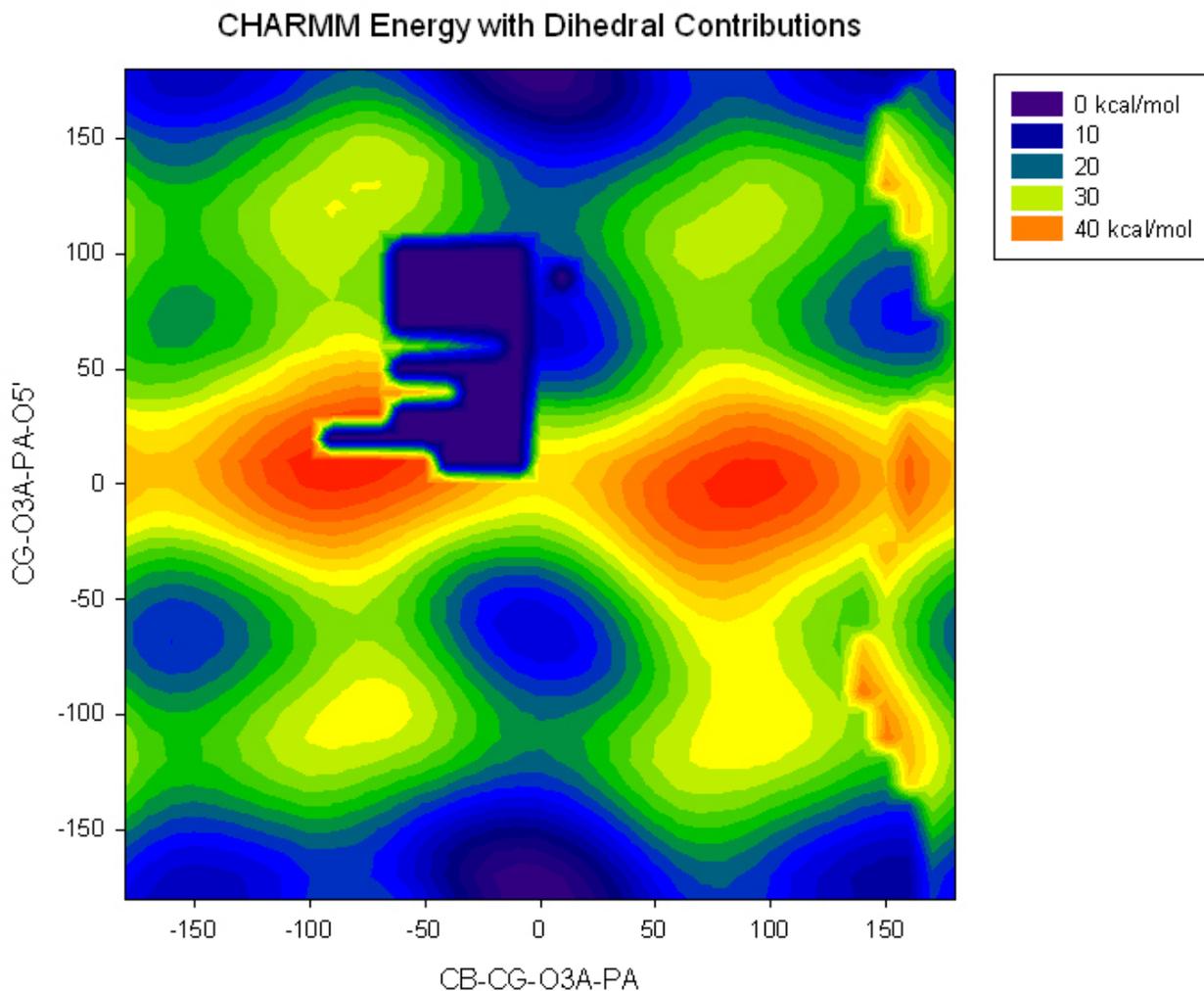


Figure 2-6. The 2-dimensional CHARMM energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex with energy contributions from the dihedrals.

of the dihedral adjustments can be seen in Figure 2-6. As illustrated in Figure 2-6, this combination of parameters produced an energy surface in CHARMM that has a global minimum at the same φ_1 _ φ_2 combination of 0° _ 180° and two other minima in the same general areas as the on the *ab initio* surface. However, the major deficiency of this energy surface is a local minimum in the area around a φ_1 _ φ_2 of 120° _ 150° where the global maximum is located on the *ab initio* surface. This was considered a significant problem as the primary objective in reproducing energy surfaces in a parameterization is to match the overall shape of the energy surface.

Reproducing the energy barriers within the surface is considered a secondary objective, as it is an extremely challenging feat to accomplish. Therefore, the glaring omission of a maximum where the global maximum should be, prompted continuing the parameterization efforts to achieve a better fit.

Because the dihedral parameters were believed to be close to their optimal values, a second look was taken at the bond and angle parameters that were estimated from already existing parameters within CHARMM. Pavelites, *et al.* determined that a negative force constant on a Urey-Bradley term added to the bridging bond in methyl diphosphate was necessary for CHARMM to be able to reproduce the dihedral rotational energy surface.¹¹⁵ After reexamining the PA-O3A-CG bond angle, the initial estimates based on the P-ON2-CN9 angle in the same methyl diphosphate were replaced with the parameters for the P-ON2-P angle. This included the negative Urey-Bradley force constant that was critical in reproducing their *ab initio* energy surface. This change left a K_θ of 15.0 with the same θ_0 as before, a K_{UB} of -40.0, and a S_0 of 2.8 as the parameters on the PA-O3A-CG angle of the model complex and β AspAMP.

Recalculating the CHARMM 2-dimensional energy surface while only changing the PA-O3A-CG angle parameters yielded the energy surface shown in Figure 2-7. The major change in the PES calculated by CHARMM between Figure 2-6 and Figure 2-7, is that now a maximum is in the area around a $\varphi_1\text{-}\varphi_2$ of $120^\circ\text{-}150^\circ$. This is a critical shift in improving the overall fit of CHARMM PES with the equivalent *ab initio* surface. There is a loss of the third minimum at a $\varphi_1\text{-}\varphi_2$ of $170^\circ\text{-}60^\circ$ in the surface calculated with the new set of parameters; however, this was believed to have been an acceptable tradeoff for having a maximum near the corresponding $\varphi_1\text{-}\varphi_2$ of $120^\circ\text{-}150^\circ$ as seen in the *ab initio* surface in Figure 2-3. This final surface does retain the global minimum at a $\varphi_1\text{-}\varphi_2$ of $0^\circ\text{-}180^\circ$ with a local minimum at a $\varphi_1\text{-}\varphi_2$ of $0^\circ\text{-}70^\circ$. With this

CHARMM energy with negative Urey-Bradley force constant

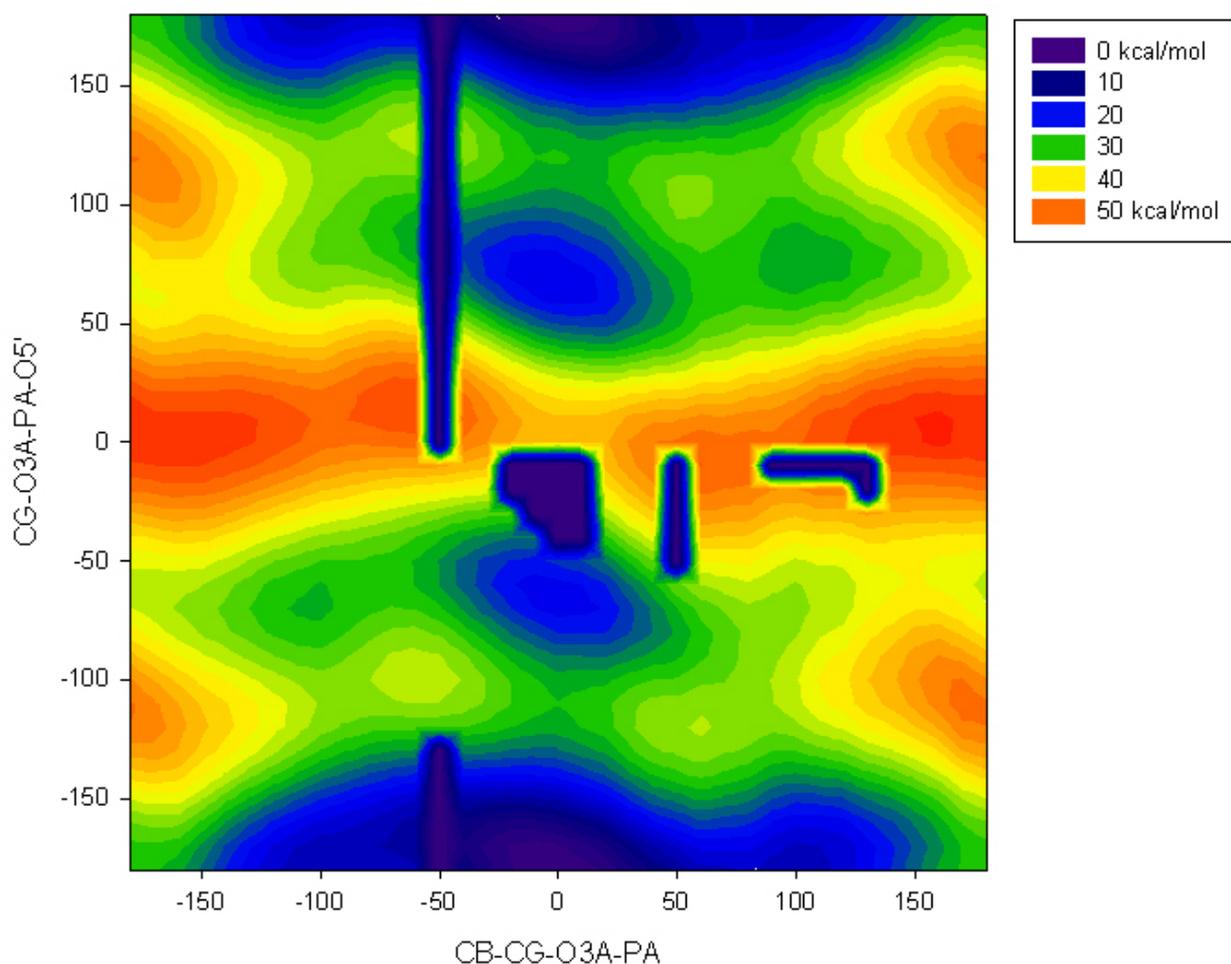


Figure 2-7. The 2-dimensional CHARMM energy surface for dihedral angle CB-CG-O3A-PA versus dihedral angle CG-O3A-PA-O5' of the model complex using the final set of developed parameters. This graph used the exact same parameters as Figure 2-6 with the exception of the changes made to the PA-O3A-CG angle parameters as described in the text.

surface believed to be the best estimate of the *ab initio* surface to this point, further refinement to the force constants could be made by comparing the vibrational spectrum for the model complex as calculated in CHARMM with the target vibrational spectrum that was calculated for the model complex.

Vibrational Spectra for the Model Complex

The target vibrational frequency data for parameterizing the model complex was calculated via GAUSSIAN 03, and was calculated at the HF/6-31+G(*d*) level of theory to maintain consistency with all previous *ab initio* calculations performed for the model complex. For comparison of the *ab initio* vibrational frequencies with those calculated by CHARMM, a scaling factor was used to reduce the HF results, which are known to be systematically overestimated.¹¹⁷ A scale factor of 0.897 was used here in accordance with the findings of Scott and Radom for HF/6-31+G(*d*) level results.¹¹⁷ Table 2-4 shows the 39 vibrational frequencies calculated by HF and then scaled, as well as the vibrational frequencies calculated by CHARMM. Generally, differences of 100 cm⁻¹ or less are considered to be in relative agreement. As seen in Table 2-4, only four CHARMM frequencies are more than 100 cm⁻¹ different from the scaled *ab initio* frequencies. The overall rms deviation is 35.08 cm⁻¹ with the average percent difference being 9.6%. The frequencies below approximately 300 cm⁻¹ are the most biologically relevant modes that normally occur in an MD simulation. These modes dominate the large-scale motions. The first eight frequencies, which comprise this low frequency region, all show good agreement with the *ab initio* results. The next ten vibrational frequencies extend from approximately 300 cm⁻¹ to approximately 1,000 cm⁻¹, and also exhibit good overall fit. There are two frequencies (modes 15 and 17) that show a difference greater than 100 cm⁻¹ from the corresponding *ab initio* modes. Although this difference is not ideal, because they first appear in the second half of the modes where the less biologically important frequencies occur, the large difference is more acceptable. Again, only two outliers are seen in the remaining span of modes that stretch from 1,000 cm⁻¹ to 3,000 cm⁻¹. In all, only four frequencies calculated by CHARMM are separated by more than 100 cm⁻¹ from the corresponding frequencies calculated by HF. The good agreement in vibrational spectra demonstrates a good level of parameterization.

Table 2-4. Vibrational frequencies for the model complex as calculated by HF and CHARMM.

Mode	<i>Ab initio</i> Frequency ^a (cm ⁻¹)	CHARMM Frequency (cm ⁻¹)	ABS(<i>AI</i> – CHARMM)
1	51.2	44.6	6.6
2	59.2	89.6	30.4
3	86.4	109.3	22.9
4	111.5	185.3	73.8
5	153.1	189.6	36.5
6	172.2	229.5	57.3
7	214.2	241.2	27.0
8	306.3	323.8	17.5
9	350.2	360.5	10.3
10	365.4	430.6	65.2
11	428.9	461.1	32.2
12	478.2	501.6	23.4
13	523.2	510.1	13.1
14	546.0	564.9	18.9
15	589.2	723.4	134.2
16	703.6	778.5	74.9
17	752.3	853.6	101.3
18	948.0	930.9	17.1
19	1018.5	1023.4	4.9
20	1063.5	1035.6	27.9
21	1071.1	1115.2	44.1
22	1099.5	1134.1	34.6
23	1163.4	1145.3	18.1
24	1190.4	1164.1	26.3
25	1259.9	1275.6	15.7
26	1289.3	1399.0	109.7
27	1403.1	1426.9	23.8
28	1452.3	1436.1	16.2
29	1453.9	1438.4	15.5
30	1455.0	1456.7	1.7
31	1473.8	1482.1	8.3
32	1487.7	1622.4	134.7
33	1743.9	1707.8	36.1
34	2880.6	2798.6	82.0
35	2897.8	2853.7	44.1
36	2934.1	2860.8	73.3
37	2952.6	2863.1	89.5
38	2964.9	2918.1	46.8
39	2975.6	2919.6	56.0

^a *Ab initio* frequencies have been scaled by 0.897 as outlined in the text.

An attempt to optimize the force constants further using the vibrational spectra was made using the Automated Frequency Matching Method (AFMM) program.¹¹⁸ AFMM uses a Monte

Carlo-like algorithm that adjusts parameters iteratively until the optimal fit of the vibrational spectra of CHARMM with the target vibrational spectra from *ab initio* methods is obtained.¹¹⁸

After multiple attempts to use AFMM to optimize the force constants, without a single example of improvement in the CHARMM energy surface, this method was abandoned and the previously reached set of parameters was kept as the final parameter set.

Conclusions

The parameterization of the model complex, and hence the β AspAMP intermediate, centered primarily on the two unique dihedrals in this structure. The initial parameters for the one bond and three angles were taken from similar structures. The equilibrium bond length and angles were taken from the globally minimum structure found in the HF/6-31+G(*d*) calculated PES (Figure 2-3). As was discussed earlier, the partial atomic charges for the model complex were determined via the CHelpG scheme and are reported in Table 2-2. This leaves the absent dihedral parameters, which were major energy contributors to the complex, as the primary focus of the parameterization efforts.

The final set of parameters for the model complex provides good agreement with the lowest energy conformation found with HF. The overall rms difference for the bonds is 0.017 Å with respect to the *ab initio* results, while the angles have an overall rms difference of 12.2°. In general, CHARMM found slightly longer bonds than were seen from the *ab initio* structure with the exception of the two bridging oxygens (O5' and O3A). In both of these cases, the parameter set produced slightly shorter bonds than were seen in the *ab initio* structure. The largest absolute deviation of CHARMM with the *ab initio* result was only about 0.07 Å. The rms difference in the angles is significantly large at 12.2°; however, this large rms can be attributed to a single significant outlier. As signified by the bolded value in Table 2-3, angle 2 (CG-O3A-PA) is found

to be 175.7° on a bent oxygen rather than the value that is expected around 120°. If this data point is removed, the rms difference for the angles drops significantly to 2.1°. This means that if angle 2 were excluded, the largest single deviation is only 7.5°. Dihedral 2 (OD1-CG-O3A-PA) is also significantly off from its target, however, this large deviation can again be attributed to the previous angle (CG-O3A-PA) that was off. Overall, the lowest energy conformation produced for the model complex by CHARMM using the final parameter set demonstrates good agreement with the lowest energy conformer found at the HF/6-31+G(*d*) level.

Overall, the parameters developed here, in accordance with the CHARMM empirical force, do an adequate job of representing the conformational properties of the model complex as established by *ab initio* means. The final set of parameters can be seen below in Table 2-5.

Table 2-5. Final parameters for the model complex.

Structure	Bonds	K_b	b_0		
	CC-ON2	650.0	1.316 Å		
Structure	Angles	K_θ	θ_0	K_{UB}	S_0
	P-ON2-CC	15.0	128.2°	-40.0	2.8
	ON2-CC-CT2	70.0	117.6°		
	ON2-CC-OC	98.0	120.3°		
Structure	Dihedrals	K_ϕ	n	δ	
	ON2-P-ON2-CC	4.1	1	0.0°	
	P-ON2-CC-CT2	3.5	3	0.0°	
	P-ON2-CC-CT2	6.0	1	180.0°	
	P-ON2-CC-OC	11.2	2	180.0°	
	P-ON2-CC-OC	0.3	2	0.0°	
	ON3-P-ON2-CC	1.0	3	60.0°	

CHAPTER 3 DEVELOPMENT OF A COMPUTATIONAL MODEL OF ASPARAGINE SYNTHETASE-B

Introduction

The development of a computational model of a protein can offer valuable insights into both the structural and thermodynamic properties of a specific system. Traditionally, the goal of any protein structure investigation, whether computational or experimental, has been to identify the native conformation of the protein. This is believed to be the global minimum in potential energy of the system and the state that is biologically relevant. However, many enzymatic systems undergo conformational changes as the reaction progresses. Thus, the term “global minimum” can actually represent multiple protein conformations depending on what substrates or ligands are bound. Identifying the global minimum of an enzymatic system at different stages of the reaction provides insight into the conformational changes the system experiences, as well as the critical active site interactions at each stage. These important active site interactions can then be exploited in the rational design of targeted enzyme inhibitors.

Various techniques have been developed for the purpose of conformational sampling with the goal of finding the global minimum conformation.^{119,120} Some of the approaches taken for structural prediction include *ab initio* simulated annealing,¹²¹ molecular dynamics-cluster analysis (MD-CA),¹²² relaxed annealing,¹²³ pressure annealing,¹²⁴ Monte Carlo with simulated annealing,¹²⁵ Monte Carlo with molecular dynamics,¹²⁶ molecular dynamics with simulated annealing,¹¹⁹ molecular dynamics with explicit water,¹²⁷ multi-conformation simulated annealing pseudo-crystallographic refinement (MCSA-PCR),¹²⁸ high temperature molecular dynamics,¹²⁹ and replica exchange simulations¹³⁰ to name a few. The common thread that appears in most of the above approaches in one way or another is the simulated annealing (SA) method.^{131,132}

Asparagine synthetase is an exceptional candidate for structural optimization via simulated annealing. AS has proven to be a difficult enzyme to crystallize; to date, only a single crystal structure exists of a glutamine-dependent prokaryotic form of AS (AS-B).⁵⁷ While extremely valuable, as the only structural insight into glutamine-dependent AS, this structure still has many shortcomings. If modeled correctly, the improved structure would provide a more nearly complete picture of the enzyme during catalysis. Thorough understanding of the synthetase active site is of particular interest, because the roles of explicit residues involved in catalysis will provide a detailed platform for rational design of inhibitors. Furthermore, the model could also be used to help understand experimental results from site-directed mutagenesis studies. Unfortunately, the crystal structure of AS-B is incomplete as it does not have the coordinates for the residues of two small loop regions embedded in the sequence, nor does it have the coordinates for the final 37 residues. The AS-B crystal structure also only offers a limited perspective of the critical synthetase active site as it only has AMP, an eventual product, bound.⁵⁷

The work presented here will expand on the initial efforts of Dr. Yun Ding towards the development of a model of AS-B.⁷⁸ Two models of AS-B will be optimized. The first model will include Gln bound in the glutaminase active site and Asp, ATP and Mg²⁺ bound in the synthetase active site of AS-B. This model represents AS-B in its presumed initial state, prior to the initiation of catalysis. The second model will show AS-B with Gln again bound in the glutaminase site and with the reaction intermediate β AspAMP, PP_i and Mg²⁺ bound in the synthetase active site. This model represents AS-B in the early stages of catalysis, after formation of the intermediate in the synthetase active site.

Methods

Simulated Annealing Methodology

In SA, the system of interest is heated to a sufficiently high temperature to give the system enough kinetic energy to surpass all barriers in the potential energy landscape before being slowly cooled to the global minimum. If the system reaches thermal equilibrium at each temperature, then the system can be compared to the Boltzmann distribution, which yields the most likely population of a state i with energy E_i at the temperature T :

$$n_i = \frac{\left[Ne^{\left(-E_i/kT\right)} \right]}{\sum_j e^{-E_j/kT}} . \quad (3-1)$$

Lower energy states become more probable as the temperature is lowered until it reaches absolute zero, where the system is in its lowest possible energy state. However, SA is in reality only an approximation, as an infinite number of time steps would be required and at each time step the system would be allowed to come to thermal equilibrium in order to reach the true global minimum.¹³³ Cooling the system becomes the most critical aspect of a SA procedure. A logarithmic cooling scheme has been shown to be the most efficient at reaching the minimum energy conformations;¹³⁴ however, it is still possible to reach the global minimum when not following a logarithmic cooling scheme if cooling is performed sufficiently slowly.¹³⁵ Another option is running multiple SA experiments to increase the likelihood that the method finds the global minimum.¹³⁶

Simulated Annealing Procedure

Three different SA procedures were utilized to arrive at the global minimum conformations for the two model systems. They all started from the same minimized model systems, which are

described in detail in a subsequent section. Long-range interactions were calculated via the Particle-mesh Ewald (PME) method¹³⁷⁻¹³⁹ within the constructs of the periodic boundary condition (PBC)^{140,141} to avoid the improper use of cut-off distances. The setup for both the PBC and the calculation of the nonbonded interactions remained unchanged in every calculation presented here. The setup and calculation can be seen in any of the CHARMM example input files in the Appendix. The three different SA methods all began from a single heating procedure; model systems were heated from 0 K to 600 K over the course of 240 ps with the temperature increasing 5 K every 2 ps at a constant pressure of 1 atm. As expected, in each case the volume of the system increased during the heating procedure.

In the first set of SA experiments (SA Run 1), the heated systems were held at constant pressure (1 atm) at the peak temperature (600 K) via the Nose-Hoover method,^{142,143} while the volume was allowed to expand as the temperature was distributed evenly throughout the system over 100 ps. The systems were then cooled linearly from 600 K to 300 K over 500 ps at constant pressure (1 atm). Ten structures were collected at 50 ps intervals and at each interval; the system was placed at constant temperature via the Nose-Hoover method for 0.1 ps to serve as a quenching mechanism to allow the temperature to quickly distribute evenly.

In the second set of SA experiments (SA Run 2), the heated systems were immediately cooled after reaching 600 K by the same protocol described for the first set of SA experiments. Once the systems were cooled to 300 K, they were equilibrated over 100 ps with a constant pressure (1 atm) and at a constant temperature (300 K), via the Nose-Hoover method.

In the final set of SA experiments (SA Run 3), after heating, 100 ps of run time allowed the volume to expand dramatically as the pressure was held constant at 0.025 atm and the temperature was held constant by Nose-Hoover methods at 600 K. The systems were then cooled

under high pressures (up to 1,000 atm) until the systems reached 300 K and were as close to their original volumes as possible. Again, in the cooling process, ten structures were collected at 50 ps intervals and quenching steps were implemented as previously described. Every SA experiment finished with the minimization of the systems at each of the ten temperatures where structures were collected. The Appendix contains the input files for the heating, equilibration, cooling and final minimization for the SA Run 1 for the model system that contains the β AspAMP intermediate. Only slight manipulations to these core input files and rearrangement of the order in which the files were run, as discussed above with the various SA protocols, were required to produce all of the remaining input files.

Model Systems

AspATP model system

The first of the two model systems that were optimized in this work was taken from the work of a former group member, Dr. Yun Ding.⁷⁸ This model consists of the *E. coli* glutamine-dependent asparagine synthetase (AS-B) with glutamine bound in the N-terminal glutaminase active site and aspartate, ATP and Mg^{2+} bound in the C-terminal synthetase active site. The protein structure contains all residues from 1 – 516, including the two small loops from the interior of the protein that were unresolved in the crystal structure of AS-B (PDB code: 1ct9).⁵⁷ The two unresolved loops were added as part of the dissertation work of Dr. Ding.⁷⁸ As in the AS-B crystal structure, the first residue of the model is an alanine rather than the naturally occurring cysteine. The initial structural conformation of the system was taken as the “best” structure as identified from the work of Dr. Ding. There were 306 water molecules added to this complex based on their positions in the crystal structure. The complete complex (AS-B/Gln/Asp/ATP/ Mg^{2+} /H₂O) has a total charge of -20 (AS-B = -17; Gln = 0; Asp = -1; ATP = -4; Mg = +2). The complex was then solvated in a cubic box of water using the TIP3 water model

and NaCl was added to the system in the form of Na^+ and Cl^- ions to neutralize the system. The system was then minimized to an energy convergence of 0.00001 with the SHAKE algorithm¹⁴⁴ in use. For the rest of this work, this system setup will be referred to as the AspATP model system.

β AspAMP model system

The second model system optimized in this work was initially built from the coordinates of the AspATP model complex described above. Using the AspATP model as a starting structure, the β AspAMP intermediate was built by manipulation of the aspartate and ATP structures. In the active site, aspartate was moved closer to ATP as the carboxylate side chain was rotated towards the α -phosphate of ATP where it attacks to form the intermediate. A bond was then drawn to join the carboxylate tail of aspartate to an oxygen on the α -phosphorus. Next, the bridging bond between the α - and β -phosphates was broken to create PP_i and the β AspAMP intermediate. Aspartate was the only molecule to move during this procedure. Both the intermediate and PP_i retained their initial coordinates from ATP in the AspATP model. Figure 3-1a below shows the original positions of the Asp and ATP and Figure 3-1b shows the newly created β AspAMP and PP_i . As in the AspATP model, the 306 crystal water molecules were added to this complex based on their positions in the crystal structure. The complete complex (AS-B/Gln/ β AspAMP/ PP_i / Mg^{2+} / H_2O) has a total charge of -20 (AS-B = -17; Gln = 0; β AspAMP = -1; PP_i = -4; Mg = +2). The complex was then solvated in a cubic box of water using the TIP3 water model and NaCl was added to the system in the form of Na^+ and Cl^- ions to neutralize the system. The system was then minimized to an energy convergence of 0.00001 with the SHAKE algorithm.¹⁴⁴ For the remainder of this work, this system setup will be referred to as the β AspAMP model system.

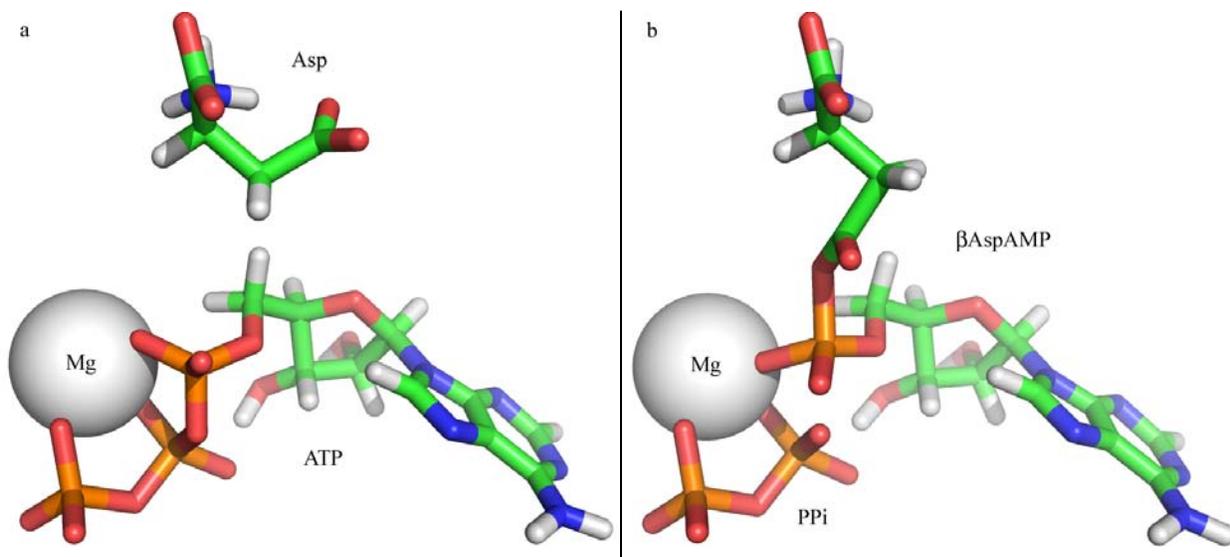


Figure 3-1. Creation of β AspAMP intermediate. (a) Original positions of the ATP and aspartate from the AspATP model. (b) The β AspAMP intermediate created by the joining of aspartate and ATP as described in the text.

Results and Discussion

AspATP Model

The three simulated annealing runs that were performed on the AspATP model system began from a single minimized initial structure and a single heating run to 600 K over 240 ps. This method was chosen to minimize the computational efforts required so that more annealing protocols could be tested. As was detailed more thoroughly in the previous section, the three runs then varied in their equilibration and cooling protocols, before all of the resulting structures were minimized. The resulting energies of the three complete SA protocols can be seen in Table 3-1. The resulting structures from the equilibrations of Run 1 and Run 3 were not minimized because the equilibrations were carried out at the peak temperature of 600 K and the resulting structures were so deviant from the initial structures that a minimization would not have produced final structures of any relevance.

Table 3-1. Final energies for the SA Runs for the AspATP model system

	Energy (kcal/mol)		
	SA Run 1	SA Run 2	SA Run 3
Initial		-275,457.76	
600 K ^a		-67,066.50	
100 ps run at 600 K ^a	-44,075.66	-284,660.89 ^b	-44,800.98
570 K	-281,372.96	-281,214.41	-281,372.96
540 K	-282,930.30	-281,497.79	-281,533.30
510 K	-281,943.36	-282,295.30	-281,943.36
480 K	-	-283,983.00	-282,403.70
450 K	-286,284.01	-284,966.86	-282,445.26
420 K	-286,707.27	-285,230.17	-282,849.96
390 K	-287,037.59	-285,303.97	-282,742.20
360 K	-287,474.56	-283,631.96	-283,194.62
330 K	-287,564.08	-283,955.49	-283,757.58
300 K	-288,046.56	-284,453.55	-284,418.12

^a Structures were not minimized; ^b This energy value was for the minimized final structure for the equilibration at 300 K, as no equilibration was done at 600 K in SA Run 2. The minimization at 480 K in SA Run 1 continually failed.

The analysis for all the SA Runs began with a look at the initial structure. Figure 3-2 shows the potential hydrogen bonds that exist between active site residues of AS-B, free aspartate, and ATP in this model. Although this type of polar contact is only one piece in the assessment of the accuracy of model systems, it is a useful first step for comparison of the other structures. For clarity, the ATP and aspartate interactions have been split, rather than shown in the same view. Many of the residues shown interacting with the ATP have been identified as part of a common ATP binding sequence. Residues Ser-234, Leu-237, Asp-238, and Ser-239 are all a part of the known pyrophosphatase loop (SGGXDS),^{59,77} while Lys-449 also has been proposed to have ATP binding implications in AS.¹⁴⁵ A structural alignment was performed in PyMOL¹⁴⁶ to compare the heated system to the initial system. PyMOL estimated an rmsd (based on the backbone coordinates of the C α s) of 2.314 Å. In the alignment (Figure 3-3), the major difference is that the opening at the synthetase active site has grown. The loop shown in red on the heated

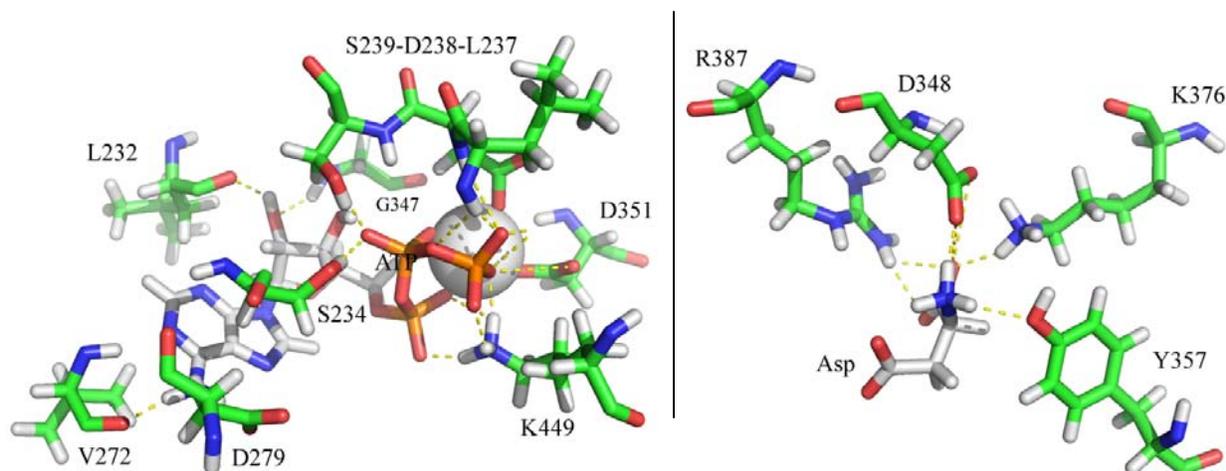


Figure 3-2. Synthetase active site interaction in the initial AspATP model system. The AS-B/ATP interactions are shown on the left with the AS-B/Asp interactions shown on the right. The AS-B is shown with the carbons colored green while both substrates are depicted with white carbons. The potential interactions are shown with the dashed yellow lines. The Mg^{2+} is also shown as a sphere in the view on the left. For the remaining atoms O = red, N = blue, P = orange, and H = white.

AspATP model (cyan) can be moved away from the active site. In addition, the entire protein structure has lost some tertiary structural features with the added energy, as can be seen with both sides of the heated protein moving away from the synthetase active site. This results in an opening of the synthetase active site, which allows the free aspartate to migrate away from the initial position. The ability of the aspartate to move freely could indicate that a second Mg^{2+} is absent in the model. This Mg^{2+} would help anchor the aspartate to prevent it from being pushed away from ATP by charge-charge repulsion. This could also suggest that the missing final 37 residues do serve a purpose in closing off the synthetase active site and restricting the position of the free aspartate.

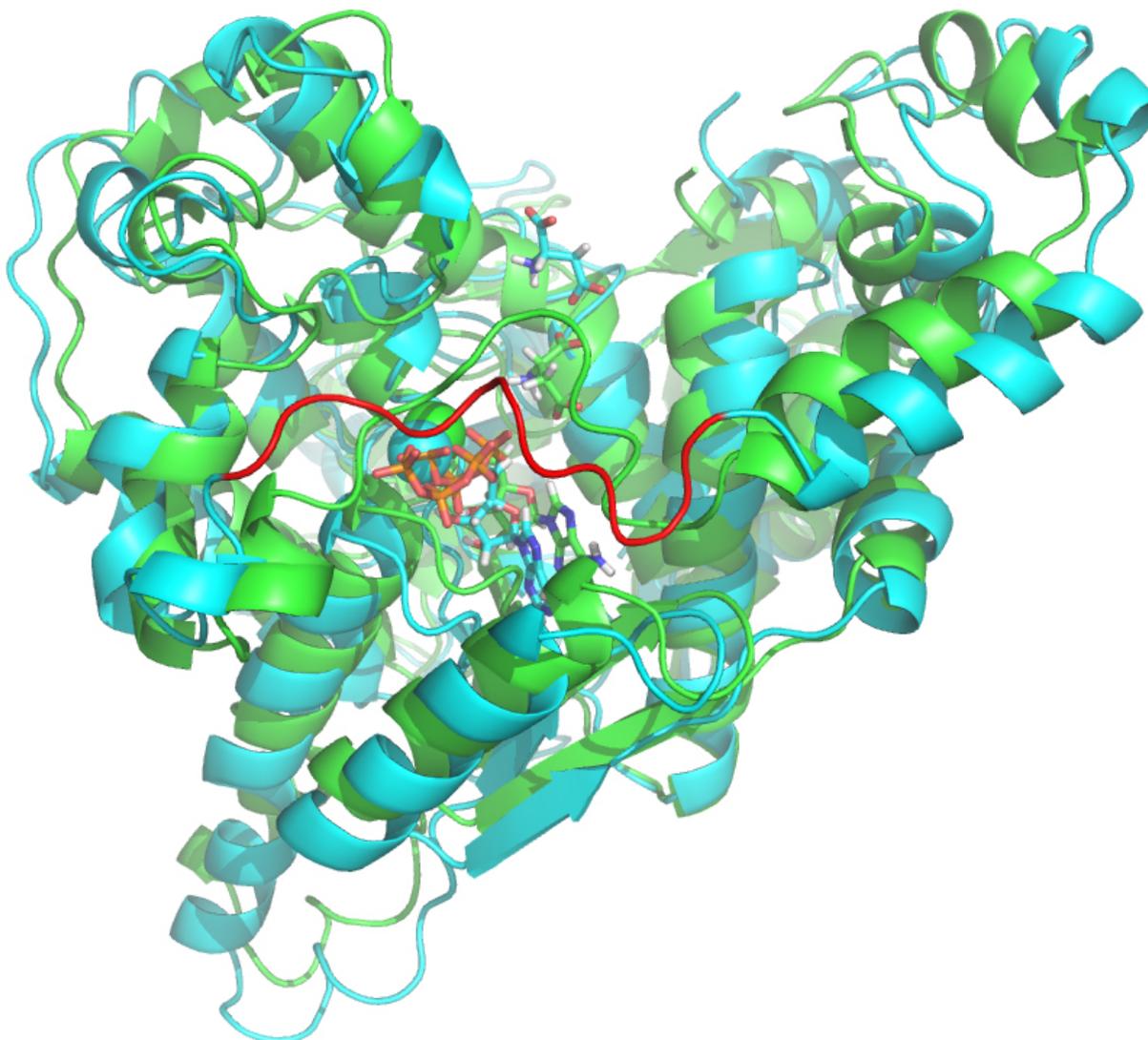


Figure 3-3. Structural comparison of the initial structure of the AspATP model with the same structure heated to 600 K. The initial structure is in green, while the heated structure is in cyan and red. The synthetase substrates, aspartate, ATP and Mg^{2+} , are shown in stick or sphere (Mg^{2+}) representation with the carbons colored to match their respective system.

Simulated annealing run 1

Upon reaching 600 K, the temperature was allowed to distribute evenly in the system for 100 ps, as described above. Finally, the system was cooled to 300 K over 500 ps, while structural snap-shots were written out in 50 ps, i.e. 30 K, increments. These 10 structures were minimized to obtain the final structures used in the analysis. Figure 3-4 shows four plots that track the

evolution of the system throughout the annealing process. The temperature, volume and energy plots were generated from the output files, while the rmsd plot was generated from the trajectory files created during the simulation. As expected, the volume of the system fluctuated since the pressure was held constant (1 atm) and the temperature was regulated.

The series of graphs shown in Figure 3-4 demonstrate that the simulation progressed as planned. The interpretation of these results must then be drawn from the final energies of the

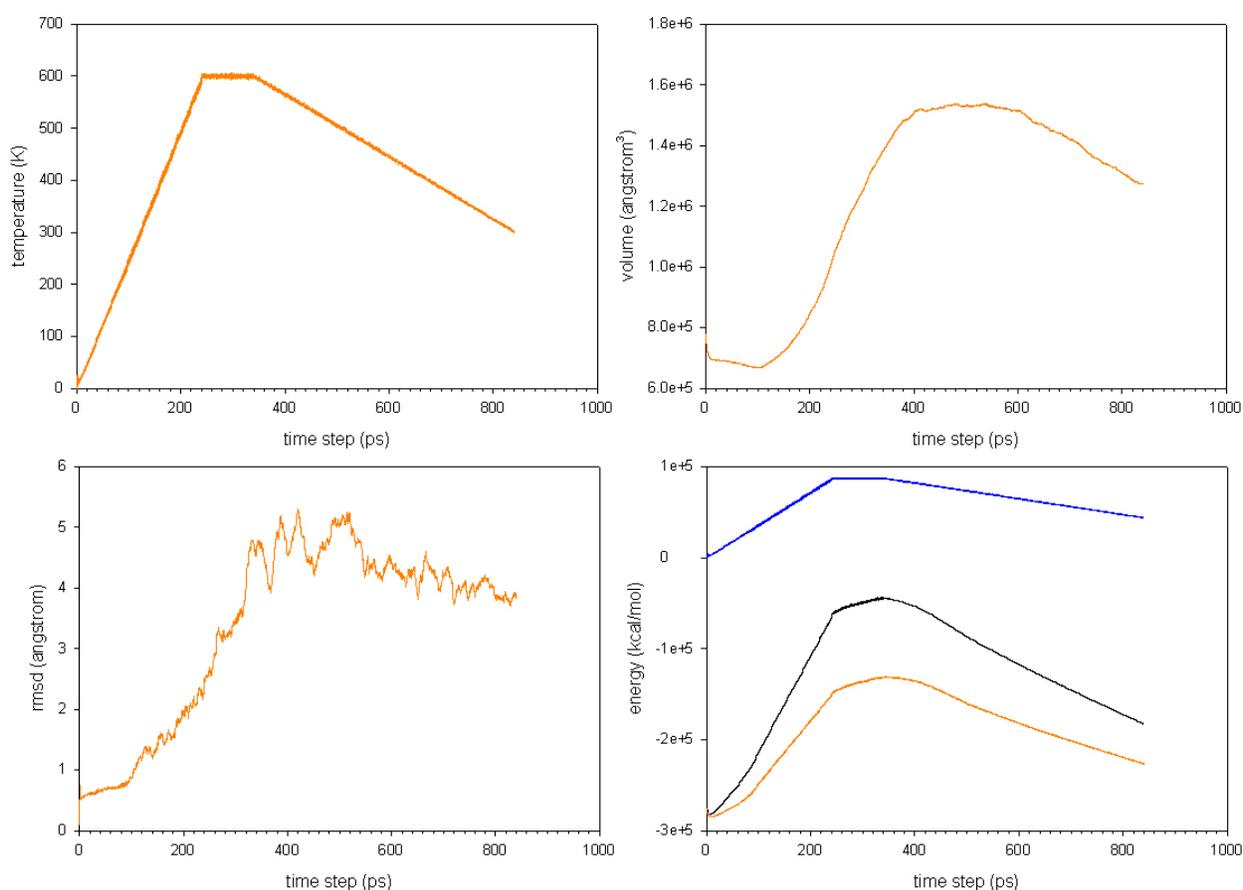


Figure 3-4. AspATP SA Run 1. Going clockwise from the top left, the first plot is a plot of the temperature of the system vs the time step. The second plot depicts the volume of the system vs the time step. The third plot shows the energy of the system vs time step. In this graph, the blue curve represents the kinetic energy of the system, while the orange curve represents the potential energy of the system, and finally the black curve follows the total energy (which is a sum of the kinetic and potential energies). The last plot shows the rmsd of the backbone atoms in the system as the simulation progresses, as compared with the initial structure vs the time step.

system as the annealing progressed and from close visual inspection of the system. Table 3-1 shows the energies of the AspATP system over the course of the annealing. The energies are those of the system at each specific temperature after the structures were minimized, effectively lowering each structure to 0 K and trapping them in the closest minima.

During the constant run at 600 K for 100 ps of SA Run 1, the free aspartate migrated completely outside of the synthetase active site, and was never able to return to the active site during the cooling process. Although these structures have the lowest overall energies for the AspATP system across all SA Runs, they cannot be regarded as valuable structures when aspartate is absent from the active site. Several constraints were tried in an attempt to pull the aspartate back into the active site; however, the constraints were elevating the energy of the system to the point of crashing the calculation, rendering the constraints not viable.

Simulated annealing run 2

In the second simulated annealing protocol, the heated AspATP model system was cooled immediately, rather than equilibrating at 600 K. The cooling process was the same as previously described, as 10 structures were collected during the cooling. Once the system reached 300 K, it was then equilibrated and the final structure of the equilibration was minimized by the same procedure as the 10 structures collected from the cooling. Figure 3-5 shows the plots made from following the systems through this specific SA procedure. One distinguishable difference in the system from SA Run 1 to SA Run 2 is the total volume of the system. In this protocol, the volume did not expand as much as in the first run and the volume of the system ended up in the same range as where it began, rather than ending at a larger total volume.

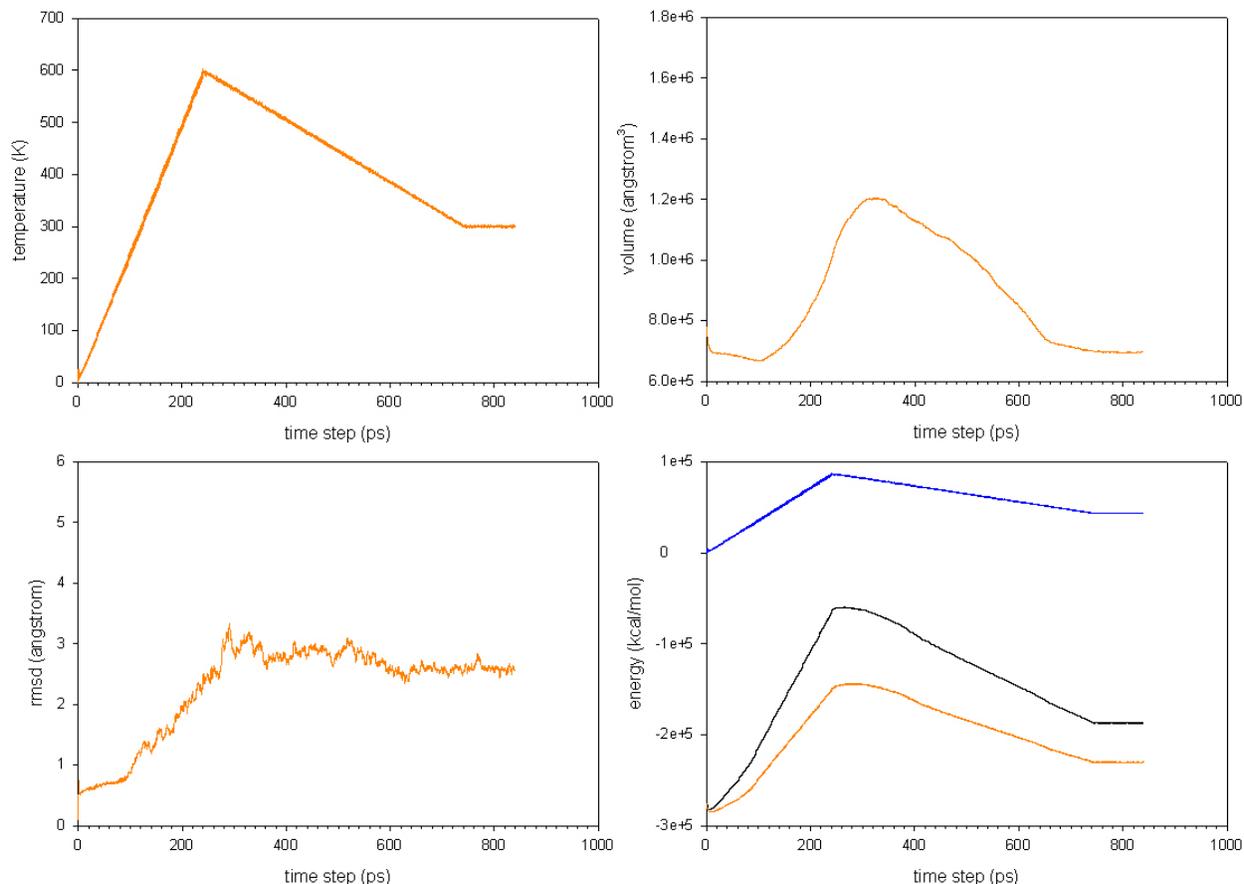


Figure 3-5. AspATP SA Run 2. Going clockwise from the top left, the first plot is a plot of the temperature of the system vs the time step. The second plot depicts the volume of the system vs the time step. The third plot shows the energy of the system vs time step. In this graph, the blue curve represents the kinetic energy of the system, while the orange curve represents the potential energy of the system, and finally the black curve follows the total energy (which is a sum of the kinetic and potential energies). The last plot shows the rmsd of the backbone atoms in the system as the simulation progresses, as compared with the initial structure vs the time step.

While all of the structures were examined in comparison to the initial structure, the two which are the focus of this analysis are the lowest total energy structure (the minimization of the structure at 390 K) as seen in Table 3-1, and the final structure that was the result of a minimization after an equilibration period of 100 ps at 300 K. Both of these structures have low rmsd values when compared to the initial structure. PyMOL estimated the rmsd of the minimized structure at 390 K to be 1.651 Å in comparison to the initial structure, while the equilibrated

structure had an rmsd of 1.578 Å. Figure 3-6 shows the structural alignments of both structures with the initial AspATP system. In both cases, the overall structure of the proteins looks very similar to the original system. The major deviations occur in the positions of the adenine group of the ATP and free aspartate. To examine the differences, the active sites were inspected in more detail.

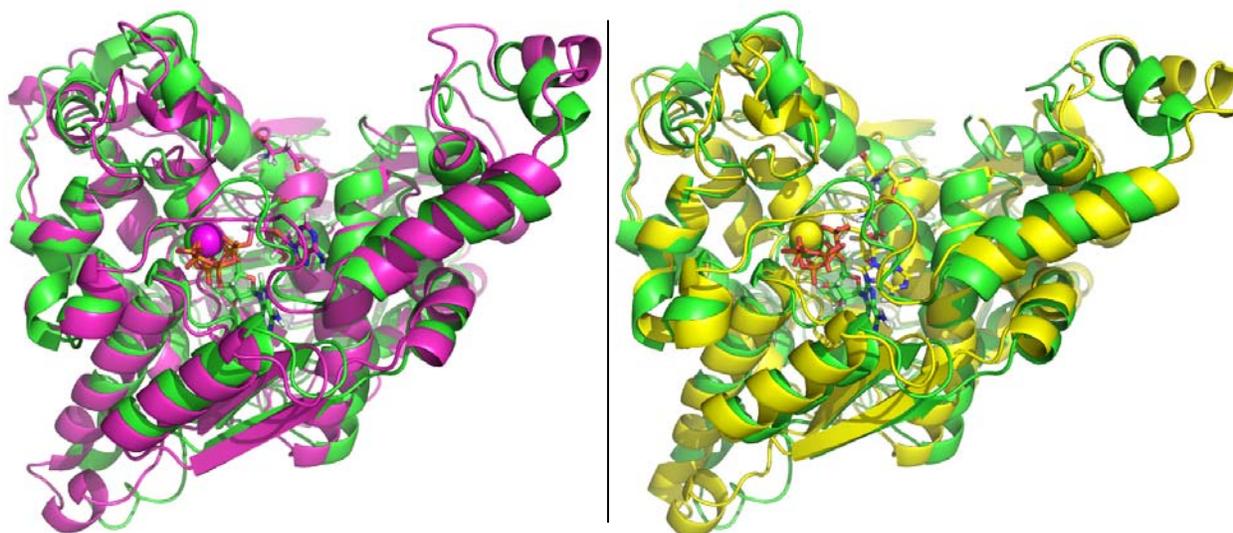


Figure 3-6. Structural alignments of the minimized structure at 390 K and the structure equilibrated at 300 K with the initial structure. On the left, the original structure is shown in green and the minimized structure at 390 K is shown in magenta, while on the right, the minimized structure of the equilibration run at 300 K is shown in yellow where the original structure remains in green.

Figure 3-7 shows a detailed view of the ATP binding site for all three structures. In a comparison of the three structures, the pyrophosphatase loop (shown with residues Ser-234, Lue-237, Asp-238 and Ser-239) has almost no change. This holds true for all of the residues surrounding the phosphate tail of ATP. The only large difference was in the position of the adenine moiety of ATP and the residues surrounding it. It rotates about the α -phosphate group in each structure, shifting the residues around it as it moves. There is also one interesting side chain shift from the initial structure as well. In both the 390 K structure and the equilibrium structure,

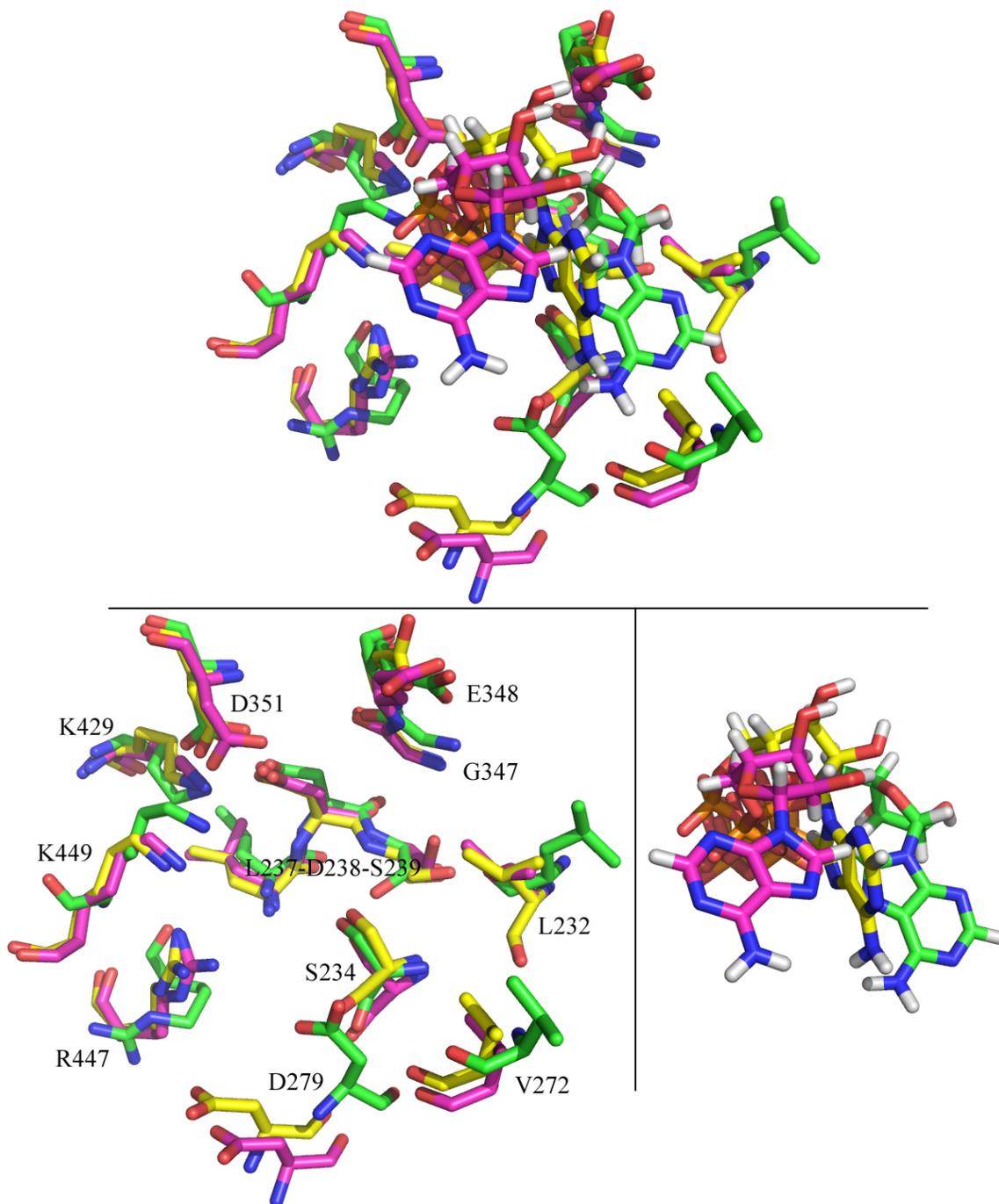


Figure 3-7. A look at ATP binding interactions in SA Run 2. The initial structure is shown with green carbons, while the 390 K minimized structure shown with magenta carbons and the minimized structure of the equilibration at 300 K is shown with yellow carbons. The top view is the complete AS-B/ATP view that is then broken down into the residues only on the bottom left and the ATP only on the bottom right. The non-carbon atoms are colored as follows: O = red, N = blue, H = white, P = orange, and Mg^{2+} = color of model carbon.

the side chain of residue Arg-447 has shifted to make a hydrogen bond with the phosphate tail of ATP.

A closer look at the free aspartate and the residues that interact with it reveals a few interesting changes. Figure 3-8 shows shifts in residues and free aspartate that occurred in each

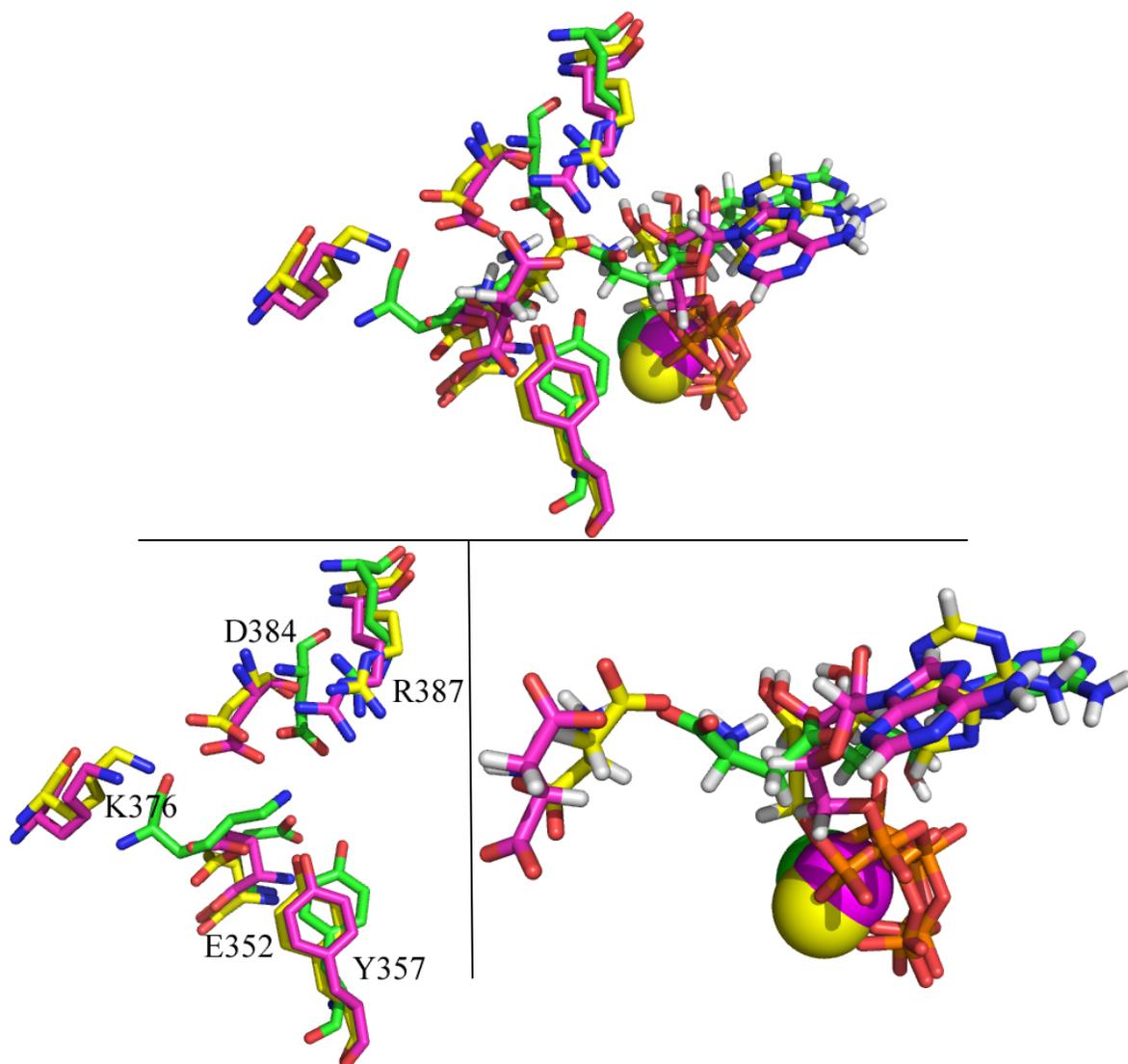


Figure 3-8. A look at free aspartate binding interactions in SA Run 2. The initial structure is shown with green carbons, while the 390 K minimized structure shown with magenta carbons and the minimized structure of the equilibration at 300 K is shown yellow carbons. The top view is the complete AS-B/Asp/ATP/Mg²⁺ view that is then broken down into the residues only on the bottom left and the Asp/ATP/Mg²⁺ only on the bottom right. The non-carbon atoms are colored as follows: O = red, N = blue, H = white, P = orange, and Mg²⁺ = color of model carbon.

model. The major changes in residues are seen in a side chain flip of Glu-352 and a large shift in Lys-376 as the adenine moiety of ATP shifts, pushing the aspartate and the Lys-376 back. In both the 390 K minimized structure and the equilibrated structure, the free aspartate is in a seemingly worse position for attack of ATP and formation of the intermediate. This seems to be more evidence that something is missing in this model. Whether the important factor omitted is a second Mg^{2+} ion or the remaining 37 residues cannot be determined from these results.

Simulated annealing run 3

In the third type of simulated annealing run, after the system was heated to 600 K, the temperature was then allowed to distribute evenly in the system under a reduced but still constant pressure of only 0.025 atm. This was done to allow the volume to expand even more. Increasing the volume of the system has the effect of lowering the density, and in principle should allow the protein more freedom to explore even more conformational space while at 600 K. By then cooling under higher pressure, the volume is forced down and can help the protein more readily adopt the lowest energy conformer.¹²⁴ Figure 3-9 shows the progression of this SA run. At the end of 340 ps, this system had reached a larger volume than either of the previous SA runs. The first 250 ps of cooling were performed under a constant pressure of 1000 atm. This led to an instant and steep decrease in the volume of the system. The pressure was then reduced to 500 atm for the next 100 ps of cooling. The transition from 1000 atm to 500 atm is where a brief increase in the volume of the system is seen near the 600 ps mark. The pressure was then reduced again to 10 atm for 50 ps, where another transition can be seen on the volume graph around 700 ps mark, before running the final 100 ps of cooling under a constant pressure of 1 atm. The initial system started with a volume of $778,688 \text{ \AA}^3$ and reached a maximum volume of approximately $1,380,471 \text{ \AA}^3$ after 340 ps of total simulation time before ending with a total system volume of approximately $700,605 \text{ \AA}^3$. Structures were collected at the same 10 points as

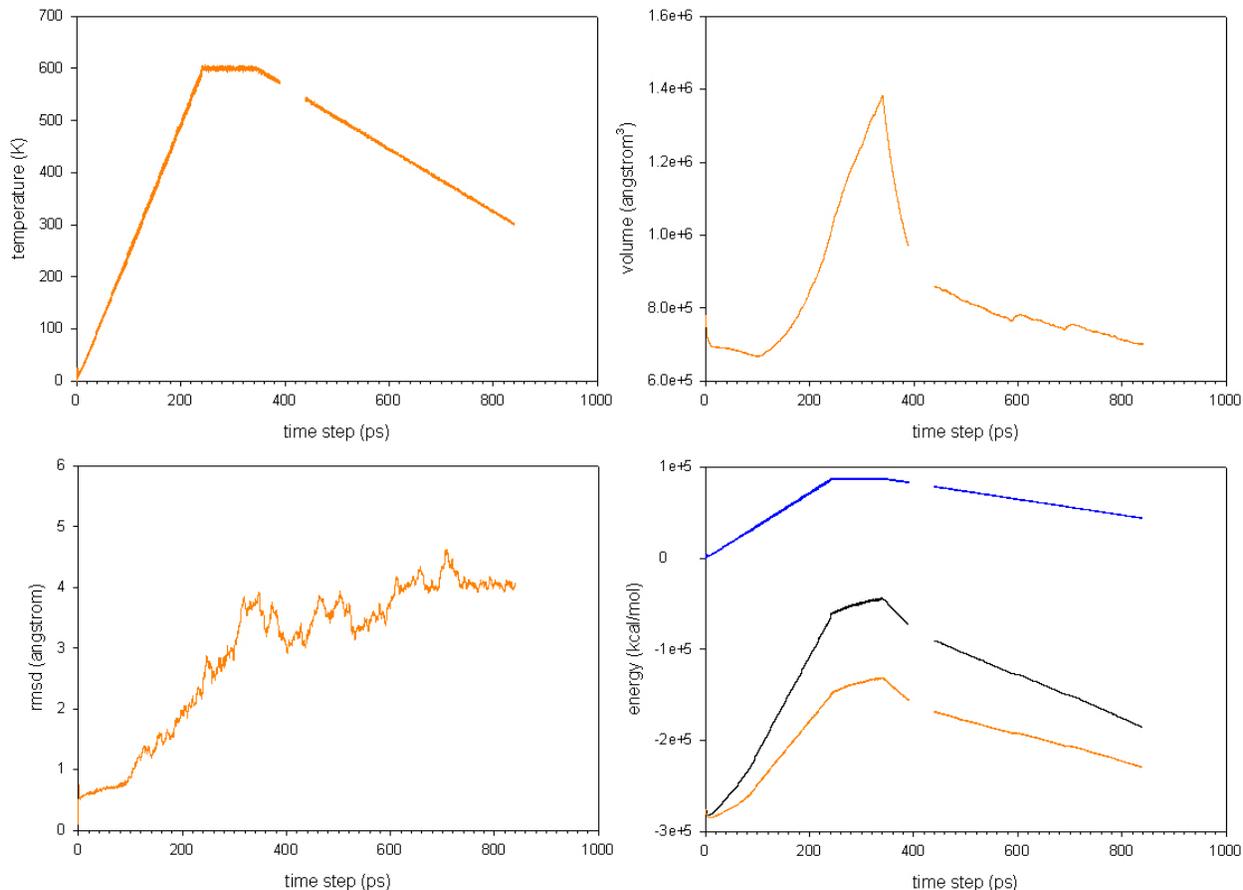


Figure 3-9. AspATP SA Run 3. Going clockwise from the top left, the first plot is a plot of the temperature of the system vs the time step. The second plot depicts the volume of the system vs the time step. The third plot shows the energy of the system vs time step. In this graph, the blue curve represents the kinetic energy of the system, while the orange curve represents the potential energy of the system, and finally the black curve follows the total energy (which is a sum of the kinetic and potential energies). The last plot shows the rmsd of the backbone atoms in the system as the simulation progresses, as compared with the initial structure vs the time step. The small gap in the temperature, volume and energy graphs is because that particular output file was lost prior to extracting the necessary information.

in the previous two SA runs and the energies of the corresponding minimized structures can be seen in Table 3-1.

Much like what occurred in SA Run 1 for this system, allowing the temperature to distribute at 600 K opened the synthetase active site and with no apparent interactions strong enough to hold the free aspartate in the active site; it migrated out. Figure 3-10 shows the

structural alignment of the lowest energy structure from Table 3-1 for SA Run 3. This structure was the minimized structure acquired at 300 K. PyMOL was used to calculate the rmsd for the minimized structure at 300 K to be 1.606 Å in comparison to the initial structure. The only major change in the overall structure is the C-terminus, where the protein structure moved out away from the synthetase site creating a large opening. The free aspartate that has completely left the active site can be seen above the opening in the active site, in the upper middle of Figure 3-10.

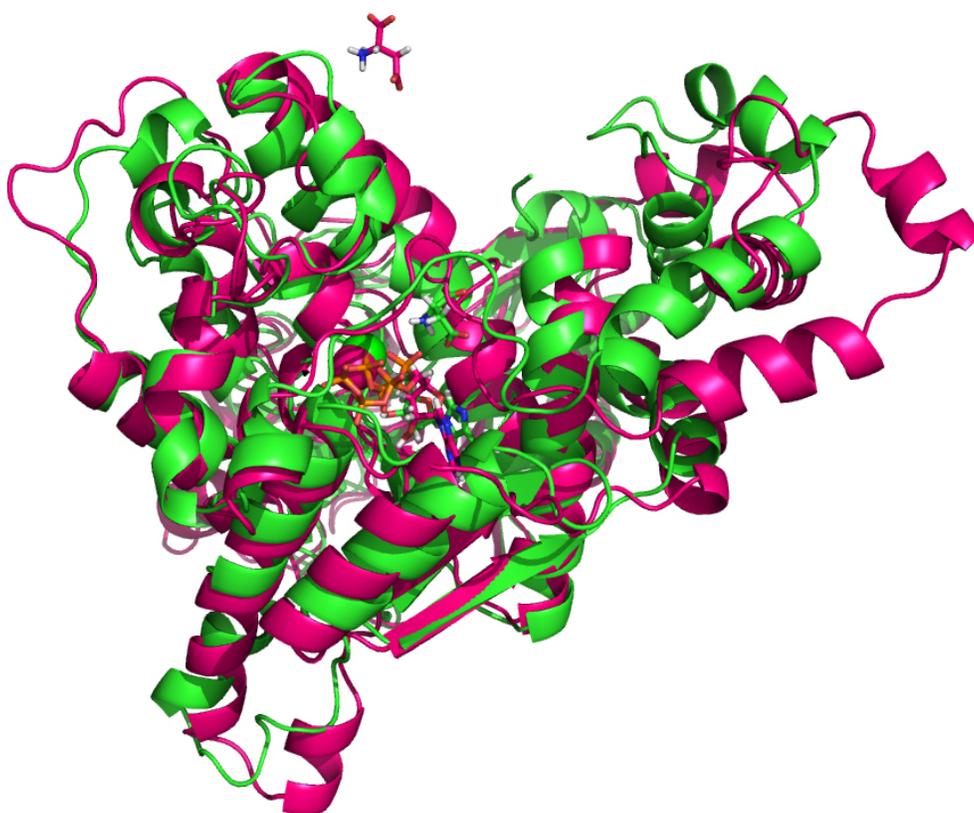


Figure 3-10. Structural comparison of the initial structure of the AspATP model with the minimized structure at 300 K. The initial structure is in green, while the minimized structure at 300 K is in hot pink. The synthetase substrates, aspartate, ATP and Mg²⁺, are shown in stick or sphere (Mg²⁺) representation with the carbons colored to match their respective system.

The sum of these three SA runs on the AspATP model system suggests that there must be some piece of this model missing that would keep the free aspartate in the synthetase active site

and in position for attack on ATP. As suggested earlier, the most plausible answer would be a second Mg²⁺ in this active site; however, the final 37 residues that are absent could also play a role in constraining aspartate to the active site.

βAspAMP Model

The three simulated annealing runs performed on the βAspAMP model system began from a single minimized initial structure and a single heating run to 600 K over 240 ps, just as was the case for the AspATP model system. As was detailed more thoroughly in the Methods section, the three runs then varied in their equilibration and cooling protocols, before all of the resulting structures were minimized in a consistent fashion as the final step. The resulting energies of all three complete SA protocols can be seen in Table 3-2.

Table 3-2. Final energies for the SA Runs for the βAspAMP model system

	Energy (kcal/mol)		
	SA Run 1	SA Run 2	SA Run 3
Initial		-290,457.77	
600 K ^a		-69,076.12	
100 ps run at 600 K ^a	-46,136.79	-292,764.37 ^b	-46,115.30
570 K	-290,097.90	-290,207.23	-289,884.32
540 K	-290,189.82	-290,240.85	-290,375.42
510 K	-292,308.88	-290,488.33	-290,555.66
480 K	-293,666.30	-292,179.31	-291,091.17
450 K	-294,375.28	-293,035.19	-291,478.55
420 K	-294,859.69	-293,509.66	-291,637.92
390 K	-294,925.12	-293,376.03	-292,007.39
360 K	-295,462.21	-292,423.96	-291,941.47
330 K	-295,938.30	-292,226.61	-292,373.55
300 K	-295,918.78	-292,501.75	-292,762.35

^a Structures were not minimized; ^b This energy value was for the minimized final structure for the equilibration at 300 K, as no equilibration was done at 600 K in SA Run 2.

A closer examination of the initial structure of AS-B with the βAspAMP intermediate and PPI bound in the synthetase active site is shown in Figure 3-11. Because the βAspAMP model

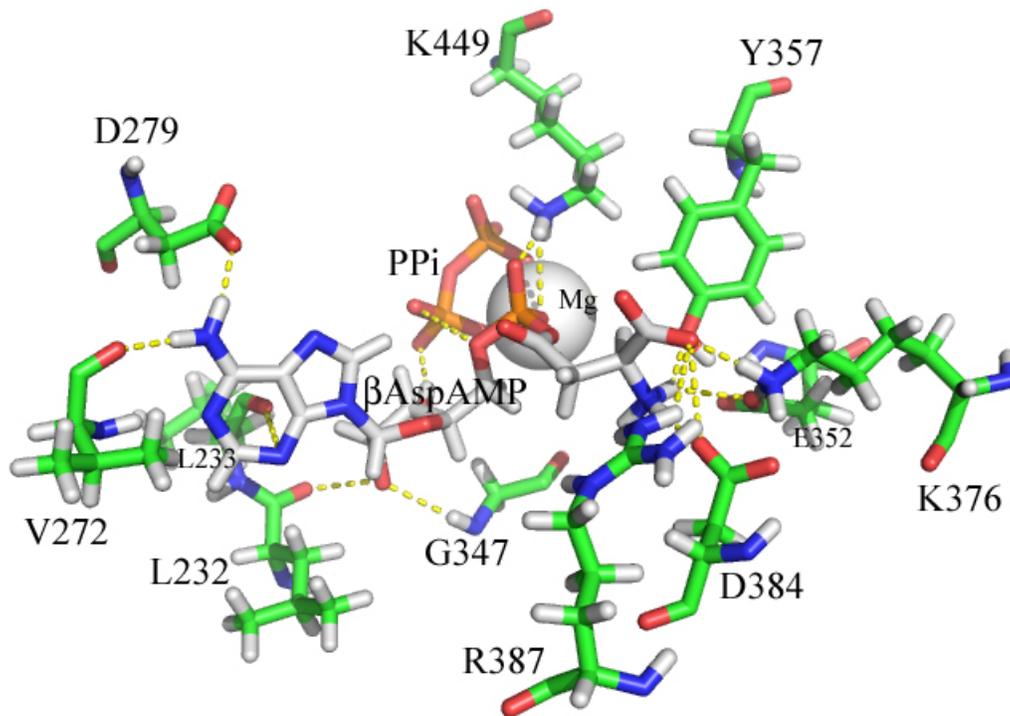


Figure 3-11. Synthetase active site interaction in the initial β AspAMP model system. The AS-B is shown with the carbons colored green while the β AspAMP intermediate is depicted with white carbons. The potential interactions are shown with the dashed yellow lines. The Mg^{2+} is also shown as a white sphere. The other non-carbon atoms are colored as follows: O = red, N = blue, H = white, and P = orange.

system was created from the AspATP model, most of the same interactions are seen here as in the previous system. A structural alignment was then performed in PyMOL to compare the initial structure with the structure after it had been heated to 600 K. PyMOL found the rmsd of the heated structure of 1.560 Å in comparison to the initial structure. The structural alignment can be seen in Figure 3-12. As can be seen in the overlay and with the rmsd, there is not a great deal of variance between the heated and initial structure. This is in sharp contrast to the AspATP system that experienced a much larger structural change with the heating. In the β AspAMP model, the β AspAMP intermediate has enough interactions with the surrounding protein structure, shown in Figure 3-11, to stay close to the position it started in. The PP_i , on the other hand, is held in place

because it is buried behind the β AspAMP intermediate inside of the synthetase active site, as well as being locked down by interactions with the Mg^{2+} .

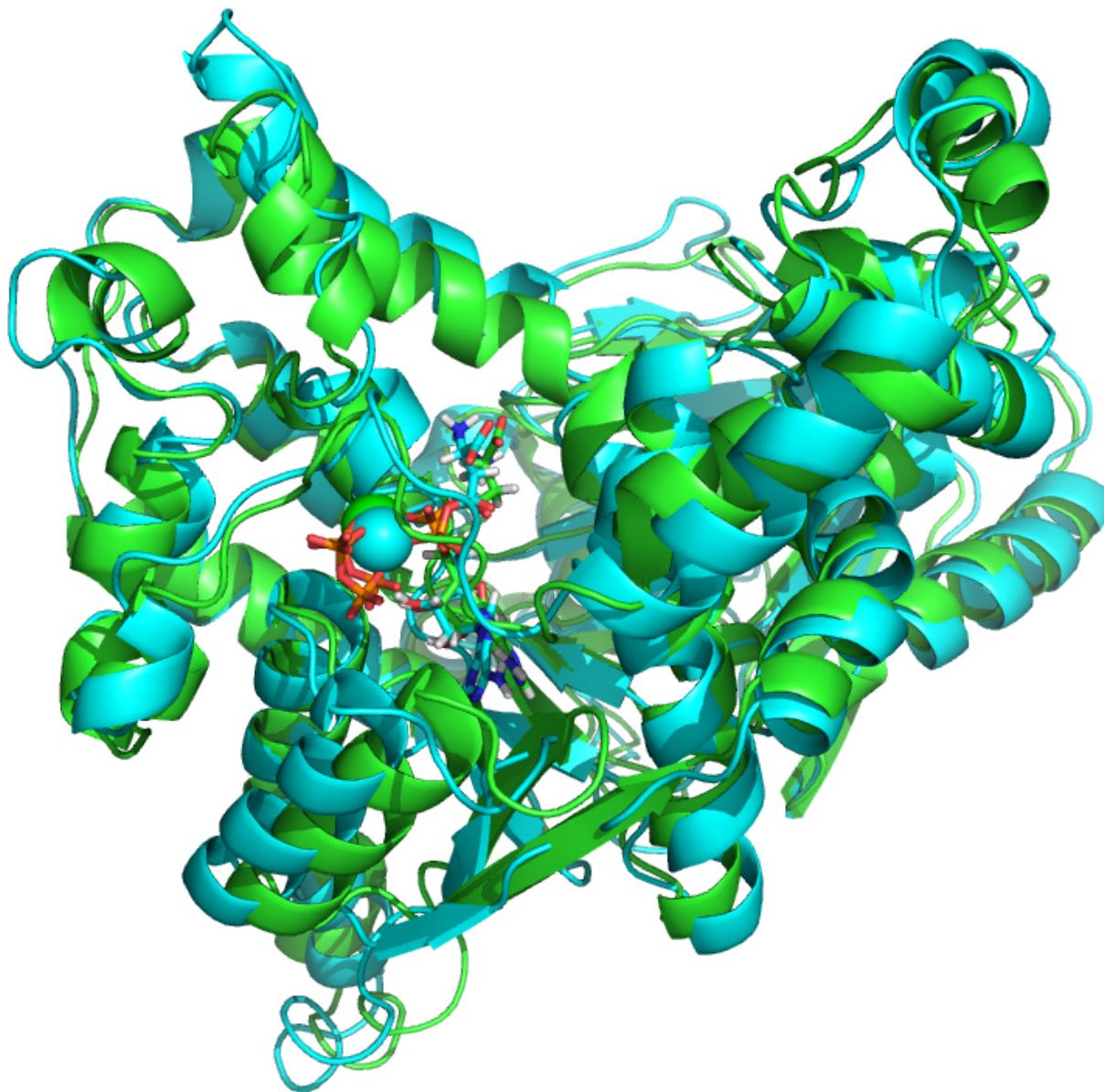


Figure 3-12. Structural comparison of the initial structure of the β AspAMP model with the same structure heated to 600 K. The initial structure is in green, while the heated structure is in cyan. The synthetase substrates, β AspAMP, PP_i and Mg^{2+} , are shown in stick or sphere (Mg^{2+}) representation with the carbons colored to match their respective system.

Simulated annealing run 1

The first simulated annealing run for the β AspAMP model system followed the same protocol that was previously described, and for the first run for the AspATP model system. The system was first heated to 600 K over 240 ps at a constant pressure of 1 atm. After reaching 600 K, the temperature was allowed to distribute evenly in the system for 100 ps at a constant pressure of 1 atm. At that point, the system was cooled down to 300 K over the course of 500 ps

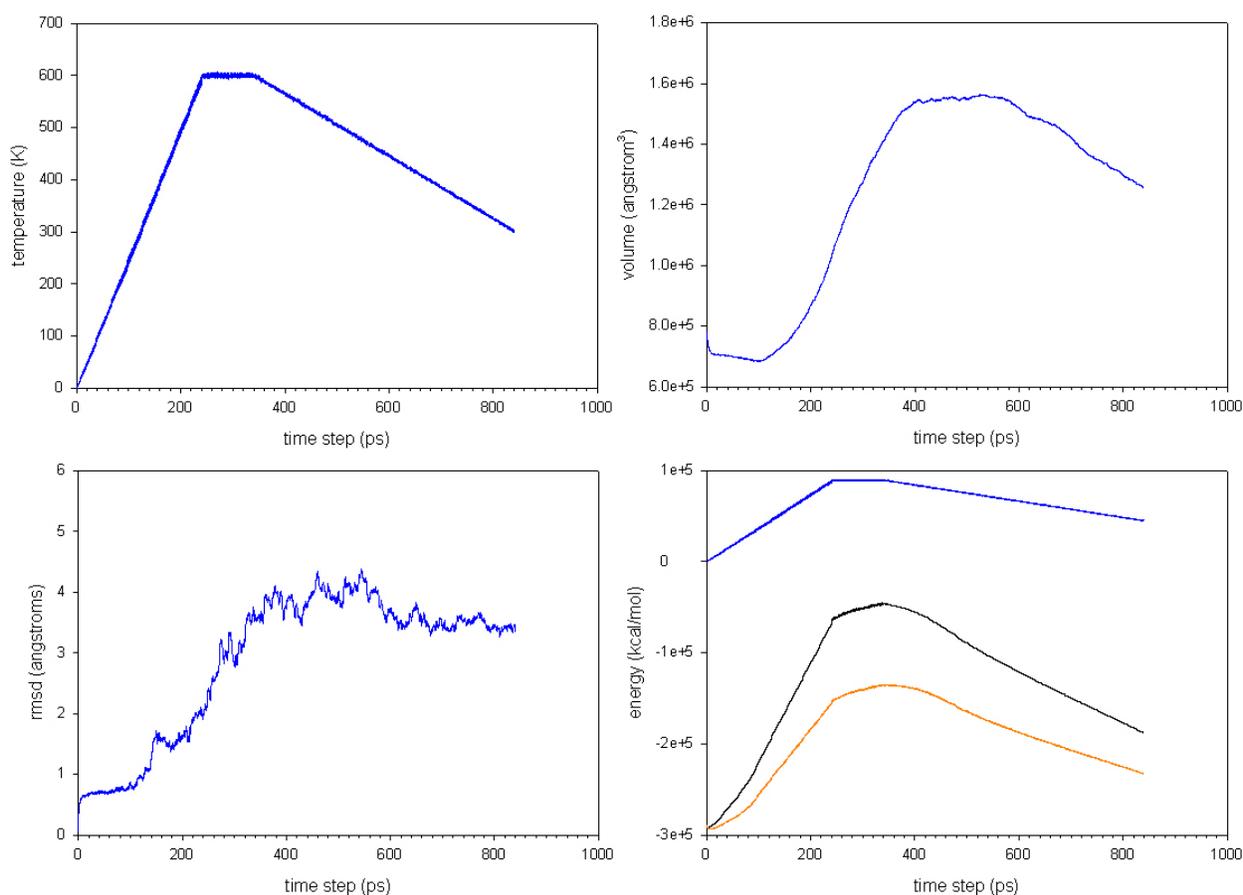


Figure 3-13. β AspAMP SA Run 1. Going clockwise from the top left, the first plot is a plot of the temperature of the system vs the time step. The second plot depicts the volume of the system vs the time step. The third plot shows the energy of the system vs time step. In this graph, the blue curve represents the kinetic energy of the system, while the orange curve represents the potential energy of the system, and finally the black curve follows the total energy (which is a sum of the kinetic and potential energies). The last plot shows the rmsd of the backbone atoms in the system as the simulation progresses, as compared with the initial structure vs the time step.

with structures being collected every 50 ps. The collected structures were then minimized and the final energies of minimizations can be seen in Table 3-2. Figure 3-13 shows the graphs of the temperature, volume, backbone rmsd and energy of the system as SA Run 1 progressed.

The lowest energy structure was for the minimized structure at 300 K (Table 3-2). A structural alignment of this structure with the initial β AspAMP structure in PyMOL gave an rmsd of 1.887 Å. This was the lowest rmsd found between the initial structure and any of the minimized structures. Figure 3-14 shows the active site comparison between the initial β AspAMP structure and the minimized structure at 300 K. As can be seen from this figure, there are a few significant differences. The β AspAMP intermediate undergoes a significant shift of the aspartyl tail. This shift seems to result in the shift of residues on the right side of the active site. Tyr-357 appears to have shifted out of the way of the intermediate tail, while the Arg-387, Asp-384 and Lys-376 have shifted slightly because they are no longer within hydrogen bonding distance of the aspartyl tail. This shift in the tail also significantly shifts Lys-449 that is located at the top of the active site in Figure 3-14. As in the AspATP model system, the side chain of Arg-447 shifts to interact with the PP_i that is present in this model.

Perhaps the most striking change is in residue Glu-348 that is seen at the center of the active site (Figure 3-14). This Glu-348 has been extensively tested experimentally as part of the dissertation work by former group member, Dr. Jemy Gutierrez.¹⁴⁷ This residue is located at the mouth of the intramolecular tunnel that connects the glutaminase active site with the synthetase active site and is believed to be involved in the formation of the β AspAMP intermediate or in active site communication. It is also possible that this residue acts as a gate for the passage of the NH_3 formed in the glutaminase active site. In the annealed model, at all stages, the side chain of Glu-348 has flipped. As can be seen in Figure 3-15, the side chain flip of Glu-348 seems to have

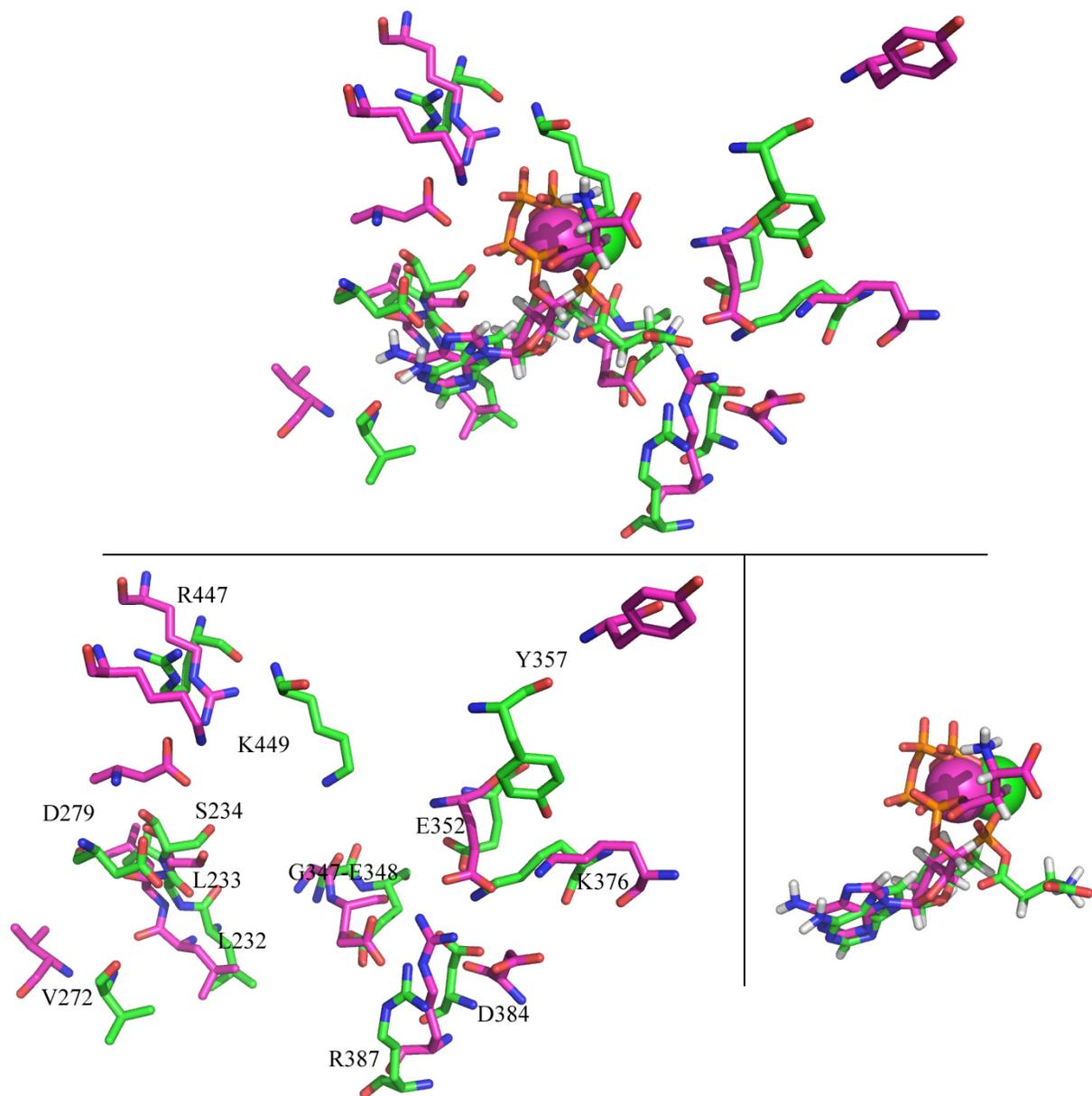


Figure 3-14. Active site comparison between the initial β AspAMP model and the model minimized at 300 K. The initial structure is shown with green carbons, while the minimized structure at 300 K is shown with magenta carbons. The top view is of the active site with the β AspAMP intermediate, PP_i and Mg^{2+} bound. The bottom left is the same view with the substrates removed for clarity and the bottom right is the only the substrates. The non-carbon atoms are colored as follows: O = red, N = blue, H = white, P = orange, Mg^{2+} = model carbon color.

positioned the negatively charges carboxylate tail away from the opening of the tunnel in the synthetase active site. Judging from the entire range of structures collected, the shift of the side

chain of Glu-348 is the most likely cause of the shift in the β AspAMP intermediate tail shift that forces the other residues to shift as well. This sequence is suggested because in the series of models, the β AspAMP intermediate tail gradually moves away from Glu-348, which has flipped in every instance. Looking at this change in the remaining two series of SA runs should confirm this change in the β AspAMP model.

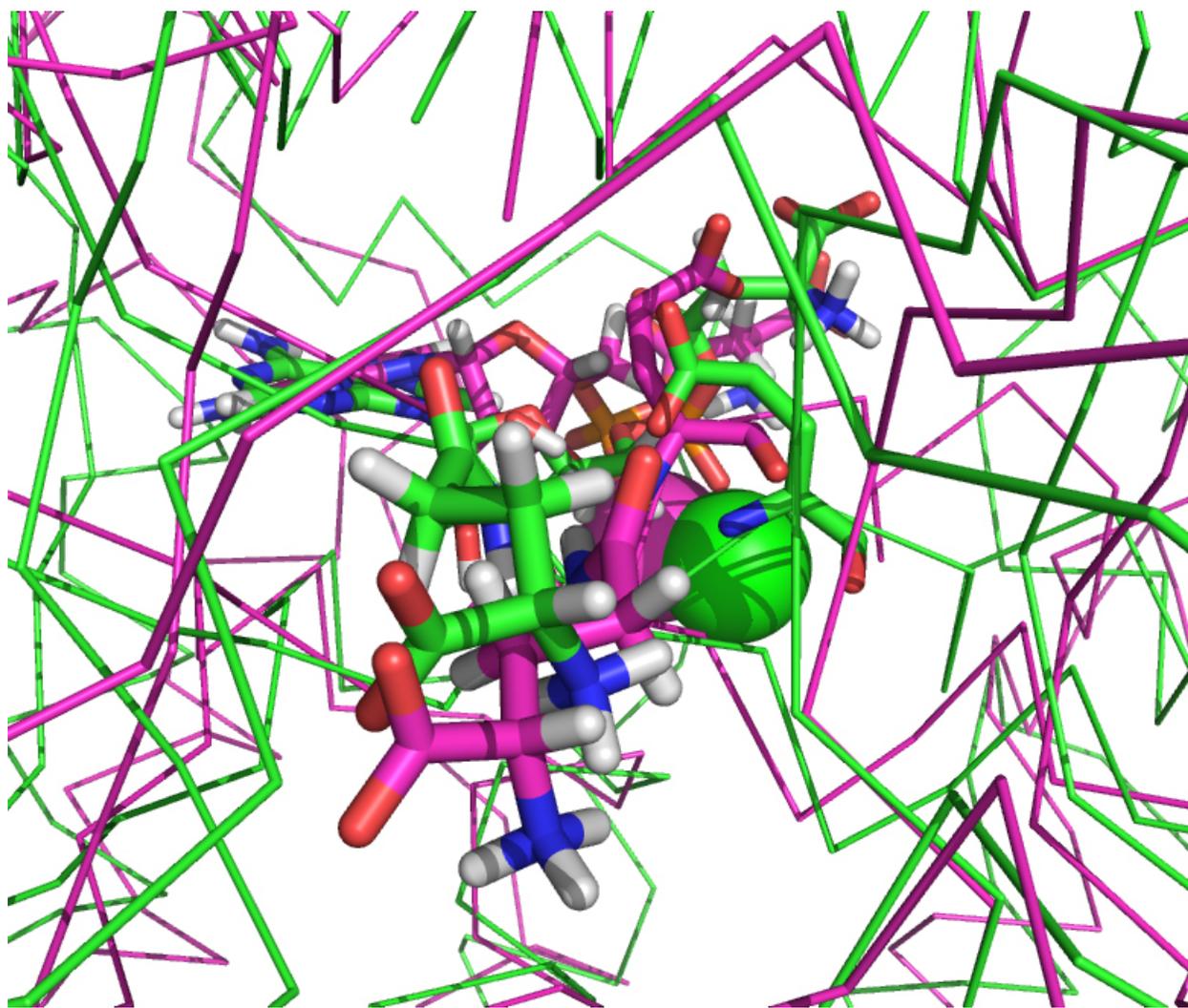


Figure 3-15. The shift of Glu-348. This view is looking down the intramolecular tunnel that connects the glutaminase site with glutamine bound (near) and the synthetase active site with the β AspAMP intermediate bound (far) in AS-B. The original structure of the model is seen in green while the annealed model is shown in magenta. The Glu-348 side chain flips from right to left, away from the opening of the tunnel.

Simulated annealing run 2

In the second simulated annealing run performed on the β AspAMP model, the cooling stage was begun immediately after the system reached 600 K. This was the same process done for the SA Run 2 for the AspATP model. The system was then equilibrated at 300 K and all of

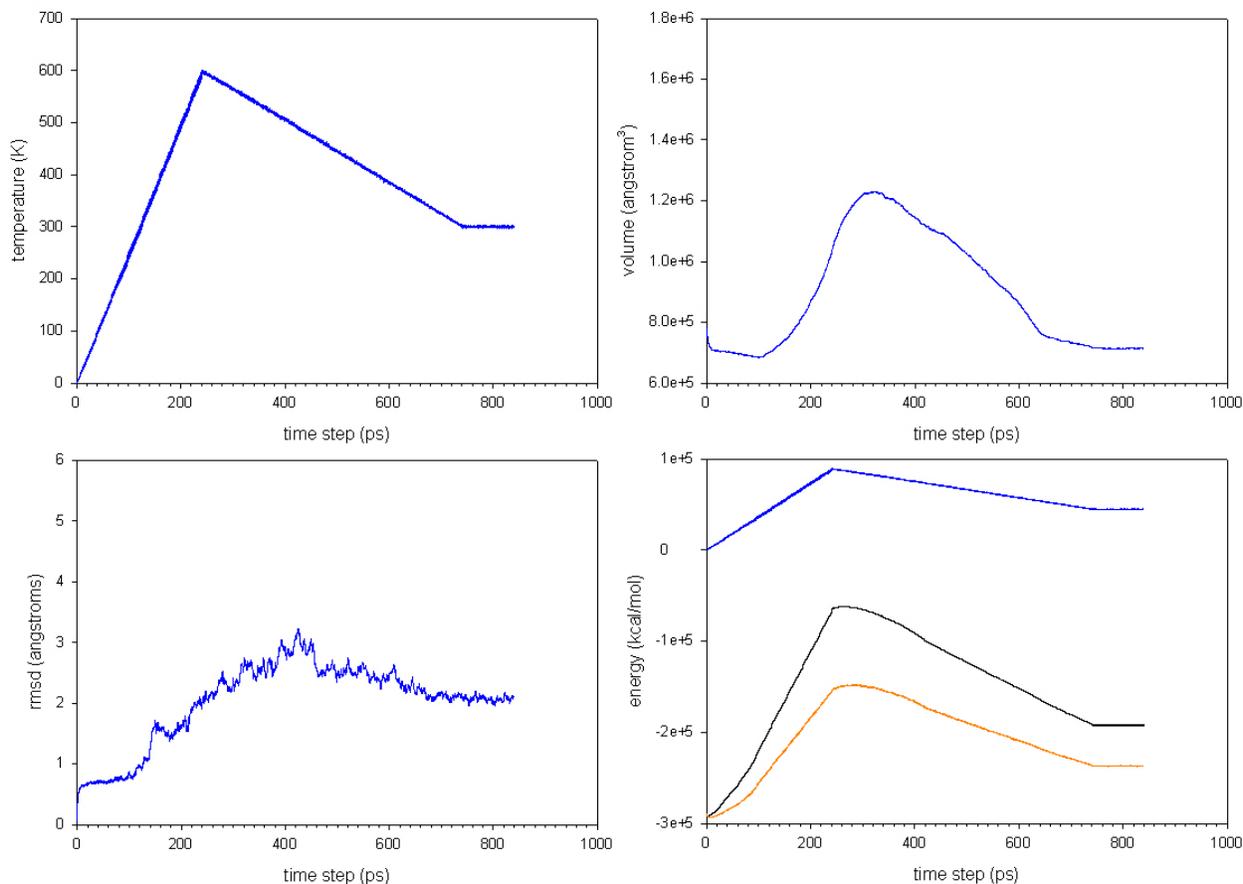


Figure 3-16. β AspAMP SA Run 2. Going clockwise from the top left, the first plot is a plot of the temperature of the system vs the time step. The second plot depicts the volume of the system vs the time step. The third plot shows the energy of the system vs time step. In this graph, the blue curve represents the kinetic energy of the system, while the orange curve represents the potential energy of the system, and finally the black curve follows the total energy (which is a sum of the kinetic and potential energies). The last plot shows the rmsd of the backbone atoms in the system as the simulation progresses, as compared with the initial structure vs the time step.

the structures were minimized. The progression of the SA run can be seen in the graphs in Figure 3-16. The system was held at a constant pressure of 1 atm for the entire SA process. The energies of the minimized structures can be seen in Table 3-2.

All of the minimized structures were then compared to the initial structure for the β AspAMP model in PyMOL. The minimized structure taken at 420 K and the minimized results of equilibration at 300 K are the key structures for SA Run 2. These were chosen because the minimized structure at 420 K was found to be the lowest overall energy structure for SA Run 2 (Table 3-2), while the structure that was minimized from the equilibration at 300 K was the final structure collected in this process. PyMOL was used to calculate an rmsd for the minimized structure at 420 K to be 2.006 Å in comparison to the initial structure, while the minimized structure from the equilibration run was found to have an rmsd of 1.609 Å. A comparison of the synthetase active sites (Figure 3-17) shows much less variation than was seen in the model from SA Run 1.

Both of the structures compared against the initial structure for SA Run2 have less variance in the active site in contrast to the results from SA Run1 for the same model system. In Figure 3-17, the residues that showed the greatest deviation in SA Run 1 (Lys-449, Tyr-357 and Asp-279) are all closer to their initial positions in SA Run 2. Even with what appears to be better structural agreement, Arg-447 and Glu-348 still experienced the same side chain flipping as seen in SA Run 1. In the case of the model minimized after equilibration at 300 K, the carboxylate of the β AspAMP intermediate is within hydrogen bonding distance of Glu-348. The β AspAMP from the minimized structure taken at 420 K is similar to the result of SA Run 1, with the aspartyl tail shifted away from Glu-348. This aspartyl tail shift is the most significant difference between the two models examined in Figure 3-17. The remaining variations from the

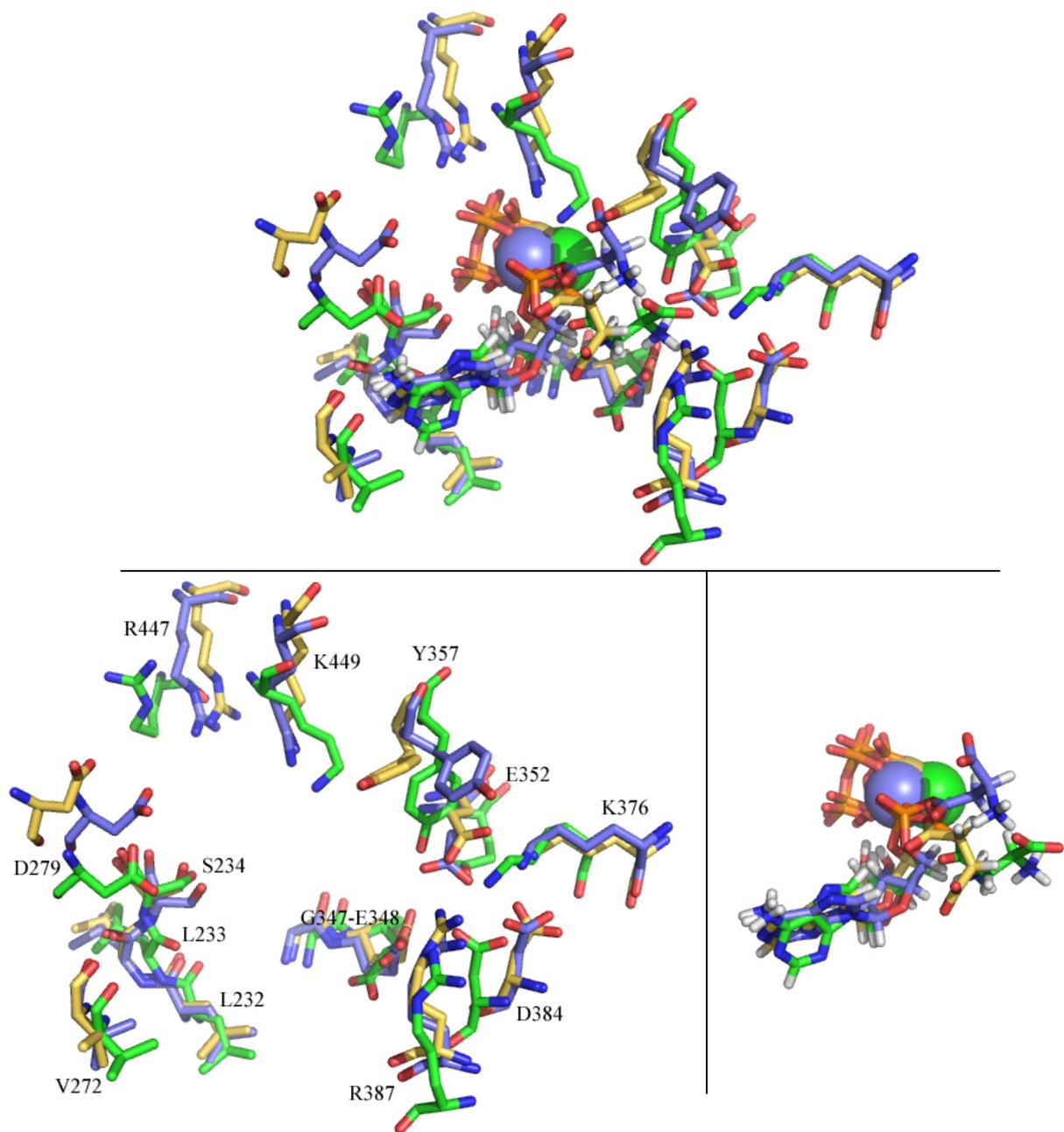


Figure 3-17. Active site comparison between the initial β AspAMP model and the model minimized at 420 K and the model minimized after equilibration at 300 K. The initial structure is shown with green carbons, while the minimized structure at 420 K is shown with slate carbons and the minimized structure after equilibration is shown with yellow-orange carbons. The top view is of the active site with the β AspAMP intermediate, PP_i and Mg^{2+} bound. The bottom left is the same view with the substrates removed for clarity and the bottom right is the only the substrates. The non-carbon atoms are colored as follows: O = red, N = blue, H = white, P = orange, Mg^{2+} = model carbon color.

initial model are consistent between the two models produced in the annealing process, as well as the other structures produced, but not shown here. These results provide evidence that residue Glu-348 undergoes a side chain flip that is independent of β AspAMP position.

Simulated annealing run 3

In the third simulated annealing run, as was done for the AspATP model, after the system was heated to 600 K, the temperature was then allowed to distribute evenly in the system under a reduced but still constant pressure of 0.025 atm. The system was then cooled under high pressure until it was at 300 K. The graphs in Figure 3-18 follow the progression of this SA run. At the end of 340 ps, this system reached a larger volume than either of the previous SA runs on the β AspAMP model. The first 300 ps of cooling were performed under a constant pressure of 1000 atm. This led to an instant and steep decrease in the volume of the system. The pressure was then reduced to 500 atm for the next 100 ps of cooling. The transition from 1000 atm to 500 atm is where a brief increase in the volume of the system is seen near the 650 ps mark. The pressure was then reduced to 1 atm for the final 100 ps, where another transition can be seen on the volume graph around 750 ps. The initial system started with a volume of $778,688 \text{ \AA}^3$ and reached a maximum volume of approximately $1,415,147 \text{ \AA}^3$ after 340 ps of total simulation time before ending with a total system volume of approximately $716,928 \text{ \AA}^3$. Structures were collected at the same 10 points as in the previous two SA runs and the energies of the corresponding minimized structures can be seen in Table 3-2.

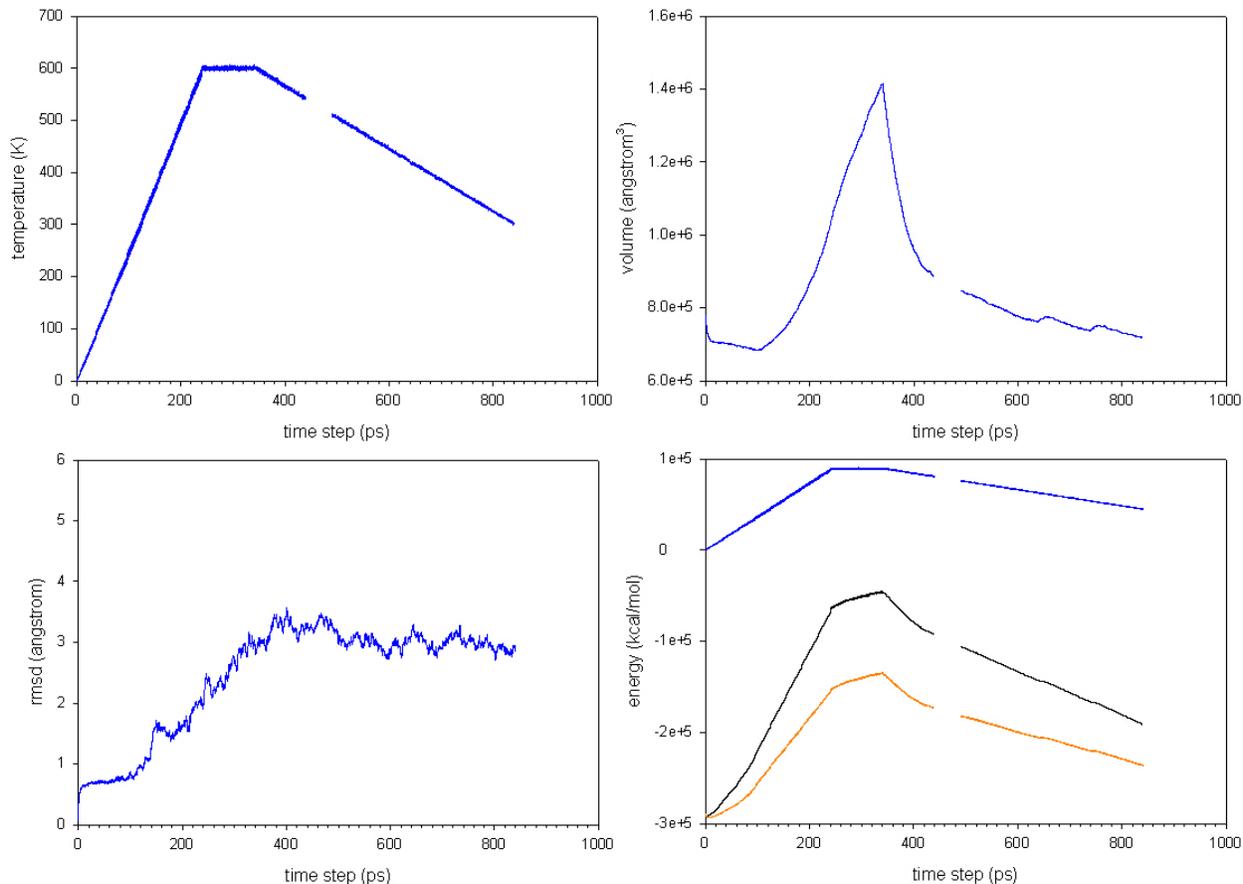


Figure 3-18. β AspAMP SA Run 3. Going clockwise from the top left, the first plot is a plot of the temperature of the system vs the time step. The second plot depicts the volume of the system vs the time step. The third plot shows the energy of the system vs time step. In this graph, the blue curve represents the kinetic energy of the system, while the orange curve represents the potential energy of the system, and finally the black curve follows the total energy (which is a sum of the kinetic and potential energies). The last plot shows the rmsd of the backbone atoms in the system as the simulation progresses, as compared with the initial structure vs the time step. The small gap in the temperature, volume and energy graphs is because that particular output file was lost prior to extracting the necessary information until it was at 300 K.

The lowest energy structure found in this annealing process was for the minimized structure at 300 K (Table 3-2). PyMOL was again used to calculate an rmsd of 2.379 Å for this structure in comparison to the initial structure. A closer look at the active site comparison between the minimized structure at 300 K and the initial structure (Figure 3-19) reveals more similarities to the model produced in SA Run 1 than those that resulted from SA Run 2 for this

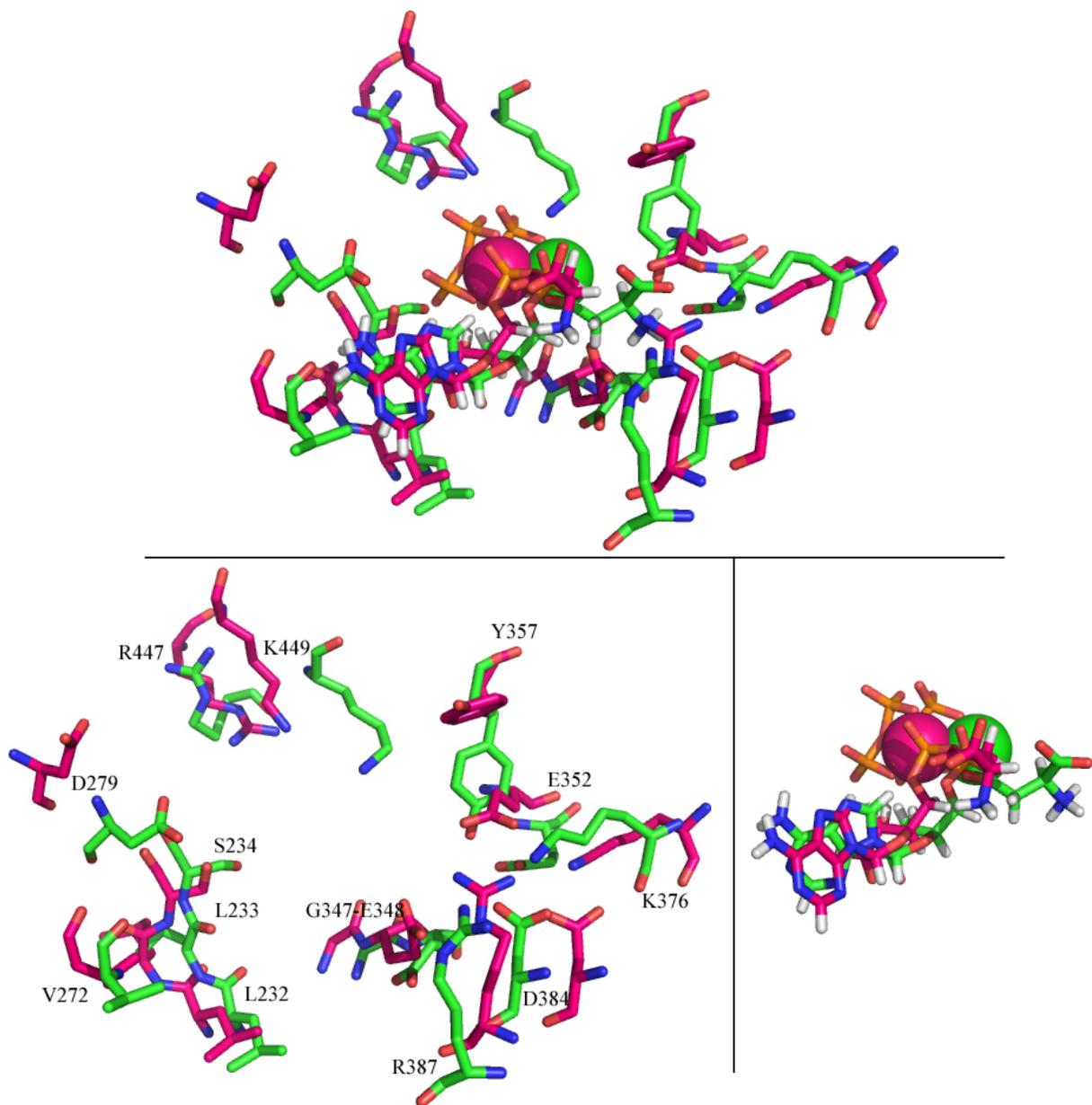


Figure 3-19. Active site comparison between the initial β AspAMP model and the model minimized at 300 K. The initial structure is shown with green carbons, while the minimized structure at 300 K is shown with hot pink carbons. The top view is of the active site with the β AspAMP intermediate, PP_i and Mg^{2+} bound. The bottom left is the same view with the substrates removed for clarity and the bottom right is the only the substrates. The non-carbon atoms are colored as follows: O = red, N = blue, H = white, P = orange, Mg^{2+} = model carbon color.

model. Large shifts are seen in the residues surround the aspartyl tail of the β AspAMP intermediate. This model actually has a slightly different position for the aspartyl tail, with it

positioned coming out of the view, rather than just being shifted above or below the initial position. Despite the large deviations in active site residue positions, the Arg-447 and Glu-348 show side chain shifts in every structure produced over the course of all three simulated annealing runs carried out on the β AspAMP model. With these shifts being a constant across all of the simulations, it seems that these two residues in particular prefer to move to these new positions in this model.

Conclusions

Simulated annealing has been used for some time as an effective tool for the optimization of model systems. Here, two different model systems, AspATP model and β AspAMP model, were optimized through three varying simulated annealing procedures. For both models, the first SA procedure produced the lowest energy structures of all three annealings. However, these systems, in particular the AspATP model, expanded so greatly that they were never quite able to recover and ultimately produced higher rms deviations from their initial systems. The second set of simulated annealing runs performed on the two model systems returned the lowest rms deviations from the original structures of all three SA procedures. In both cases, the minimized structures from the equilibrations at 300 K appeared to be the best overall optimized structures. The final SA procedure that included high pressures in the cooling phase seemed to produce structures and energies that were in between the first two SA procedures in both structure rms deviations and in energies.

The best structure for the AspATP model system came as a result of the structure equilibrated at 300 K of SA Run 2. This system had the lowest overall rmsd in comparison to the initial structure and this entire SA run was the only one performed on AspATP in which the free aspartate did not leave the active site. However, even in the best structures, the position of the

free aspartate is poorly positioned for attack on ATP to form the β AspAMP intermediate. This result and the continued escape of the aspartate from the active site suggests that the model has shortcomings; it is either missing a Mg^{2+} or the final 37 residues that play a role in containing the free aspartate and preparing it for attack. Very recent unpublished work by Megan Meyer on the final 37 residues of AS-B found that the enzyme was still active even (activity was about half that of the wild-type enzyme) when the final 37 residues are removed. This may suggest that these residues are not critical for catalysis. This result in conjunction with the results of the SA would make the presence of a second Mg^{2+} the most likely explanation.

The best structure for the β AspAMP model system also came as a result of the structure equilibrated at 300 K of SA run 2. This structure has the lowest rms deviation in the protein structure from the initial structure of all the trials. One consistent change across every SA run was the flipping of the side chain for the Glu-348 residue (Figure 3-15). One putative role of this residue is that it acts as a gate at the mouth of the intramolecular tunnel connecting the glutaminase and synthetase site. The flip seems to position the carboxylate group, which could block the translocation of the NH_3 by forming a hydrogen bond with it, away from the opening. This movement could allow the NH_3 to pass by and attack the β AspAMP intermediate to form the tetrahedral transition state more easily. This optimized model of AS-B with the β AspAMP intermediate bound is believed to represent a significant improvement in any previous versions of the same model system.

CHAPTER 4 ENHANCING THE PMF99 SCORING FUNCTION FOR MOLECULAR DOCKING AND VIRTUAL SCREENING

Introduction

Recently, the human genome project and efforts in high-throughput crystallography have led to a number high-resolution crystal structures that can be used in drug discovery projects. A routine first approach in drug discovery is the *in silico* screening of a database of molecules into a receptor of interest, using a scoring function to estimate the fit and possibly the binding affinity.^{148,149} Multiple docking programs are currently used in the biotech and pharmaceutical industries.¹⁵⁰⁻¹⁵⁹ Over the past decade, there have been multiple reports of virtual screening applications being successfully employed.¹⁶⁰⁻¹⁶² Despite the success, there is still not one definitive methodology for accurate docking and scoring across a wide array of protein/ligand complexes.

The success of any docking study is constrained by the choice of (i) the intramolecular potential energy description, or scoring function, and (ii) the algorithm for exploring molecular conformations that are accessible to the ligand and, in principle, the residues defining the binding site in the drug target. While a number of solutions have been developed that address the second problem,^{163,164} the difficulty in constructing generally robust scoring functions is illustrated by the existence of the bewildering variety of potentials that have been implemented in software packages such as DOCK,¹⁵⁰ GOLD,¹⁵¹ FlexX,¹⁵² Glide,¹⁶⁵ and AutoDock.¹⁶⁶ The situation is also complicated by the need to devise scoring functions that accurately describe protein/ligand interactions while being sufficiently simple for very rapid evaluation, especially given the number of structures that must be examined for any adequate sampling of the conformational space.¹⁶⁷ Thus, while force field-based potentials may give very accurate estimates of interaction energies for the protein/ligand complex, they possess a relatively complicated mathematical

form. In addition, including the effects of solvation and conformational entropy changes into the calculated interaction enthalpies introduces an additional layer of complexity into the calculation.¹⁶⁸

Given its speed and functional simplicity, a Potential of Mean Force (PMF) scoring potential is ideal for *in silico* screening experiments.¹⁶⁹⁻¹⁷² Like other knowledge-based potentials, such as DrugScore¹⁷³ or SMOG,¹⁷⁴ the simple interaction-pair potentials in PMF are parameterized to reproduce experimentally observed structures of protein/ligand complexes, and therefore implicitly include solvation and entropy effects that are difficult to model with force field-based strategies. On the other hand, the procedure used to parameterize the pairwise PMF potentials is limited by the availability of crystallographic information. The inclusion of nonbonded intramolecular interactions within the ligand is also not accomplished in a straightforward manner in a PMF scoring function.¹⁷¹ Thus, the behavior of a modified PMF99 scoring function (ePMF99) in which the van der Waals terms in the original PMF99 implementation are replaced by pairwise non-bonded interactions was examined.¹⁷⁵ As discussed below, this change permits the use of repulsion radii that are more appropriate for protein/ligand atom pairs that form hydrogen bonding or electrostatic interactions. The effects of this simple modification were demonstrated on the behavior of the ePMF99 potential in reproducing the structures of a well-defined test set of approximately 170 protein/ligand complexes taken from the Protein Data Bank.¹⁷⁶ In addition, the new treatment of non-bonded repulsion was tested for improvements in the utility of the ePMF99 potential in cross-docking studies of thymidine kinase (TK) inhibitors.¹⁶⁰

Methods

Protein/Ligand Structure Preparation

All of the protein/ligand complexes were processed for subsequent docking experiments following an identical procedure to that recommended in previous efforts to assess the performance of the GLIDE scoring function.¹⁶⁵ This procedure therefore yielded partially optimized structures from the original crystallographic coordinates that contained hydrogen atoms, with suitable adjustments to the protonation states of ionizable residues on the protein, optimization of tautomeric forms of histidine, and repositioning of reorientable hydrogen atoms to optimize hydrogen bonding interactions. Although explicit hydrogen atoms are ignored in the enhanced PMF99 scoring function, this procedure was performed to ensure a valid comparison of results obtained using the enhanced PMF99 and GLIDE scoring functions. In addition, this procedure removes steric clashes within the protein/ligand complex. Processing to obtain all structures used in the ligand docking studies was performed using the FirstDiscovery suite of software packages for molecular modeling (V2.5; Schrödinger LLC, New York). Although details of the preparation procedure have been described elsewhere, it is important to note that crystallographic water molecules are often removed from the structure so as to increase the volume of the ligand binding pocket. When cofactors were present in the crystal structure, they were assigned chemical bonds and formal charges consistent with standard Lewis structures, and treated as part of the protein in the docking studies (i.e. their positions were fixed). After the addition of hydrogen atoms to the cofactors, protein residues, and ligand molecule, constrained energy minimizations were performed on the resulting complex, using Macromodel⁹⁷ as implemented within the FirstDiscovery V2.5 package. In these structural optimizations, which employed the MMFF94s force field,¹⁷⁷ side chain hydroxyls (Ser, Thr and Tyr residues) and cysteine thiols were reoriented so as to optimize hydrogen bonding and remove steric clashes by

strongly constraining all non-hydrogen atoms, and allowing hydrogen atoms to move freely in the absence of torsional interactions. Optimization is then repeated using weaker constraints until the rmsd of the non-hydrogen atom positions relative to their initial crystallographic coordinates is less than 0.3 Å.

In preparation for docking, the crystal structure of TK complexed to thymidine (1kim)¹⁷⁸ was processed following a protocol identical to that described above. Crystal structures of TK complexed to nine other purine and pyrimidine analogs (Figure 4-3) were then employed in a standard superimposition procedure to obtain initial poses for the nine ligands within the 1kim TK active site. All ten ligands were then re-docked into the rigid active site and their positions compared with those observed in the original crystal structures.

Docking Algorithm

Ligands were docked into their active sites in the protein structures using a Lamarckian genetic algorithm (LGA) similar to that implemented within the AutoDock 3.0.5 package.¹⁶⁶ Each search was performed for 50,000 generations, or until the rmsd population fitness converged to a value less than 0.001 kcal/mol. Each generation consisted of the following steps performed on the population of genomes (ligand poses): selection using a roulette-wheel strategy,¹⁷⁹ crossover (two-point), mutation (Cauchy-based operator), elitism, and local search. In these calculations, the probabilities of gene mutation (changes in ligand conformation) and gene crossover (generating a new ligand conformation by combining dihedral angles from multiple prior conformations) were defined as 0.02 and 0.80, respectively. Local searches (Solis-West method)¹⁸⁰ were performed on an individual with a probability of 0.06, and were performed for a maximum of 500 iterations. The population comprised 100 individuals (ligand poses) and an elitism algorithm was employed in which the 10 best scoring individuals in each generation were preserved. All genetic algorithm parameter values correspond to those found to be optimal

in prior studies employing the LGA search algorithm in AutoDock.¹⁶⁶ On completion of the LGA search, the population was cleaned by clustering using a 0.3 Å rmsd cutoff in the coordinates of any two ligand conformations.

Results and Discussion

The Enhanced PMF99 Scoring Function

As described in detail elsewhere,¹⁶⁹⁻¹⁷² the PMF99 score for a given protein/ligand complex is computed by the addition of knowledge-based, pairwise atomic potentials for all intermolecular interactions, according to the expression:

$$PMF_score = \sum_{\substack{kl \\ r < r_{cut-off}^{ij}}} A_{ij}(r) \quad (4-1)$$

where $A_{ij}(r)$ is the PMF99 interaction energy of a protein/ligand atom pair separated by a distance r that lies within a user-defined cutoff, r_{cutoff}^{ij} . The PMF interaction energy $A_{ij}(r)$ is calculated as:

$$A_{ij}(r) = -k_B T \ln \left[f_{Vol_corr}^j(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right] \quad (4-2)$$

where k_B is the Boltzmann factor, T is the absolute temperature and the volume correction factor $f_{Vol_corr}^j$ ^{169,171} is included to account for the omission of ligand/ligand interactions. $\rho_{seg}^{ij}(r)$ is the number density of ligand/protein atom pairs of type ij seen at an atom pair distance r . ρ_{bulk}^{ij} is the number density of the atom pair type ij in the reference sphere that has a radius of 12 Å.

Although the magnitude of the PMF99 score likely has little physical meaning, it exhibits a correlation with experimental binding affinities for a diverse set of protein/ligand complexes.¹⁸¹⁻

¹⁸⁴ Evaluating the PMF score is computationally efficient, however, because each pair potential can be pre-computed at 0.2 Å intervals and stored in a lookup table.

As in calculations using empirical force fields, evaluating PMF99 scores requires the assignment of specific atom types. PMF99 currently employs 34 ligand and 16 protein atom types, resulting in 544 unique intermolecular pair potentials.¹⁶⁹ PMF04 has four significant changes in atom types as compared to PMF99. The first change was for the protein ring nitrogen atom type NR as applied to histidine ring nitrogens. It was replaced for the more specific NA atom type for a nitrogen hydrogen bond acceptor and the already existing ND atom type for a hydrogen bond donating nitrogen. A general atom type ME was introduced for metal ions of Zn, Ca, K, Mg, Mn and Fe. This atom type is only invoked for metal ions that are recorded as individual residues in the PDB. The ligand atom type OR for oxygen in a ring structure was merged into the oxygen in an ether bond atom type OE. The final significant change was the addition of the SO atom type for sulfur bonded to more than two atoms or to atoms other than carbon or hydrogen. These changes resulted in PMF04 having one additional protein atom type in comparison to the original PMF99. Thus, PMF04 has 578 unique intramolecular pair potentials.¹⁷²

Since these numerical PMF potentials are obtained from crystal structures of protein/ligand complexes, they contain information about the free energy of interaction between all protein/ligand atom pairs as a function of distance, including the short distances that are typical for covalent bonds. On the other hand, the repulsive contribution to the pairwise potential is not modeled correctly because atom pairs do not get sufficiently close together in crystal structures used in parametrizing the interaction. A non-bonded repulsion term must therefore be added to the knowledge-based portion of the PMF to prevent atoms from being positioned at unphysically short distances. Standard implementations of the PMF99 scoring potential therefore employ van der Waals pairwise potentials from the AMBER force field in order to overcome this limitation

of the knowledge-based method.¹⁸⁵ In the AMBER potential, however, each atom type is assigned a single fixed hardness and van der Waals radius to generate the repulsive and attractive parts of the non-bonded potential, irrespective of the type of atom with which it forms a non-bonded interaction. Consequently, the non-bonded potential in the original version of PMF99 often does not describe hydrogen bonding interactions involving donor-acceptor, donor-donor, and acceptor-acceptor atoms equally well, and can minimize the importance of key hydrogen bonds in modeling protein/ligand complexes.

There is a second problem in employing non-bonded repulsion potentials that are implemented in empirical force fields, such as AMBER, for modeling hydrogen bonds in knowledge-based PMF99 scoring function. This arises from the fact that in molecular mechanics methods,¹⁸⁶ hydrogen bonding is represented as either a composite of electrostatic and van der Waals interactions, or by specific functional forms, such as the 10,12-potential.^{101,104,187} As a result, standard repulsion/dispersion terms are overridden to ensure that atoms are placed at the correct hydrogen bond distance in any energy-minimized structure. The procedure for parameterizing PMF99 pair potentials cannot compensate for such errors in the AMBER repulsion potential, leading to hydrogen bonding distances that may be too long when compared with experimental data (*vide infra*).

One solution to these problems is to employ specific 6,12 Lennard-Jones potentials for the pairwise interactions in the standard PMF99 implementation, each of which can then be fit to account for all of the effects that contribute to a non-bonded interaction, i.e. van der Waals, electrostatics, and hydrogen bonds.¹⁷⁵ This strategy permits hydrogen bonding and charge-charge interactions that would be prohibited on the basis of the van der Waals potential, as well as descriptions of donor-acceptor, donor-donor, and acceptor-acceptor pairwise interactions that are

of equal accuracy. This approach, which is implemented in the enhanced PMF99 scoring function, can be illustrated by considering the hydrogen bonding interaction between a positively charged amino group, such as that present on a lysine side chain (NC), and the oxygen atom of a secondary alcohol (OD). Site-directed mutagenesis experiments have established that such an interaction can contribute up to 3 kcal/mol of the ligand binding energy.^{188,189} Statistical analysis of structural data also supports the importance of this type of hydrogen bond, with the observed N-O distance being less than 3.8 Å.^{190,191} Plotting the AMBER non-bonded potential for the NC-OD interaction, however, shows that the repulsion term dominates if the heavy atoms are placed at a distance of less than 3 Å, as seen in Figure 4-1. This is not a problem in force-field calculations because electrostatic terms that contribute to hydrogen bonding are also included, and the favorable charge-charge interactions release sufficient energy to pay for the interatomic repulsion energy. Hence, the heavy-atom N-O distance in optimized structures obtained using the AMBER force field reproduces experimental values. When the standard PMF99 potential for the same interaction is plotted (Figure 4-1), however, the effects of errors in the repulsive term in the AMBER 6,12 Lennard-Jones potential are not corrected due to the absence of observed protein/ligand atom pairs that are placed too close together. As a result, the standard PMF99 score becomes unfavorable at an N-O distance that is longer than that seen crystallographically for this type of hydrogen bonding interaction. On the other hand, in the enhanced PMF99 potential, the repulsion term is modified thereby allowing the two heavy atoms to get closer together before the structure receives an unfavorable PMF99 score. This specific example illustrates the key problem with the original implementation of the non-bonded potential in the PMF99 method. The key role of the non-bonded potentials is to ensure that bad structures in which atoms are too close together receive a high score. Unfortunately, flaws in the repulsion

term can lead to good poses being scored unfavorably, resulting in their PMF99-derived free energy being higher than other incorrectly docked structures.

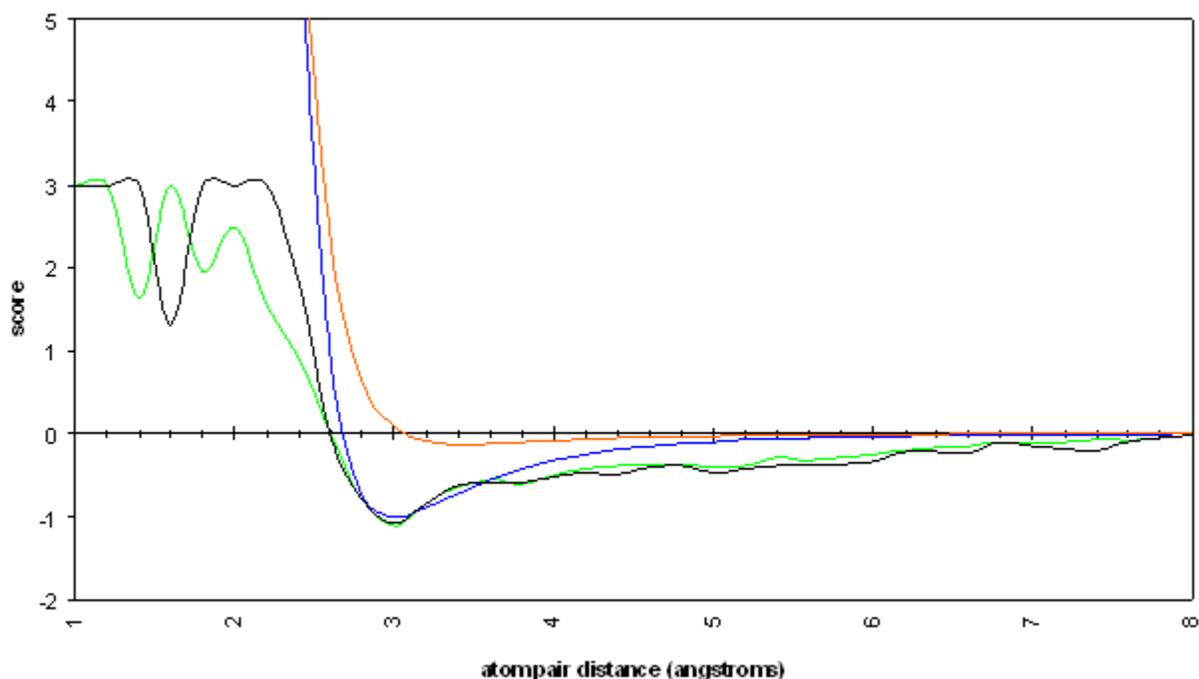


Figure 4-1. A plot of the pairwise scores for protein/ligand atom pair of NC-OD. The original PMF99 score is shown in black with the updated PMF04 shown in green. The non-bonded correction is shown in blue, and the Amber vdW contribution for the original PMF99 is shown in orange for comparison.

Molecular Docking Algorithm

Degrees of freedom in the docking consisted of translation of and rotation about the center of mass of the ligand, and the dihedral angles of all rotatable bonds that were not located in rings. Atoms in the protein were fixed at their crystallographic coordinates during docking and scoring. All calculations in this study employed a Lamarckian genetic algorithm (LGA) search strategy for generating ligand conformations and their positions within the defined protein active sites. In this method, which was first implemented in the AutoDock 3.0.5 package,¹⁶⁶ the Cartesian coordinates of each individual docked ligand conformation (phenotype) are obtained from the degrees of freedom (genes) that are varied in the search algorithm, and the “fitness” is

determined by the PMF99 score computed for that complex. The population size was defined to contain 100 docked ligand conformations, and the 10 best scoring structures were used to generate the ligand population in the subsequent round of optimization. All simulations employed a maximum of 50,000 generations, with an energy convergence value of 0.001 kcal/mol.

Although many values of variables such as population size, crossover rate, and elitism gave similar results for the complexes investigated, systematic investigations of the dependence of optimized ligand pose upon user-defined parameters revealed the critical importance of defining the binding site volume in the protein, especially for solvent exposed sites. Thus, defining an active site region, or box, that is much larger than the ligand creates an empty space that would normally be occupied by solvent and counter-ions outside the protein. In the absence of these molecules in the docking simulation, ligands tend to occupy this empty space so as to avoid enclosure in the active site given that there is no energetic penalty imposed by the PMF99 scoring function to prevent this. The use of a box defining the active site within which the ligand must be located overcomes this problem. On the other hand, the active site volume obviously needs to be large enough to accommodate the maximally extended conformation of the ligand because all relevant intermolecular interactions must be included in the PMF99 scoring function.

The initial ligand pose used in these docking experiments corresponded to the starting pose in the cleaned crystal structure of the protein/ligand complex. While it has been argued that such a choice will bias the search and prevent exploration of alternate ligand conformations within the binding site,¹⁶⁵ in many drug discovery studies there is a significant amount of structure-activity and mutagenesis data that provide some knowledge of the initial ligand pose. Choosing the starting pose of the protein/ligand complex on this basis therefore enhances the probability that

the final docked pose is at least as good as the initial pose by biasing the GA search process. On the other hand, since identifying bound conformations that may not be in accord with chemical intuition, the use of random ligand conformations to generate the starting complex was tested to examine if that alters the outcome of these docking experiments. It was found that employing an initial conformation which differs significantly from the observed structure impacts only the number of steps required in the LGA search that is required to position the ligand correctly within the defined active site.

Assessing the Performance of the ePMF99 Scoring Function in Molecular Docking

The evaluation of the combination of the enhanced PMF99 scoring function in conjunction with the LGA for docking was conducted with a test set of over 170 noncovalently bound cocrystallized protein/ligand complexes. The test set was chosen because it is large and encompassing as is the case in most evaluations of docking potential,^{148,164,192,193} and more specifically the test set was chosen to closely mimic the published test set results of Glide,¹⁶⁵ GOLD¹⁵¹ and FlexX.¹⁵² As in other studies, the strategy employed was to examine whether the combination of the ePMF99 scoring potential with the LGA could reproduce the crystal structures for each of the protein/ligand complexes. Success was evaluated on the basis of the root-mean-square deviation (rmsd) in the coordinates of the docked ligand relative to those observed experimentally. So as to simplify comparisons with previous docking studies, all of the complexes were cleaned using a well-defined procedure in which hydrogen atoms were added to the structure, with a subsequent constrained energy minimization being used to remove physically unreasonable steric clashes.¹⁶⁵ These calculations employed the MMFF94s force field,¹⁰¹ as implemented in the MacroModel software package.⁹⁷ In these calibration studies, all crystallographic ordered water molecules present in the binding site of each complex were removed. This ensured that the initial protein structures were similar to those employed in recent

studies of the Glide algorithms for molecular docking, simplifying any comparison of results between the enhanced PMF and Glide Score potentials. In addition, the presence of these water molecules in virtual screening studies may discriminate against ligands that would displace them from the active site. Finally, removal of the waters aided the investigation of the importance of water-mediated, protein/ligand interactions, and the existence of alternate binding modes in the enlarged active site volume.

For the test set of complexes, the ePMF99 potential employing specifically parameterized repulsion potentials gave a significant improvement in the average RMSD over the original version of the PMF scoring function as seen in Table 4-1. As can also be seen in the same table, PMF04 appears to underperform in comparison to ePMF99 and PMF99. This remained true when the behavior of the two scoring functions was assessed for ligands grouped according to the number of rotatable bonds (Table 4-1). PMF04 is suppose to represent a significant improvement to the PMF99 scoring function due to the reorganization of atom types and due to the increased PDB sample size from which the knowledge-based potential is derived. There are two potential explanations for the underachievement of PMF04 on this particular test set. One possibility is that PMF04 is not implemented correctly in CAChe (the software package used for all the ePMF99, PMF99 and PMF04 docking simulations presented here). A second explanation for the apparent lack of improvement from PMF99 to PMF04 could be in the criteria used for the assessment rendered in this work. In all of the published original validation work by Muegge done on PMF99 and PMF04, he never assessed the validation of his PMF with rms deviations from the original crystal structures as is commonplace for the validation of most docking potentials. Rather, Muegge chose to present his validation in terms of the correlation between the PMF scores and experimental binding constants. Rough comparisons were made between the

PMF04 scores of the final structures docked in CAChe and the PMF04 scores be estimated from Figure 8 of the *Journal of Medicinal Chemistry* paper “PMF Scoring Revisited” by Muegge. CAChe seemed to reproduce the PMF scores implied in this paper.¹⁷² Although very thin in quantity, this comparison seems to suggest that PMF04 has been implemented correctly in CAChe, and that PMF04 does not broadly represent an improvement to PMF99.

Table 4-1. Comparison of rms deviation (Å) for flexible docking

Method	≤10 rotatable bonds		≤20 rotatable bonds		all ligands	
	av rmsd	max rmsd	av rmsd	max rmsd	av rmsd	max rmsd
PMF99 ^a	2.91	13.29	2.96	13.29	2.98	16.32
PMF04 ^b	3.29	15.74	3.59	15.74	3.51	15.74
ePMF99	1.47	7.67	1.58	7.67	1.52	7.67
Glide	1.34	7.55	1.70	11.88	1.98	13.10
GOLD ^c	2.62	14.01	3.01	14.01	3.16	14.01
FlexX ^d	3.07	12.55	3.51	13.41	3.73	15.54

The ePMF99 and Glide test set used for this comparison are identical and comprised of 120 protein/ligand complexes with fewer than 10 rotatable bonds, 159 complexes with fewer than 20 rotatable bonds, and 172 total complexes. ^a PMF99 only varies from ePMF99 in its exclusion of 2mcp (4 rotatable bonds) and lake (22 rotatable bonds); ^b PMF04 varies from ePMF99 in its exclusion of 1byb (24 rotatable bonds), 1frp (7 rotatable bonds), 1hdy (0 rotatable bonds), 1lic (14 rotatable bonds), 1pbd (2 rotatable bonds), 1pph (9 rotatable bonds), 1slt (13 rotatable bonds), 1srj (4 rotatable bonds), 1tpp (5 rotatable bonds), 1ulb (1 rotatable bond), 1xid (6 rotatable bonds), 2ak3 (6 rotatable bonds), 2xis (9 rotatable bonds), 3hvt (1 rotatable bond), 3tpi (7 rotatable bonds), 4fab (4 rotatable bonds) and 8gch (9 rotatable bonds); ^c The GOLD test set contains 64 complexes (≤10 rotatable bonds), 78 complexes (≤20 rotatable bonds), and 85 total complexes that are all taken from the larger ePMF99 test set; ^d FlexX only varies from the ePMF99 test set in its exclusion of 1hdy (0 rotatable bonds). The values exhibited in the table above were calculated from values in Table 4-3.

One complex that was initially poorly reproduced (2.87 Å rmsd) using the ePMF99 scoring function was that between the β-anomer of L-arabinose (abe_b) and the arabinose-binding protein (ABP) as visualized in Figure 4-2a. This was a particularly surprising result for two reasons. First, this small cyclic monosaccharide has relatively limited conformational freedom. Second, PMF calculations for the α-anomer of L-arabinose (abe_a) bound to ABP, which is observed in the same crystal structure,^{56,57} gave a predicted structure, Figure 4-2b, that was

almost identical to the crystal structure (0.20 Å rmsd). On inspection, the best docked pose for β -L-arabinose placed this ligand in an orientation that was rotated by one atom with respect to the oxygen in the ring in its original crystal position (Figure 4-2a). Given the high proportion of hydrogen bonding groups in the ligand, it was noticed that the scoring error was associated primarily with the interactions between β -L-arabinose and the positively charged side chains of Lys-10 and Arg-151. As discussed previously, this problem is compensated for in the ePMF99 implementation if the nonbonded NC-OD and NC-OE potentials are correctly parameterized. Thus, the hydrogen bonding heavy atoms were prevented from reaching their optimal hydrogen bonding distances. Adjusting the aforementioned non-bonded parameters gave docked ligand poses that were almost identical to the experimental structures for both the α - and β -anomers of the ligand (Figure 4-2b and Figure 4-2c, respectively).

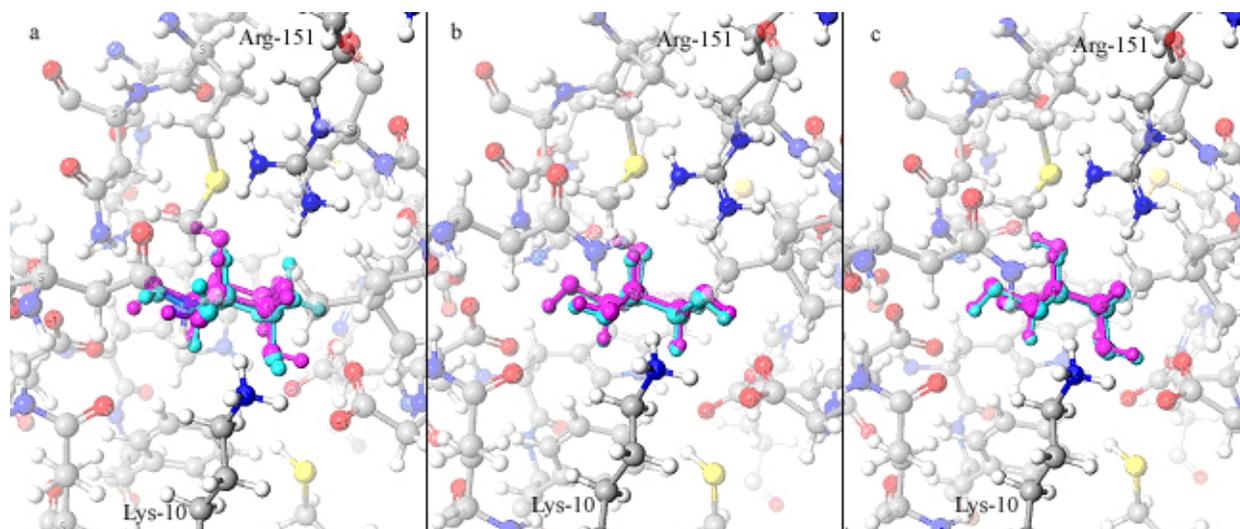


Figure 4-2. Graphical visualization of optimal ligand poses for arabinose in the arabinose-binding protein. (a) Incorrectly docked β -anomer of arabinose (turquoise) on the ligand at its crystallographic position (fuchsia). (b) Correctly docked α -anomer (turquoise) relative to the ligand at its observed position (fuchsia) after optimization of the enhanced PMF scoring function. (c) Correctly docked β -anomer (turquoise) relative to the ligand at its observed position (fuchsia) after optimization of the enhanced PMF scoring function. Protein atoms are rendered as “ball-and-stick” representations (C, grey; H, white; O, red; N, blue; S, yellow).

Comparison to Other Docking/Scoring Algorithms

The large test set used in this study was chosen so that a direct and adequate comparison could be made for the ePMF99/LGA dockings with other widely accepted and implemented methods for docking.^{148,160,161,194,195} A comparison was facilitated by the availability of results collected in recent studies of the Glide docking algorithm. In common with other studies, the comparisons are based solely on the rms deviation in the position of the docked ligand relative to that in the original crystal structure. The results show that the ePMF99 scoring function yields docked structures that are as good as those obtained using Glide for fewer than 10 rotatable bonds. This group is the most relevant to database screening applications that have the goal of finding relatively inflexible leads (Table 4-1).¹⁶⁵ This same group of complexes vastly outperformed the results of GOLD and FlexX. For all ligands with fewer than 20 rotatable bonds, the ePMF99 scoring potential was found to outperform even Glide. That difference grew even larger for the entire test set, once the complexes that contained ligands with more than 20 rotatable bonds were added. The larger and more flexible the ligand, the more the ePMF99 scoring potential outperformed Glide, GOLD and FlexX, in general. On the whole, only approximately 34% of the docked complexes had rmsd values greater than the average total rmsd for all of the complexes of about 1.5 Å. Of those, 73% of the poorly docked complexes contained metal ions, which PMF99 (or ePMF99) does not explicitly treat, while a separate 7% of the ligands that were poorly docked were primarily hydrophobic in nature. Because ePMF99 was developed to improve hydrogen bond type interaction, ligands that are primarily hydrophobic do not benefit from the changes made for the ePMF99 scoring function. This leaves only approximately 20% of the poorly docked ligands with no explicit explanation as to their poor results.

The ePMF99 scoring function was able to adequately reproduce about 40% of the total metal-containing complexes, of which over half contained Ca²⁺ ions (which is a commonly added counter-ion resulting in some of the aforementioned ions not being located in the identified active site). The inability of PMF99 and PMF04 to treat heme- and FeS-groups led to the exclusion of all heme- and FeS-containing complexes from the test set. Three additional complexes (1imb, 1lna and 3cla) were also excluded from the test set because the metal ions they contained were not accounted for in the PMF atom types. The last eight complexes unaccounted for were not included due to errors in the docking or scoring procedure.

Virtual Screening for Thymidine Kinase

Having established that the ePMF99 potential performed well in reproducing the structures of known protein/ligand complexes, the next test was to examine its behavior in virtual screening studies using thymidine kinase (TK) as a model system. This choice was motivated by several factors. First, there are a substantial number of small molecule complexes of this enzyme for

Table 4-2. Accuracy in cross-docking of thymidine kinase inhibitors to the 1kim active site.

Ligand ^a	Bioactivity K _i (μM)	rms deviation of Best Scoring Pose (Å)						
		ePMF99	ePMF99/H ₂ O	Glide ¹⁶⁵	DOCK ¹⁶⁰	FlexX ¹⁶⁰	GOLD ¹⁶⁰	Surflex ¹⁹⁷
dT	0.2 ²⁰³	0.39	0.39	0.45	0.82	0.78	0.72	0.74
ahiu	-	0.40	0.52	0.54	1.16	0.88	0.63	0.87
idu	0.09 ²⁰⁴	0.33	0.55	0.35	9.33	1.03	0.77	1.05
mct	11.4 ²⁰⁵	0.90	0.65	0.79	7.56	1.11	1.19	0.87
hpt	26.6 ²⁰⁶	4.57	0.87 ^b	1.58	1.02	4.18	0.49	1.90
dhbt	-	3.47	1.21	0.68	2.02	3.65	0.93	0.96
hmtt	30.9 ²⁰⁶	2.78	1.49	2.83	9.62	13.30	2.33	1.78
acv	170.0 ²⁰⁷	3.56	5.01	4.22	3.08	2.71	2.74	3.51
gcv	48.0 ²⁰⁸	2.24	2.88	3.19	3.01	6.07	3.11	3.54
pcv	1.5 ²⁰⁵	3.70	3.18	4.10	4.10	5.96	3.01	3.84

^a Literature references for crystal structures of these ligands bound to thymidine kinase: dT (1kim),¹⁷⁸ ahui (1ki6),²⁰⁸ idu (1ki7),²⁰⁸ mct (1e2k),¹⁹⁹ hpt (1e2m),²⁰⁶ dhbt (1e2p),²⁰⁶ hmtt (1e2n),²⁰⁶ acv (2ki5),²⁰⁷ gcv (1ki2),²⁰⁸ pcv (1ki3).²⁰⁸ ^b Docking studies employed an additional water molecule in the ligand binding site of the TK crystal structure (1kim). See text for details.

which high-resolution crystal structures are available, making it a popular system for validating scoring potentials and docking algorithms.^{160,165,196-198} In addition, these ligands exhibit a relatively wide range of activities as TK substrates or inhibitors (Table 4-2),¹⁹⁹⁻²⁰¹ thereby permitting an evaluation of the correlation between scoring differences and biological activity. Finally, TK presents a significant challenge for validating *in silico* virtual screening strategies because (i) the active site is solvent accessible and can undergo conformational changes to accommodate a wide variety of structures, (ii) known ligands are nucleotides and/or nucleotide analogs, and (iii) the differential participation of water molecules in mediating binding depending upon the structural class of ligand being docked into the active site.²⁰²

Coordinates are available for thymidine kinase complexed with ten different ligands.^{178,199,200,206-208} As in previous studies, and since the coordinates of the protein main chain atoms in all these structures are almost superimposable, the ePMF99 potential and LGA algorithm were used to dock the ten ligands into a single structure (1kim)¹⁷⁸ and the X-ray pose of each ligand was inserted into this dT-bound conformation of the TK active site for evaluation purposes. In preparing the initial structures for these docking studies, the choice was made to remove all crystallographically observed water molecules from the TK active site so as to facilitate comparison of our docking results with those reported in studies employing the Glide software package.¹⁶⁵ Since the TK site is designed to bind 2'-deoxythymidine (dT), it was anticipated that the six pyrimidine-based inhibitors (ahiu, mct, dhbt, idu, hmtt, hpt) would dock well within the 1kim structure. In contrast, the remaining three purine derivatives (acv, gcv, pcv) are larger and may require the side chains of the protein to adopt alternate conformations (Figure 4-3). Initial docking experiments using the ePMF99 potential gave structures that were in good agreement with the crystallographic poses for only the substrate dT and four related pyrimidine-

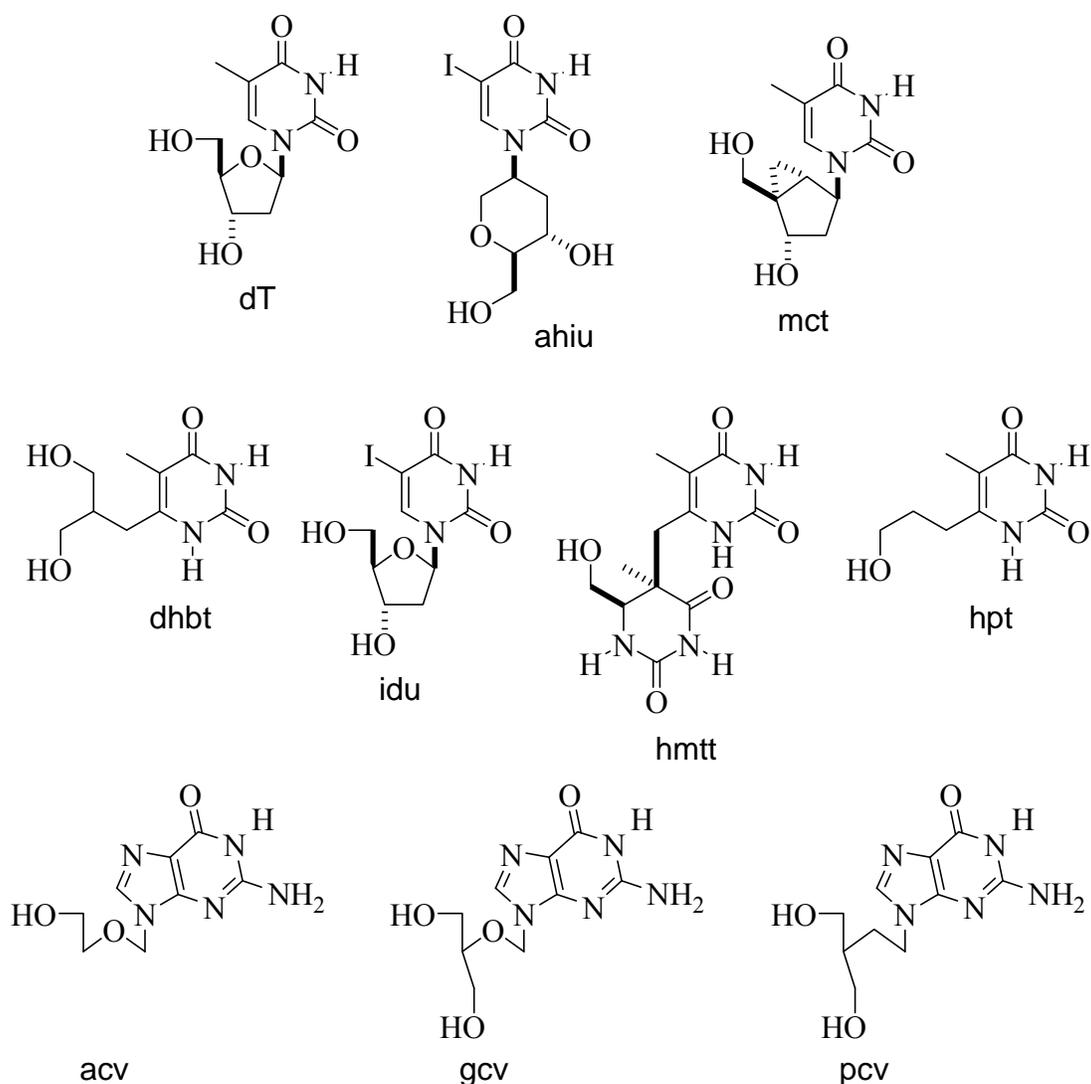


Figure 4-3. Structures of ligands employed in cross-docking studies on thymidine kinase. Compounds are named by the abbreviated ligand names.

based inhibitors (ahiu, mct, idu, dhbt) (Table 4-2). The remaining pyrimidine derivatives (hmtt, hpt), however, differ from dT in lacking the deoxyribose ring and having a bulky substituent connected to C-6 rather than N-1 (Figure 4-3). In the case of hmtt, which is the largest of the pyrimidine-derived ligands, the key problem that precludes locating its optimum pose in the TK active site is most likely associated with the conformation of the Gln-125 side chain in the 1kim protein structure, as noted in similar cross-docking experiments performed using the Glide software package. Thus, the amide side chain of Gln-125 hydrogen bonds to hmtt active site. As

a result, the LGA docking strategy failed to locate a ligand conformation in which the pyrimidine ring of this compound forms the observed hydrogen bonding network with the side chains of residues Gln-125 and Arg-176. With this, the performance of the docking/scoring algorithms was examined using an active site structure in which a third crystallographic water was retained for this problematic ligand (hpt), since such interactions were reported to be important in previous structural studies.²⁰² Thus, for this ligand, an additional water molecule was placed in the 1kim active site at its position in the 1e2m crystal structure as were the three waters observed in the TK/hmtt complex (1e2n) (Figure 4-4a).²⁰⁶ However, a similar interaction cannot be formed when this active site residue is maintained in the conformation observed when TK binds to its substrate dT (Figure 4-4b). Other scoring potentials and docking algorithms also fail to position this TK inhibitor within the 1kim crystal structure, confirming the importance of active site flexibility in cross-docking studies. In the case of the hpt ligand, graphical visualization of the incorrectly docked structure showed that the bulky C-6 substituent was placed within an active site cavity that accommodates the deoxyribose moiety of dT in 1kim. The ligands hpt, dhbt and hmtt were docked into the resulting active site structure, which contained three water molecules. The third water molecule occupied a region of the active site that normally contains the deoxyribose ring of dT and the pyrimidine derivatives (Figure 4-4b). In this case, the use of the ePMF99 potential and LGA docking algorithm yielded a model of the TK/hpt complex with a much lower rmsd value (Table 4-2). Interestingly, the introduction of the third water molecule was not necessary for the docking of the dhbt and hmtt ligands because these exhibited intramolecular hydrogen bonds to N-1 of the pyrimidine moiety thereby preventing any rotation of the ligand so as to occupy the cavity for the sugar ring of dT.

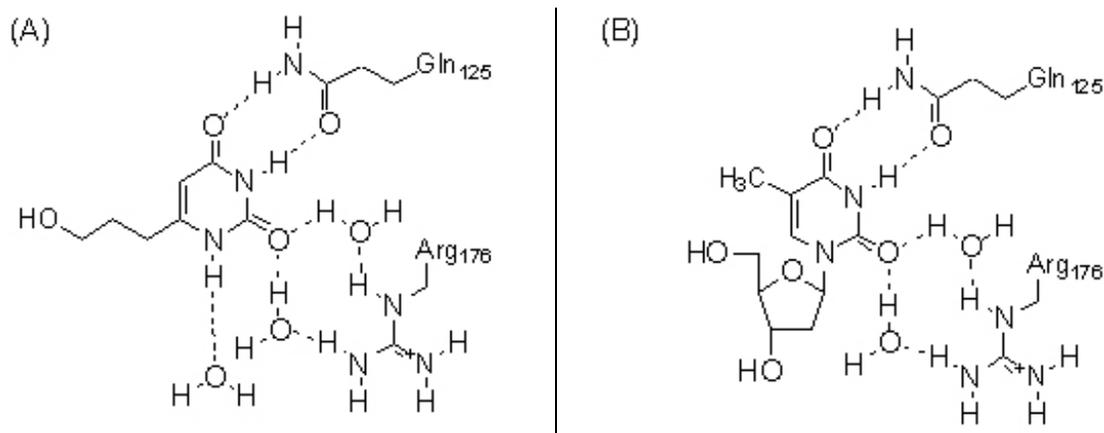


Figure 4-4. The effect of water on the binding of hpt and dT with TK. (A) Representation of the role of water molecules in mediating protein/ligand interactions in the complex between hpt and thymidylate kinase after docking the ligand into the 1kim protein structure. Note that the additional water molecules occupy the space that is usually occupied by the deoxyribose ring of dT in the 1kim crystal structure. (B) The water-ligand interactions seen in the TK/hpt complex are absent in the cognate complex containing dT, the natural substrate for the enzyme, so that protein residue side chains are directly hydrogen bonded to the pyrimidine ring.

Correlation of the ePMF99 Score and Biological Activity

Given that the goal of cross-docking studies is often to predict molecular structures that exhibit biological activity, the extent to which the ePMF99 scores for the set of pyrimidine and purine ligands were correlated with binding affinity was examined. Although the size of the test set obviously is limited, the data showed a qualitative correlation of the ePMF99 scores and $\log K_i$ for the pyrimidine-derived ligands (Figure 4-5). On the other hand, this was not the case for the purine-based analogs, although this likely reflects the fact that these compounds (acv, pcv, gcv) were docked into a rigid active site optimized for interactions with dT.

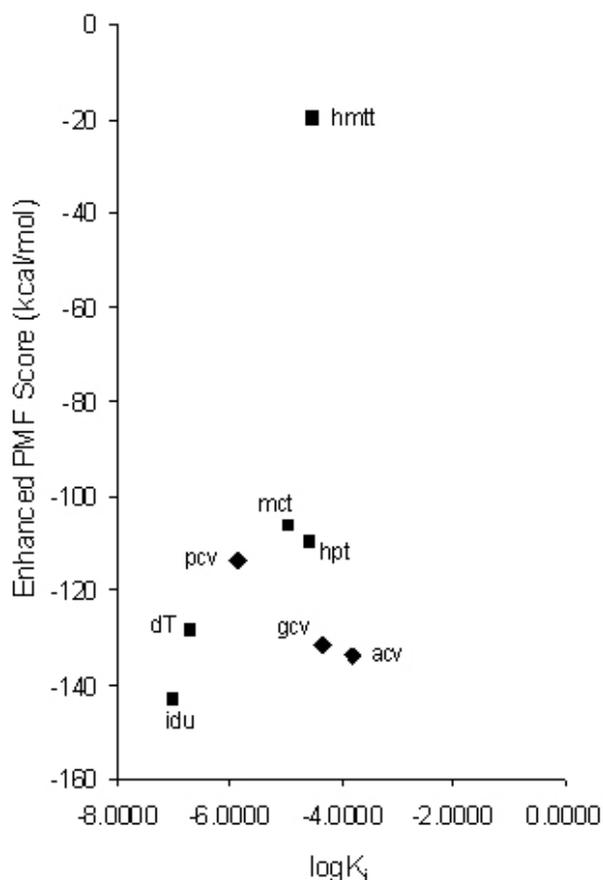


Figure 4-5. Plot showing the correlation of the ePMF99 score vs. logK_i for the pyrimidine (■) and purine (◆) analogs used in the cross-docking studies.

Conclusions

In summary, these studies suggest that modification of the PMF99 scoring function so as to treat the repulsion term in a more physically reasonable manner permits considerable improvement in reproducing observed crystal structures. As noted previously in work employing the Glide algorithm,¹⁶⁵ it is important to employ carefully prepared initial structures if accurate ligand docking is to be attained using the LGA/ePMF99 strategy. Furthermore, the docked ligands of lowest rmsd when compared to the crystal structure do, in general, exhibit the best ePMF99 score. Importantly, the use of the ePMF99 scoring potential yields optimized ligand poses that are comparable to those obtained by the scoring function implemented in Glide for the

protein/ligand complexes examined in this study (Table 4-3). In addition, the ePMF99 performed well in cross-docking studies on TK for which extensive structural and binding data are available for a variety of ligands. As noted in previous studies, the key problems encountered in modeling this system were associated with choosing the appropriate number of active site water molecules, and the failure to include protein side chain flexibility during the LGA search. Thus, it is important to perform a series of calculations that systematically vary the number of active site waters since large changes in ligand structure result in the creation of active site voids (when the protein is held rigid) that, if not filled by water, will bias the final docked pose of the ligand. Finally, as reported for the original implementation of the PMF99 scoring potential,¹⁶⁹ obtaining a good correlation between the ePMF99 score and observed biological activity when flexible ligands are docked into rigid protein cavities requires the comparison of ligands that do not differ radically in size and/or overall structure. On this point, it is noted that this problem does not seem solvable using the available scoring methods, about which considerable controversy still seems to exist in the literature,^{148,194} since it primarily arises from failing to include active site flexibility during ligand docking.

Table 4.3 Complete docking test set and calculated rmsd of docked ligand vs original crystal structure ligand.

complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX	complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX
121p (Mg)	5.72	5.66	7.83	1.57	n/a	1.29	1aaq	0.37	0.70	0.70	1.30	12.85	1.75
1abe	0.16	0.25	0.16	0.17	0.86	1.16	1abf	0.19	0.25	0.15	0.20	n/a	1.27
1acj	0.36	0.42	0.99	0.28	4.00	0.49	1acm (Zn)	4.73	5.41	3.28	0.29	0.81	1.39
1aco (FeS)	n/a	n/a	n/a	1.02	0.86	0.96	1aha	0.16	2.74	4.80	0.11	0.51	0.56
1ake	0.39	n/a	n/a	3.35	n/a	1.18	1apt	0.54	0.52	0.87	0.58	1.62	1.89
1atl (Zn)	2.79	2.72	4.63	0.94	n/a	2.06	1avd	0.59	3.18	3.46	0.52	n/a	1.22
1azm (Zn)	2.66	2.69	2.74	1.87	2.52	2.37	1baf	0.75	0.96	1.64	0.76	6.12	8.27
1bbp	0.73	8.72	2.24	4.96	n/a	3.75	1bma (Ca)	1.09	6.84	8.45	9.31	n/a	13.41
1byb	0.35	12.49	n/a	10.49	n/a	1.62	1cbs	0.52	0.75	0.84	1.96	n/a	1.68
1cbx (Zn)	1.66	6.29	6.24	0.36	0.54	1.35	1cde	2.27	2.15	0.60	1.29	n/a	7.45
1cdg (Ca)	n/a	n/a	n/a	3.98	n/a	4.87	1cil (Zn)	1.18	1.18	1.77	3.82	n/a	3.85
1com	0.92	0.60	2.80	3.64	n/a	1.62	1coy	0.68	0.97	1.05	0.28	0.86	1.06
1cps (Zn)	6.59	6.59	3.75	3.00	0.84	0.99	1ctr (Ca)	4.31	5.16	7.05	3.56	n/a	2.82
1dbb	0.30	0.29	7.28	0.41	1.17	0.81	1dbj	0.75	0.75	0.94	0.20	0.72	1.22
1dbk	0.68	0.67	4.33	0.47	n/a	0.76	1dbm	0.72	0.79	3.11	1.97	n/a	2.08
1did (Mn)	4.08	4.91	6.13	3.82	3.72	4.22	1die (Mg)	2.68	2.67	2.91	0.79	1.03	4.71
1dr1 (Ca)	0.33	0.37	0.68	1.47	1.41	5.64	1dwb	3.62	3.19	4.80	0.25	n/a	0.54
1dwc	0.73	0.84	5.38	0.87	n/a	1.19	1dwd	1.24	1.30	5.72	1.32	1.71	1.66
1eap	0.82	1.60	5.24	2.32	3.00	3.72	1eed	0.75	1.97	2.43	5.90	12.43	9.78
1ela (Ca)	4.52	5.15	2.77	1.60	n/a	9.71	1elb (Ca)	4.99	10.57	4.08	4.40	n/a	7.17
1elc (Ca)	6.82	6.94	5.85	8.22	n/a	4.74	1eld (Ca)	4.16	2.40	7.47	0.67	n/a	6.98
1ele (Ca)	4.89	4.32	4.52	2.52	n/a	10.73	1epb	0.70	1.79	1.94	1.78	2.08	2.77
1eta	2.82	7.62	7.74	2.92	11.21	8.46	1etr	0.99	1.99	6.04	1.48	4.23	7.24
1fen	0.96	1.81	1.86	0.66	n/a	1.39	1fkg	1.10	1.21	2.03	1.25	1.81	7.59
1fki	0.10	0.29	0.14	1.92	0.71	0.59	1frp (Zn)	0.64	0.65	n/a	0.27	n/a	1.89
1ghb	0.37	0.40	5.80	1.89	1.45	1.33	1glp	5.27	7.68	7.36	0.34	n/a	0.47
1glq	n/a	n/a	n/a	0.29	1.35	6.43	1hdc	3.60	4.64	4.39	0.58	10.49	11.74
1hdy (Zn)	2.57	2.65	n/a	1.74	0.94	n/a	1hef	1.05	16.32	6.68	5.30	1.87	15.32
1hfc (Zn)	6.43	3.83	7.90	2.24	n/a	2.51	1hgg	n/a	n/a	n/a	2.10	n/a	10.05
1hgh	n/a	n/a	n/a	0.28	n/a	4.14	1hgi	n/a	n/a	n/a	0.28	n/a	0.97
1hgj	n/a	n/a	n/a	0.18	n/a	3.98	1hri	1.07	2.36	1.25	1.59	14.01	10.23
1hsl (Cd)	0.26	0.26	0.32	1.31	0.97	0.59	1hti	1.74	1.83	1.72	4.40	n/a	1.54
1hvr	0.56	0.65	15.74	1.50	n/a	3.35	1hyt (Zn)	0.73	0.72	3.24	0.28	1.10	1.62
1icn	0.87	3.57	7.97	2.34	8.63	10.52	1ida	1.54	1.51	1.44	11.88	12.12	11.95

Table 4-3. Continued.

complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX	complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX
ligj	0.58	10.19	5.32	1.30	9.42	7.17	limb (Gd)	n/a	n/a	n/a	0.89	n/a	4.71
livb (Ca)	0.25	0.75	0.45	4.97	n/a	1.29	livc (Ca)	2.24	2.25	1.86	1.94	n/a	2.21
livd (Ca)	0.56	3.09	3.42	0.72	n/a	5.42	live (Ca)	0.89	1.29	1.57	2.61	2.16	5.34
livf (Ca)	0.57	0.55	1.32	0.53	n/a	6.97	llah	0.15	0.18	0.17	0.13	n/a	0.28
llcp (Zn)	3.25	11.50	3.96	1.98	n/a	1.65	lldm	0.45	8.32	8.60	0.30	1.00	0.74
llic	3.07	3.55	n/a	4.87	10.78	5.07	llmo	1.00	1.03	2.93	0.93	n/a	4.49
llna (Co)	n/a	n/a	n/a	0.95	n/a	5.40	llst	0.14	0.12	0.18	0.14	0.87	0.71
lmbi (heme)	n/a	n/a	n/a	1.68	n/a	0.47	lmcr	1.71	1.82	1.18	4.33	6.23	10.04
lmdr (Mg)	0.97	0.97	1.16	0.52	0.36	0.88	lmlc	0.69	1.77	1.25	0.32	n/a	1.45
lmmq (Zn)	7.67	6.10	7.28	0.92	n/a	0.52	lmrg	0.12	0.41	3.99	0.30	n/a	0.81
lmrk	1.12	1.15	1.39	1.20	1.01	3.55	lmup (Cd)	1.42	1.42	2.01	4.37	3.96	3.82
lnco	1.22	9.66	9.84	6.99	n/a	5.85	lnis (FeS)	n/a	n/a	n/a	0.97	4.29	1.41
lpsc (Ca)	0.38	0.38	1.23	1.21	n/a	2.12	lpbd	0.35	0.35	n/a	0.21	0.57	0.33
lpha (heme)	n/a	n/a	n/a	0.69	1.24	n/a	lphd (heme)	n/a	n/a	n/a	1.22	0.85	0.65
lphf (heme)	n/a	n/a	n/a	1.14	n/a	4.23	lphg (heme)	n/a	n/a	n/a	4.32	1.35	4.74
lpoc (Ca)	1.12	0.88	1.29	5.09	1.27	9.25	lppc (Ca)	2.44	5.42	6.45	7.92	n/a	3.05
lpph (Ca)	4.58	6.44	n/a	4.31	n/a	4.91	lppi (Ca)	0.78	0.82	1.31	6.24	n/a	6.91
lppk	0.52	0.69	0.69	0.45	n/a	1.54	lppl	0.51	0.65	0.77	2.82	n/a	5.62
lppm	0.92	1.15	1.11	0.62	n/a	8.27	lpso	1.19	1.68	9.41	13.10	n/a	1.61
lrbp	0.96	1.33	2.50	0.96	n/a	1.13	lrds	0.86	0.78	5.84	3.75	4.78	4.89
lrne	0.66	1.04	0.64	10.08	2.00	12.24	lrnt	1.09	1.03	1.05	0.72	n/a	1.90
lrob	0.36	0.37	4.41	1.85	3.75	7.70	lslt	1.44	1.62	n/a	0.51	0.78	1.63
lsnc (Ca)	0.65	0.66	3.89	1.91	n/a	7.48	lsrj	0.28	7.66	n/a	0.58	0.42	2.36
lstp	1.51	1.28	4.42	0.59	0.69	0.65	ltdb	2.49	3.31	4.22	1.46	10.48	10.10
lthy	3.20	2.71	3.56	2.31	n/a	2.67	ltka (Ca)	0.50	0.63	0.46	2.28	1.88	1.17
ltlp (Zn)	1.58	1.70	3.19	1.86	n/a	2.85	ltmn (Ca)	0.91	1.77	2.53	2.80	1.68	0.86
ltng (Ca)	2.64	2.63	6.22	0.19	n/a	1.93	ltnh (Ca)	2.43	2.46	6.46	0.33	n/a	0.56
ltni (Ca)	2.69	2.67	6.29	2.18	n/a	2.71	ltnj (Ca)	2.37	1.87	5.24	0.35	n/a	0.89
ltnk (Ca)	1.44	1.99	5.94	0.87	n/a	1.41	ltnl (Ca)	1.65	1.68	7.71	0.23	n/a	0.71

Table 4-3. Continued.

complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX	complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX
1tph	0.57	10.41	8.03	0.20	n/a	1.50	1tpp (Ca)	2.45	2.54	n/a	1.12	0.43	1.11
1trk (Ca)	0.88	0.86	0.36	1.64	n/a	1.57	1tyl (Zn)	1.71	10.87	4.48	1.06	n/a	2.34
1ukz	1.95	4.79	3.26	0.37	n/a	0.94	1ulb	4.68	5.72	n/a	0.28	0.32	3.37
1wap	n/a	n/a	n/a	0.12	n/a	0.57	1xid (Mn)	2.79	4.12	n/a	4.30	0.92	2.01
1xie (Mn)	2.66	2.72	2.46	3.86	0.69	1.94	2ack	0.68	0.71	1.02	0.97	4.99	2.21
2ada (Zn)	0.36	0.45	0.25	0.53	0.40	0.67	2ak3	1.10	10.85	n/a	0.71	5.08	0.91
2cgr	0.50	0.32	0.40	0.38	0.99	3.53	2cht	0.30	0.48	3.82	0.42	0.59	4.58
2cmd	0.83	13.29	3.82	0.65	n/a	3.75	2cpp (heme)	n/a	n/a	n/a	0.17	n/a	2.94
2ctc (Zn)	1.31	1.85	6.77	1.61	0.32	1.97	2dbl	0.31	3.12	9.04	0.69	1.31	1.49
2fox (4fxn)	0.31	12.94	1.25	0.44	n/a	1.04	2gbp (Zn)	0.18	0.20	0.16	0.15	n/a	0.92
2lgs (Mn)	1.72	4.44	6.13	7.55	n/a	4.63	2mcp	0.99	n/a	1.82	1.30	4.37	2.07
2phh	0.28	0.99	0.23	0.38	0.72	0.43	2pk4	0.40	0.55	1.01	0.86	1.34	1.66
2plv	1.72	1.95	1.78	1.88	13.92	7.85	2r04	0.73	0.86	1.68	0.80	n/a	12.55
2r07	0.71	11.82	12.42	0.48	8.23	11.63	2sim	0.70	1.04	1.57	0.92	0.92	1.99
2tmn (Zn)	1.11	1.11	1.24	0.58	n/a	5.16	2xis (Mg)	1.71	4.45	n/a	0.85	n/a	1.54
2yhx	n/a	n/a	n/a	3.84	1.19	2.25	2ypi	1.04	8.26	1.42	0.31	n/a	1.22
3cla (Co)	n/a	n/a	n/a	8.51	5.45	6.42	3cpa (Zn)	1.71	3.24	3.53	2.40	1.58	2.53
3hvt	0.48	0.78	n/a	0.77	1.12	10.26	3mth (Zn)	3.72	3.75	1.10	5.48	10.12	1.59
3ptb (Ca)	3.24	4.40	4.42	0.27	0.96	0.55	3tpi (Ca)	0.52	0.52	n/a	0.49	0.80	1.07
4aah (Ca)	0.39	0.45	0.93	0.30	0.42	5.93	4cts	0.64	0.66	0.55	0.19	1.57	1.53
4dfr (Ca)	1.12	1.06	1.23	1.12	1.44	1.40	4fab	2.55	5.58	n/a	4.50	5.69	4.95
4fbp	0.27	6.19	8.46	0.56	n/a	1.78	4hmg	0.55	0.66	4.70	0.78	n/a	5.74
4phv	0.32	0.40	0.51	0.38	1.11	1.12	4tim	0.59	0.67	1.47	1.32	n/a	4.09
4tln (Zn)	4.29	5.59	6.50	2.24	n/a	3.68	4tmn (Zn)	0.97	1.00	1.19	1.87	n/a	8.35
4ts1	0.50	6.31	0.97	0.85	n/a	1.41	5abp	0.17	0.37	0.21	0.21	n/a	1.17
5cpp (heme)	n/a	n/a	n/a	0.59	n/a	1.49	5cts	0.59	1.40	0.36	0.28	n/a	11.61
5p2p (Ca)	1.19	1.25	1.13	1.82	1.55	1.00	5tim	2.60	2.61	3.89	0.58	n/a	1.99
5tmn (Zn)	2.47	1.97	2.39	2.43	n/a	4.38	6abp	0.16	0.29	0.24	0.40	1.08	1.12
6cpa (Zn)	0.68	0.85	6.81	4.58	n/a	6.61	6rnt (Ca)	1.39	1.14	8.41	2.22	1.20	4.79
6tim	0.72	0.91	0.91	1.73	n/a	1.60	6tmn (Zn)	2.13	2.19	8.96	2.66	n/a	5.10
7cpa (Zn)	0.93	0.99	1.56	4.14	n/a	9.11	7tim	0.49	0.48	1.45	0.14	0.78	1.49

Table 4-3. Continued.

complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX	complex	ePMF99	PMF99	PMF04	Glide	GOLD	FlexX
8atc (Zn)	1.96	6.05	2.05	0.37	n/a	0.62	8gch	1.38	6.16	n/a	0.30	0.86	8.91
9hyp	1.01	0.76	1.25	2.68	n/a	15.54							

CHAPTER 5 CONCLUSIONS

Concluding Remarks

A great deal of progress has been made in identifying potent inhibitors for human AS, however, the absence of a well-defined structure or structures of AS has slowed these efforts. The development of a model of AS-B with the β AspAMP intermediate bound at synthetase active site has the potential for aiding the inhibitor design process in the future.

Parameters have been developed in CHARMM for the β AspAMP intermediate that allowed for a thoroughly optimized model of AS-B to be obtained. This model represents a significant improvement in the structural relationship between the active site and the β AspAMP intermediate. Significant side chain reorientation was seen in the model, in which Arg-447 and Glu-348 flipped the positions of their side chains to form new interaction with the β AspAMP intermediate. More significantly, the Glu-348 side chain movement appears to have opened up the intramolecular tunnel that connects the glutaminase active site with the synthetase active for NH_3 translocation. This result could explain previous finding that suggest that Glu-348 acts as a gate for NH_3 translocation and is important in β AspAMP intermediate formation.

The refinement of this model paves the way for the virtual screening of compound libraries on AS-B. One method that was proven to be very effective in both molecular docking and virtual screening was the LGA search algorithm coupled with the ePMF99 scoring function. The improvement made to the ePMF99 scoring function allows or allowed for more accurate representation of the non-bonded repulsion term in comparison to the standard PMF99. This docking and scoring combination was shown to work well when implemented in the CAChe program and could be used for the virtual screening of compound libraries on the newly refined AS-B model.

The Future of the AS-B Project from a Computational Perspective

The model of AS-B presented in this work paves the way for the refinement of a full quantitative model of the enzyme with an array of possible ligands bound in the active sites. In order to be able to confidently present a full quantitative model of AS-B, a few barriers must still be surpassed. First, a more detailed refinement of the parameters of the β AspAMP intermediate should be accomplished. For this to be achieved, the accurate assignment of the vibrational spectra calculated in CHARMM and via *ab initio* methods of the model complex is necessary. An accurate assignment of the vibrational spectra would assist in directing the refinement of the current parameters, so that the working 2-dimensional energy surface of the model complex calculated by CHARMM (Figure 2-7) would better reproduce the calculated *ab initio* surface (Figure 2-3). Second, building the final 37 residues, whose purpose and structure remain unclear, on to the model is a challenging task that would yield a completed picture as to the structure of AS-B. Correctly modeling the tail of AS-B would answer the questions that remain concerning the importance of this region of the sequence. It is possible that these residues play no role in the binding of substrates or mediation of catalysis, however, until experiment or theory definitively answers that question, the importance of these residues cannot be taken for granted. A full quantitative model would be a good target for *in silico* library screening, as well as a good starting point for the docking and parameterization of the known inhibitors of glutamine-dependent AS, as seen in Figure 5-1.

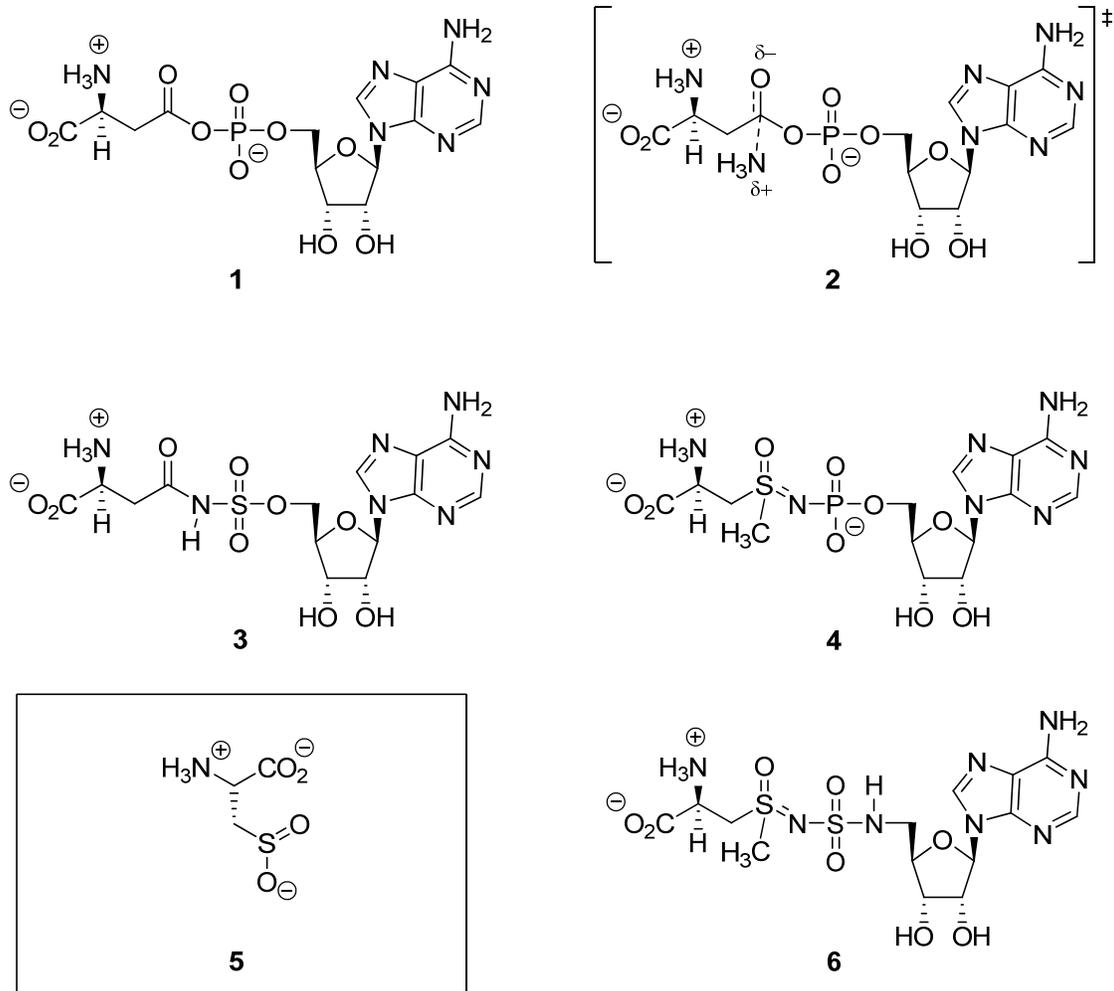


Figure 5-1. Structures of the AS reaction intermediate, β AspAMP **1**, transition-state **2**, and known synthetase site inhibitors. The N-acylsulfonamide **3**²⁰⁹ was modeled after the β AspAMP **1**, whereas the sulfoximine **4**^{23,210,211} and the sulfoximine derivative **6**¹⁴⁵ were modeled after the transition-state. L-cysteine sulfinic acid (CSA) **5**^{210,212,213} is a competitive inhibitor with respect to aspartate.

Modeling the Inhibitors

Including the inhibitors in the model of AS-B will require a significant effort due to the need for novel parameters for each unique inhibitor. The parameterization of the sulfoximine **4** have already begun, and as was detailed in Chapter 2, the first steps taken were to identify the necessary parameters (Table 5-1) and calculate the first of two time-intensive *ab initio* PES for the corresponding model complex, seen in Figure 5-2. A preliminary CHARMM atom type

Table 5-1. Sulfoximine **4** model complex and CHARMM parameters necessary.

	Bonds		
	SA-CB	PA-NA	SA-OD1
	SA-NA	SA-CG	
Angles			
	O5'-PA-NA	O2A-PA-NA ^a	PA-NA-SA
	NA-SA-CG	NA-SA-OD1	NA-SA-CB
	CG-SA-OD1	CG-SA-CB	OD1-SA-CB
	SA-CG-HG1 ^b	SA-CB-HB1 ^c	SA-CB-CA
	HA-CA-HA2 ^d		
Dihedrals			
	O5'-PA-NA-SA	O2A-PA-NA-SA ^e	PA-NA-SA-OD1
	PA-NA-SA-CG	NA-SA-CG-HG1 ^f	PA-NA-SA-CB
	NA-SA-CB-CA	NA-SA-CB-HB1 ^g	

^a The O1A-PA-NA angle is included; ^b All three SA-CG-HG angles; ^c The SA-CB-HB2 angle is included; ^d All three HA-CA-HA angles; ^e The O1A-PA-NA-SA dihedral is included; ^f All three NA-SA-CG-HG dihedrals; ^g The NA-SA-CB-HB2 dihedral is also included. The three dihedrals in bold are the three primary dihedrals of the model complex (and sulfoximine) that would be the focus of the parameter development.

assignment for the model complex is shown in Table 5-2 with partial charges calculated

according to the CHelpG scheme¹¹⁶ at the HF/6-31+G(d) level in GAUSSIAN 03. This partial charge calculation was performed in the same manner as the calculation of partial charges for the β AspAMP **1** intermediate detailed in Chapter 2.

Table 5-2. Initial CHARMM atom types and CHelpG charges assigned for the sulfoximine model complex.

Atom Name	Atom Type	Charge	Atom Name	Atom Type	Charge
C5'	CN8B	0.21	HG1	HA	0.17
H5'	HN8	-0.02	HG2	HA	0.13
H5''	HN8	-0.01	HG3	HA	0.16
HC5'	HN8	0.01	OD1	OS	-0.74
O5'	ON2	-0.58	CB	CT2	-0.02
PA	P	1.70	HB1	HA	0.07
O1A	ON3	-0.92	HB2	HA	-0.01
O2A	ON3	-0.98	CA	CT1	-0.08
NA	NN1C	-0.97	HA	HB	0.04
SA	SS	1.34	HA2	HB	0.02
CG	CT3	-0.53	HA3	HB	0.01

The three consecutive primary dihedrals that are unique to the sulfoximine **4** inhibitor represent the largest barrier to incorporation of the sulfoximine **4** in the AS-B model. Calculating two separate 2-dimensional *ab initio* energy surfaces for the three adjacent dihedrals is necessary for the accurate parameterization of this compound. Figure 5-2 shows the CA-CB-SA-NA versus CB-SA-NA-PA energy surface as calculated in GAUSSIAN 03 at the HF/6-31+G(*d*) level, as was done for the β AspAMP **1**. This energy surface was chosen as the first of the two possible

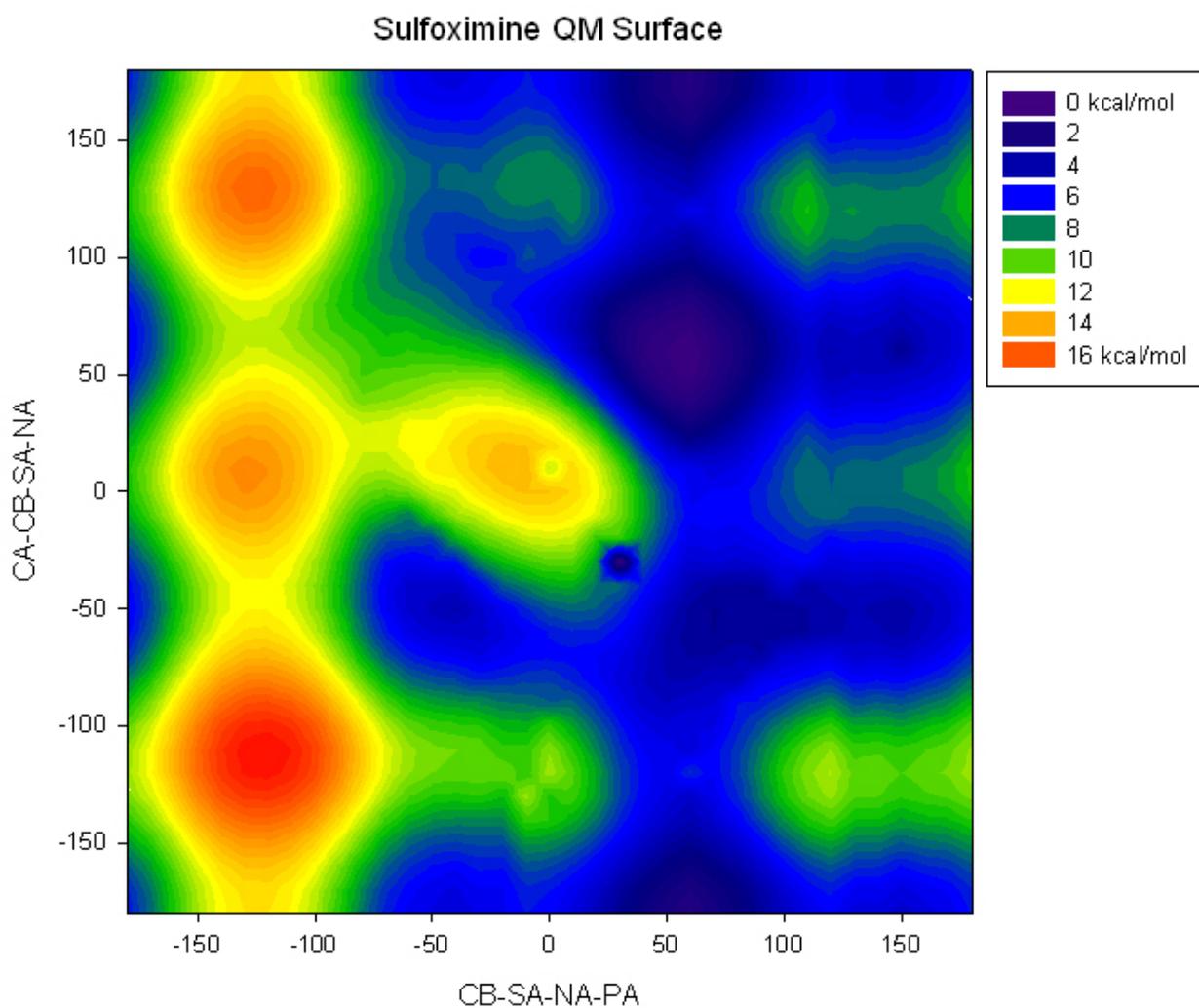


Figure 5-2. The 2-dimensional *ab initio* energy surface for dihedral angle CB-SA-NA-PA versus dihedral angle CA-CB-SA-NA of the sulfoximine model complex. The global minimum energy (kcal/mol) was set to zero with all other energies offset relatively.

surfaces calculated with the idea that the SA-NA-PA-O5' dihedral could be modeled off of the CB-O3A-PA-O5' dihedral from β AspAMP **1** as an initial guess. Ideally, a second *ab initio* surface should be calculated for the adjacent dihedrals SA-NA-PA-O5' versus CB-SA-NA-PA in order to more accurately parameterize the primary dihedrals of the sulfoximine **4** inhibitor. Developing a model with the sulfoximine **4** bound to AS-B would yield a stepwise picture of the enzyme as the reaction it catalyzes progresses from bound substrates (Gln/Asp/ATP) to formations of the intermediate (β AspAMP **1**) to tetrahedral transition-state (sulfoximine **4** mimics the transition-state) before the release of the final products. These models could then be used in much the same way a series of x-ray crystal structures would be in showing the progression of catalysis and the critical residues and potential conformational changes involved. Fully refined parameters for any of the inhibitors would also allow for an additional level of model validation, as a free energy perturbation (FEP) calculation could be run for comparison with the experimental binding constant for the particular inhibitor.

Modeling the Residue Mutations

The introduction of residue mutations into the AS-B model is another avenue that should be explored in the future of this project. Several mutations have been made in the AS-B enzyme that have aided in explaining the mechanism of catalysis. Mutations within the AS-B model could also help explain exactly what takes place in the enzyme as a result of the point residue mutations. Any mutation could result in a breakdown of the structure of AS-B or simply a breakdown of the interaction between AS-B and the substrates or intermediate during catalysis. Table 5-3 lists important residue mutations that have been experimentally tested thus far.

Table 5-3. AS-B synthetase domain mutants with the putative relevance

Residue	Mutations Tested	Relevance
Cys-523	Ala ²¹⁴	Aspartate binding
Arg-325	Ala & Lys ²¹⁵	β AspAMP formation
Thr-322	Ala, Ser, Val & Tyr ²¹⁵	Aspartate binding or β AspAMP formation
Thr-323	Ala, Ile, Leu, Ser & Val ²¹⁵	Aspartate binding or β AspAMP formation
Ser-346	Ala & Thr ¹⁴⁷	Structurally important
Glu-348	Ala, Asp & Gln ¹⁴⁷	β AspAMP formation or active site coordination
Glu-352	Ala, Asp & Gln ¹⁴⁷	Stabalizing the aspartate α -amino group
Asp-384	Ala & Asn ¹⁴⁷	Catalysis
Arg-387	Ala & Lys ¹⁴⁷	Catalysis
Lys-449	Ala & Arg ¹⁴⁵	ATP binding or catalysis

Without structural validation, it is difficult to prove unequivocally the specific role of any single residue. The mutations shown in Table 5-3 have not had their associated relevance proven. Glu-348 is the most well-characterized, as Dr. Jemy Gutierrez was able to validate the importance of this residue through ¹⁸O-labeled transfer studies and ³¹P NMR assays. Molecular dynamics simulations of single mutations of the model would allow for a more definitive structural picture of how these mutations affect the structure of AS-B. They would also shed light on the how these mutations affect the interaction of the protein with the β AspAMP or aspartate, as many of these residues are believed to be important in substrate binding or intermediate formation.

APPENDIX
CHARMM SIMULATED ANNEALING EXAMPLE INPUT FILES

Heating Input File

```
* GENERATED BY CHARMM-GUI (http://www.charmm-gui.org) on Apr, 24, 2009.
* INPUT FILE FOR EQUILIBRATION OF SOLVATED GLOBULAR PROTEIN
*

!Read topology
open read unit 10 card name top_all27_prot_na.rtf
read rtf unit 10 card

!Read parameters
open read unit 20 card name par_all27_prot_na.prm
read para unit 20 card

!Stream in BAA top/par
stream baa_stream.str

!Read PSF and Coordinates
open read unit 10 card name baa_model_mini_shake.psf
read psf unit 10 card

open read unit 10 card name baa_model_mini_shake.crd
read coor unit 10 card

!
! Setup PBC (Periodic Boundary Condition)
!

stream step3_pbcsetup.str

open read unit 10 card name crystal_image.str
CRYSTAL DEFINE @XTLtype @A @B @C @alpha @beta @gamma
CRYSTAL READ UNIT 10 CARD

!Image centering by residue
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele resname TIP3 end
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele ( segid @posid .or. segid @negid ) end

!
! Nonbonded Options
!

nbonds atom vatom vswitch bycb -
  ctonnb 10.0 ctofnb 12.0 cutnb 16.0 cutim 16.0 -
  inbfrq -1 imgfrq -1 wmin 1.0 cdie eps 1.0 -
  ewald pmew fftx @fftx ffty @ffty fftz @fftz kappa .34 spline order 6

shake bonH para ! take reference values for bond lengths from param file

!!*****
! heating to 600K
!!*****
```

```
! estimate Pmass from SYSmass (total system mass)
! [there could be problems with extreme values, such as Pmass << SYSmass or Pmass >> SYSmass
scalar mass stat
calc Pmass = int ( ?stot / 50.0 )
```

```
open write card unit 31 name "heat_600k_baa.rst"
open write file unit 32 name "heat_600k_baa.dcd"
```

```
DYNA CPT start time 0.001 nstep 240000 iseed 31415 -
firstt 0.0 finalt 600.0 teminc 5.0 ihtfrq 2000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ichecw 1 twindl -5.0 twindh +5.0 iasors 0 -
isvfrq 500 nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunwrit 31 iuncrd 32 kunit -1 nsavc 100 !ie update nonbond list when any atom has move
```

```
! write out heated coordinates in CHARMM format
open write unit 41 card name "baa_heat_600k.crd"
write coor card unit 41
```

```
! write out heated coordinates in pdb format
open write unit 42 card name "baa_heat_600k.pdb"
write coor pdb unit 42
```

```
STOP
```

Equilibration Input File

```
* GENERATED BY CHARMM-GUI (http://www.charmm-gui.org) on Apr, 24, 2009.
* INPUT FILE FOR EQUILIBRATION OF SOLVATED GLOBULAR PROTEIN
*
```

```
!Read topology
open read unit 10 card name top_all27_prot_na.rtf
read rtf unit 10 card
```

```
!Read parameters
open read unit 20 card name par_all27_prot_na.prm
read para unit 20 card
```

```
!Stream in BAA top/par
stream baa_stream.str
```

```
!Read PSF and Coordinates
open read unit 10 card name baa_model_mini_shake.psf
read psf unit 10 card
```

```
open read unit 10 card name baa_model_mini_shake.crd
read coor unit 10 card
```

```
!
! Setup PBC (Periodic Boundary Condition)
!
```

```
stream step3_pbcsetup.str
```

```

open read unit 10 card name crystal_image.str
CRYSTAL DEFINE @XTLtype @A @B @C @alpha @beta @gamma
CRYSTAL READ UNIT 10 CARD

!Image centering by residue
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele resname TIP3 end
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele ( segid @posid .or. segid @negid ) end

!
! Nonbonded Options
!

nbonds atom vatom vswitch bycb -
  ctonnb 10.0 ctofnb 12.0 cutnb 16.0 cutim 16.0 -
  inbfrq -1 imgfrq -1 wmin 1.0 cdie eps 1.0 -
  ewald pmew fftx @fftx ffty @ffty fftz @fftz kappa .34 spline order 6

! We need to keep the water molecules rigid, and also want to
! keep covalent X-H bonds fixed, so use SHAKE
shake bonH para ! take reference values for bond lengths from param file

!!*****
! equilibrate at 600K
!!*****

! estimate Pmass from SYSmass (total system mass)
! [there could be problems with extreme values, such as Pmass << SYSmass or Pmass >> SYSmass
scalar mass stat
calc Pmass = int ( ?stot / 50.0 )

open read card unit 31 name "heat_600k_baa.rst"
open write card unit 32 name "eq_600k_baa.rst"
open write file unit 33 name "eq_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100000 -
isvfrq 500 nprint 100 iprfrq 500 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 600.0 tmass 2000.0 firstt 600.0 finalt 600.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

! write out equilibrated coordinates in CHARMM format
open write unit 41 card name "baa_equi_600k.crd"
write coor card unit 41

! write out equilibrated coordinates in pdb format
open write unit 42 card name "baa_equi_600k.pdb"
write coor pdb unit 42

STOP

```

Annealing Input File

```

* GENERATED BY CHARMM-GUI (http://www.charmm-gui.org) on Apr, 24, 2009.
* INPUT FILE FOR EQUILIBRATION OF SOLVATED GLOBULAR PROTEIN
*

```

```

!Read topology
open read unit 10 card name top_all27_prot_na.rtf
read rtf unit 10 card

!Read parameters
open read unit 20 card name par_all27_prot_na.prm
read para unit 20 card

!Stream in BAA top/par
stream baa_stream.str

!Read PSF and Coordinates
open read unit 10 card name baa_model_mini_shake.psf
read psf unit 10 card

open read unit 10 card name baa_model_mini_shake.crd
read coor unit 10 card

!
! Setup PBC (Periodic Boundary Condition)
!

stream step3_pbcsetup.str

open read unit 10 card name crystal_image.str
CRYSTAL DEFINE @XTLtype @A @B @C @alpha @beta @gamma
CRYSTAL READ UNIT 10 CARD

!Image centering by residue
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele resname TIP3 end
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele ( segid @posid .or. segid @negid ) end

!
! Nonbonded Options
!

nbonds atom vatom vswitch bycb -
  ctonnb 10.0 ctofnb 12.0 cutnb 16.0 cutim 16.0 -
  inbfrq -1 imgfrq -1 wmin 1.0 cdie eps 1.0 -
  ewald pmew fftx @fftx ffty @ffty ffitz @fftz kappa .34 spline order 6

! We need to keep the water molecules rigid, and also want to
! keep covalent X-H bonds fixed, so use SHAKE
shake bonH para ! take reference values for bond lengths from param file

! estimate Pmass from SYSmass (total system mass)
! [there could be problems with extreme values, such as Pmass << SYSmass or Pmass >> SYSmass
scalar mass stat
calc Pmass = int ( ?stot / 50.0 )

!!*****
! Annealing down to 300K & 10 Quenching steps
!!*****

!Step1

```

```
open read card unit 31 name "eq_600k_baa.rst"
open write card unit 32 name "run1_600k_baa.rst"
open write file unit 33 name "run1_600k_baa.dcd"
```

```
DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfreq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 600.0 tmass 2000.0 firstt 600.0 finalt 600.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iunercd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move
```

```
open read card unit 31 name "run1_600k_baa.rst"
open write card unit 32 name "ann1_600k_baa.rst"
open write file unit 33 name "ann1_600k_baa.dcd"
open write file unit 34 name "ann1_600k_baa.dvl"
open write card unit 35 name "ann1_600k_baa.ene"
```

```
DYNA CPT start time 0.001 nstep 50000 -
firstt 600.0 finalt 570.0 teminc -0.6 tstruc 600.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfreq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iunercd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move
```

```
! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final1.crd"
write coor card unit 41
```

```
! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final1.pdb"
write coor pdb unit 42
```

!Step2

```
open read card unit 31 name "ann1_600k_baa.rst"
open write card unit 32 name "run2_600k_baa.rst"
open write file unit 33 name "run2_600k_baa.dcd"
```

```
DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfreq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 570.0 tmass 2000.0 firstt 570.0 finalt 570.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iunercd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move
```

```
open read card unit 31 name "run2_600k_baa.rst"
open write card unit 32 name "ann2_600k_baa.rst"
open write file unit 33 name "ann2_600k_baa.dcd"
open write file unit 34 name "ann2_600k_baa.dvl"
open write card unit 35 name "ann2_600k_baa.ene"
```

```
DYNA CPT start time 0.001 nstep 50000 -
```

```
firstt 570.0 finalt 540.0 teminc -0.6 tstruc 570.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move
```

```
! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final2.crd"
write coor card unit 41
```

```
! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final2.pdb"
write coor pdb unit 42
```

!Step3

```
open read card unit 31 name "ann2_600k_baa.rst"
open write card unit 32 name "run3_600k_baa.rst"
open write file unit 33 name "run3_600k_baa.dcd"
```

```
DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 540.0 tmass 2000.0 firstt 540.0 finalt 540.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move
```

```
open read card unit 31 name "run3_600k_baa.rst"
open write card unit 32 name "ann3_600k_baa.rst"
open write file unit 33 name "ann3_600k_baa.dcd"
open write file unit 34 name "ann3_600k_baa.dvl"
open write card unit 35 name "ann3_600k_baa.ene"
```

```
DYNA CPT start time 0.001 nstep 50000 -
firstt 540.0 finalt 510.0 teminc -0.6 tstruc 540.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move
```

```
! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final3.crd"
write coor card unit 41
```

```
! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final3.pdb"
write coor pdb unit 42
```

!Step4

```
open read card unit 31 name "ann3_600k_baa.rst"
open write card unit 32 name "run4_600k_baa.rst"
open write file unit 33 name "run4_600k_baa.dcd"
```

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 510.0 tmass 2000.0 firstt 510.0 finalt 510.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 -1 -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

open read card unit 31 name "run4_600k_baa.rst"
open write card unit 32 name "ann4_600k_baa.rst"
open write file unit 33 name "ann4_600k_baa.dcd"
open write file unit 34 name "ann4_600k_baa.dvl"
open write card unit 35 name "ann4_600k_baa.ene"

DYNA CPT start time 0.001 nstep 50000 -
firstt 510.0 finalt 480.0 teminc -0.6 tstruc 510.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 -1 -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move

! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final4.crd"
write coor card unit 41

! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final4.pdb"
write coor pdb unit 42

!Step5

open read card unit 31 name "ann4_600k_baa.rst"
open write card unit 32 name "run5_600k_baa.rst"
open write file unit 33 name "run5_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 480.0 tmass 2000.0 firstt 480.0 finalt 480.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 -1 -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

open read card unit 31 name "run5_600k_baa.rst"
open write card unit 32 name "ann5_600k_baa.rst"
open write file unit 33 name "ann5_600k_baa.dcd"
open write file unit 34 name "ann5_600k_baa.dvl"
open write card unit 35 name "ann5_600k_baa.ene"

DYNA CPT start time 0.001 nstep 50000 -
firstt 480.0 finalt 450.0 teminc -0.6 tstruc 480.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -

inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iunerd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move

! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final5.crd"
write coor card unit 41

! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final5.pdb"
write coor pdb unit 42

!Step6

open read card unit 31 name "ann5_600k_baa.rst"
open write card unit 32 name "run6_600k_baa.rst"
open write file unit 33 name "run6_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 450.0 tmass 2000.0 firstt 450.0 finalt 450.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iunerd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

open read card unit 31 name "run6_600k_baa.rst"
open write card unit 32 name "ann6_600k_baa.rst"
open write file unit 33 name "ann6_600k_baa.dcd"
open write file unit 34 name "ann6_600k_baa.dvl"
open write card unit 35 name "ann6_600k_baa.ene"

DYNA CPT start time 0.001 nstep 50000 -
firstt 450.0 finalt 420.0 teminc -0.6 tstruc 450.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iunerd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move

! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final6.crd"
write coor card unit 41

! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final6.pdb"
write coor pdb unit 42

!Step7

open read card unit 31 name "ann6_600k_baa.rst"
open write card unit 32 name "run7_600k_baa.rst"
open write file unit 33 name "run7_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -

HOOVER reft 420.0 tmass 2000.0 firstt 420.0 finalt 420.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

open read card unit 31 name "run7_600k_baa.rst"
open write card unit 32 name "ann7_600k_baa.rst"
open write file unit 33 name "ann7_600k_baa.dcd"
open write file unit 34 name "ann7_600k_baa.dvl"
open write card unit 35 name "ann7_600k_baa.ene"

DYNA CPT start time 0.001 nstep 50000 -
firstt 420.0 finalt 390.0 teminc -0.6 tstruc 420.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move

! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final7.crd"
write coor card unit 41

! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final7.pdb"
write coor pdb unit 42

!Step8

open read card unit 31 name "ann7_600k_baa.rst"
open write card unit 32 name "run8_600k_baa.rst"
open write file unit 33 name "run8_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 390.0 tmass 2000.0 firstt 390.0 finalt 390.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

open read card unit 31 name "run8_600k_baa.rst"
open write card unit 32 name "ann8_600k_baa.rst"
open write file unit 33 name "ann8_600k_baa.dcd"
open write file unit 34 name "ann8_600k_baa.dvl"
open write card unit 35 name "ann8_600k_baa.ene"

DYNA CPT start time 0.001 nstep 50000 -
firstt 390.0 finalt 360.0 teminc -0.6 tstruc 390.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move

! write out Annealed coordinates in CHARMM format

open write unit 41 card name "baa_ann-final8.crd"
write coor card unit 41

! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final8.pdb"
write coor pdb unit 42

!Step9

open read card unit 31 name "ann8_600k_baa.rst"
open write card unit 32 name "run9_600k_baa.rst"
open write file unit 33 name "run9_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 360.0 tmass 2000.0 firstt 360.0 finalt 360.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 -! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

open read card unit 31 name "run9_600k_baa.rst"
open write card unit 32 name "ann9_600k_baa.rst"
open write file unit 33 name "ann9_600k_baa.dcd"
open write file unit 34 name "ann9_600k_baa.dvl"
open write card unit 35 name "ann9_600k_baa.ene"

DYNA CPT start time 0.001 nstep 50000 -
firstt 360.0 finalt 330.0 teminc -0.6 tstruc 360.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 -! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move

! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final9.crd"
write coor card unit 41

! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final9.pdb"
write coor pdb unit 42

!Step10

open read card unit 31 name "ann9_600k_baa.rst"
open write card unit 32 name "run10_600k_baa.rst"
open write file unit 33 name "run10_600k_baa.dcd"

DYNA CPT restart time 0.001 nstep 100 -
isvfrq 100 nprint 100 iprfrq 100 ntrfrq 100 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
HOOVER reft 330.0 tmass 2000.0 firstt 330.0 finalt 330.0 -
ieqfrq 0 ichew 0 twindl -5.0 twindh +5.0 iasors 0 -
inbfrq -1 ihbfrq 0 imgfrq -1 -! -1 for inbfrq/imgfrq means heuristic update
iunread 31 iunwrit 32 iuncrd 33 kunit -1 nsavc 100 !ie update nonbond list when any atom has move

```
open read card unit 31 name "run10_600k_baa.rst"
open write card unit 32 name "ann10_600k_baa.rst"
open write file unit 33 name "ann10_600k_baa.dcd"
open write file unit 34 name "ann10_600k_baa.dvl"
open write card unit 35 name "ann10_600k_baa.ene"
```

```
DYNA CPT start time 0.001 nstep 50000 -
firstt 330.0 finalt 300.0 teminc -0.6 tstruc 330.0 ihtfrq 1000 -
PCONst pref 1.0 pmass @Pmass pgamma 20.0 -
ieqfrq 0 ichecw 0 twindl -5.0 twindh +5.0 iasors 0 -
nprint 100 iprfrq 500 ntrfrq 100 -
inbfrq -1 ihbfrq 0 imgfrq -1 - ! -1 for inbfrq/imgfrq means heuristic update
iunrea 31 iunwrit 32 iuncrd 33 iunvel 34 kunit 35 nsavc 100 !ie update nonbond list when any atom has move
```

```
! write out Annealed coordinates in CHARMM format
open write unit 41 card name "baa_ann-final10.crd"
write coor card unit 41
```

```
! write out Annealed coordinates in pdb format
open write unit 42 card name "baa_ann-final10.pdb"
write coor pdb unit 42
```

STOP

Final Minimization Input File

```
* GENERATED BY CHARMM-GUI (http://www.charmm-gui.org) on Apr, 24, 2009.
* INPUT FILE FOR EQUILIBRATION OF SOLVATED GLOBULAR PROTEIN
*
```

```
!Read topology
open read unit 10 card name top_all27_prot_na.rtf
read rtf unit 10 card
```

```
!Read parameters
open read unit 20 card name par_all27_prot_na.prm
read para unit 20 card
```

```
!Stream in BAA top/par
stream baa_stream.str
```

```
open read card unit 30 name baa_model_mini_shake.psf
read psf card unit 30
```

```
open read card unit 10 name baa_ann-final10.pdb
read coor pdb unit 10 resid
```

```
!
! Setup PBC (Periodic Boundary Condition)
!
```

```
stream step3_pbcsetup.str
```

```
open read unit 10 card name crystal_image.str
CRYSTAL DEFINE @XTLtype @A @B @C @alpha @beta @gamma
CRYSTAL READ UNIT 10 CARD
```

```

!Image centering by residue
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele resname TIP3 end
IMAGE BYRESID XCEN 0.0 YCEN 0.0 ZCEN 0.0 sele (segid @posid .or. segid @negid ) end

!
! Nonbonded Options
!

nbonds atom vatom vswitch bycb -
  ctonnb 10.0 ctofnb 12.0 cutnb 16.0 cutim 16.0 -
  inbfrq -1 imgfrq -1 wmin 1.0 cdie eps 1.0 -
  ewald pmew fftx @fftx ffty @ffty ffitz @fftz kappa .34 spline order 6

energy

coor copy comp

! We need to keep the water molecules rigid, and also want to
! keep covalent X-H bonds fixed, so use SHAKE
shake bonH para ! take reference values for bond lengths from param file

!!!! MINIMIZATION
! allow water to adjust around protein
cons harm force 50.0 sele segid asb end
mini SD nstep 1000 tolgrd 0.01
cons harm force 20.0 sele segid asb end
mini CONJ nstep 2500 tolgrd 0.01
cons harm force 10.0 sele segid asb end
mini CONJ nstep 5000 tolgrd 0.01
cons harm force 5.0 sele segid asb end
mini CONJ nstep 10000 tolgrd 0.001
cons harm force 0.0 sele segid asb end
mini ABNR nstep 100000 tolgrd 0.00001

coor orient rms mass

! write out heated coordinates in CHARMM format
open write unit 41 card name "baa_ann-final10_mini.crd"
write coor card unit 41

! write out heated coordinates in pdb format
open write unit 42 card name "baa_ann-final10_mini.pdb"
write coor pdb unit 42

STOP

```

LIST OF REFERENCES

- (1) "Cancer Facts and Figures, 2009," American Cancer Society, 2009.
- (2) Cooney, D. A.; Handschumacher, R. E. *Annu. Rev. Pharmacol.* **1970**, *10*, 421.
- (3) Swain, A. L.; Jaskolski, M.; Housset, D.; Rao, J. K. M.; Wlodawer, A. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 1474.
- (4) Aghaiypour, K.; Wlodawer, A.; Lubkowski, J. *Biochemistry* **2001**, *40*, 5655.
- (5) Tallal, L.; Tan, C.; Oettgen, H.; Wollner, N.; McCarthy, M.; Helson, L.; Burchenal, J.; Karnofsky, D.; Murphy, M. L. *Cancer* **1970**, *25*, 306.
- (6) Sutow, W. W.; Garcia, F.; Starling, K. A.; Williams, T. E.; Lane, D. M.; Gehan, E. A. *Cancer* **1971**, *28*, 819.
- (7) Terebelo, H. R.; Anderson, K.; Wiernik, P. H.; Cuttner, J.; Cooper, R. M.; Faso, L.; Berenberg, J. L. *Am. J. Clin. Oncol.* **1986**, *9*, 411.
- (8) Barr, R. D.; DeVeber, L. L.; Pai, K. M.; Andrew, M.; Halton, J.; Cairney, A. E.; Whitton, A. C. *Am. J. Pediatr. Hematol. Oncol.* **1992**, *14*, 136.
- (9) Kiriyaama, Y.; Kubota, M.; Takimoto, T.; Kitoh, T.; Tanizawa, A.; Akiyama, Y.; Mikawa, H. *Leukemia* **1989**, *3*, 294.
- (10) Lobel, J. S.; O'Brien, R. T.; McIntosh, S.; Aspnes, G. T.; Capizzi, R. L. *Cancer* **1979**, *43*, 1089.
- (11) Capizzi, R. L.; Bertino, J. R.; Skeel, R. T.; Creasey, W. A.; Zanes, R.; Olayon, C.; Peterson, R. G.; Handschumacher, R. E. *Ann. Intern. Med.* **1971**, *74*, 893.
- (12) Chakrabarti, R.; Schuster, S. M. *Int. J. Pediatr. Hematol. Oncol.* **1997**, *4*, 597.
- (13) Pieters, R.; Klumper, E.; Kaspers, G. J.; Veerman, A. J. *Crit. Rev. Oncol. Hematol.* **1997**, *25*, 11.
- (14) Haskell, C. M.; Canellos, G. P. *Biochem. Pharmacol.* **1969**, *18*, 2578.
- (15) Martin, J. K.; Sun, W.; Moraga-Amador, D.; Schuster, S. M.; Wylie, D. E. *Amino Acids* **1993**, *5*, 51.
- (16) Hutson, R. G.; Kitoh, T.; Amador, D. A. M.; Cosic, S.; Schuster, S. M.; Kilberg, M. S. *Am. J. Physiol.* **1997**, *272*, C1691.
- (17) Aslanian, A. M.; Fletcher, B. S.; Kilberg, M. S. *Biochem. J.* **2001**, *357*, 321.

- (18) Aslanian, A. M.; Kilberg, M. S. *Biochem. J.* **2001**, 358, 59.
- (19) Minowada, J.; Onuma, T.; Moore, G. E. *J. Natl. Cancer Inst.* **1972**, 49, 891.
- (20) Verma, N.; Kumar, K.; Kaur, G.; Anand, S. *Crit. Rev. Biotechnol.* **2007**, 27, 45.
- (21) Richards, N. G. J.; Kilberg, M. S. *Annu. Rev. Biochem.* **2006**, 75, 629.
- (22) Su, N.; Pan, Y.-X.; Zhou, M.; Harvey Richard, C.; Hunger Stephen, P.; Kilberg Michael, S. *Pediatr. Blood Cancer* **2008**, 50, 274.
- (23) Gutierrez, J. A.; Pan, Y.-X.; Koroniak, L.; Hiratake, J.; Kilberg, M. S.; Richards, N. G. J. *Chem. Biol.* **2006**, 13, 1339.
- (24) Milman, H. A.; Cooney, D. A. *Biochem. J.* **1979**, 181, 51.
- (25) Hinchman, S. K.; Schuster, S. M. *Protein Eng.* **1992**, 5, 279.
- (26) Cedar, H.; Schwartz, J. H. *J. Biol. Chem.* **1969**, 244, 4112.
- (27) Cedar, H.; Schwartz, J. H. *J. Biol. Chem.* **1969**, 244, 4122.
- (28) Scofield, M. A.; Lewis, W. S.; Schuster, S. M. *J. Biol. Chem.* **1990**, 265, 12895.
- (29) Horowitz, B.; Meister, A. *J. Biol. Chem.* **1972**, 247, 6708.
- (30) Patterson, M. K., Jr.; Orr, G. R. *J. Biol. Chem.* **1968**, 243, 376.
- (31) Boehlein, S. K.; Richards, N. G. J.; Schuster, S. M. *J. Biol. Chem.* **1994**, 269, 7450.
- (32) Zalkin, H.; Smith, J. L. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, 72, 87.
- (33) Smith, J. L. *Biochem. Soc. Trans.* **1995**, 23, 894.
- (34) Smith, J. L. *Curr. Opin. Struct. Biol.* **1998**, 8, 686.
- (35) Tso, J. Y.; Zalkin, H.; Van Cleemput, M.; Yanofsky, C.; Smith, J. M. *J. Biol. Chem.* **1982**, 257, 3525.
- (36) Badet-Denisot, M. A.; Rene, L.; Badet, B. *Bull. Soc. Chim. Fr.* **1993**, 130, 249.
- (37) Badet, B.; Vermoote, P.; Haumont, P. Y.; Lederer, F.; Le Goffic, F. *Biochemistry* **1987**, 26, 1940.
- (38) Vanoni, M. A.; Curti, B. *Cell. Mol. Life Sci.* **1999**, 55, 617.

- (39) Vanoni, M. A.; Edmondson, D. E.; Rescigno, M.; Zanetti, G.; Curti, B. *Biochemistry* **1991**, *30*, 11478.
- (40) van den Heuvel Robert, H. H.; Ferrari, D.; Bossi Roberto, T.; Ravasio, S.; Curti, B.; Vanoni Maria, A.; Florencio Francisco, J.; Mattevi, A. *J. Biol. Chem.* **2002**, *277*, 24579.
- (41) Boehlein, S. K.; Schuster, S. M.; Richards, N. G. J. *Biochemistry* **1996**, *35*, 3031.
- (42) Andrulis, I. L.; Chen, J.; Ray, P. N. *Mol. Cell. Biol.* **1987**, *7*, 2435.
- (43) Hongo, S.; Sato, T. *Biochimica et Biophysica Acta, Protein Structure and Molecular Enzymology* **1983**, *742*, 484.
- (44) Luehr, C. A.; Schuster, S. M. *Archives of Biochemistry and Biophysics* **1985**, *237*, 335.
- (45) Dang, V.-D.; Valens, M.; Bolotin-Fukuhara, M.; Daignan-Fornier, B. *Mol. Microbiol.* **1996**, *22*, 681.
- (46) Reitzer, L. J.; Magasanik, B. *J. Bacteriol.* **1982**, *151*, 1299.
- (47) Mehlhaff, P. M.; Schuster, S. M. *Arch. Biochem. Biophys.* **1991**, *284*, 143.
- (48) Mehlhaff, P. M.; Luehr, C. A.; Schuster, S. M. *Biochemistry* **1985**, *24*, 1104.
- (49) Markin, R. S.; Luehr, C. A.; Schuster, S. M. *Biochemistry* **1981**, *20*, 7226.
- (50) Ciustea, M.; Gutierrez, J. A.; Abbatiello, S. E.; Eyler, J. R.; Richards, N. G. J. *Arch. Biochem. Biophys.* **2005**, *440*, 18.
- (51) Fresquet, V.; Thoden, J. B.; Holden, H. M.; Raushel, F. M. *Bioorg. Chem.* **2004**, *32*, 63.
- (52) Tesson, A. R.; Soper, T. S.; Ciustea, M.; Richards, N. G. J. *Arch. Biochem. Biophys.* **2003**, *413*, 23.
- (53) Boehlein, S. K.; Stewart, J. D.; Walworth, E. S.; Thirumoorthy, R.; Richards, N. G. J.; Schuster, S. M. *Biochemistry* **1998**, *37*, 13230.
- (54) Rognes, S. E. *Phytochemistry (Elsevier)* **1975**, *14*, 1975.
- (55) Hongo, S.; Sato, T. *Arch. Biochem. Biophys.* **1985**, *238*, 410.
- (56) Li, K. K.; Beeson, W. T.; Ghiviriga, I.; Richards, N. G. J. *Biochemistry* **2007**, *46*, 4840.
- (57) Larsen, T. M.; Boehlein, S. K.; Schuster, S. M.; Richards, N. G. J.; Thoden, J. B.; Holden, H. M.; Rayment, I. *Biochemistry* **1999**, *38*, 16146.

- (58) Richards, N. G. J.; Schuster, S. M. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, 72, 145.
- (59) Bork, P.; Koonin, E. V. *Proteins: Struct., Funct., Genet.* **1994**, 20, 347.
- (60) Hinchman, S. K.; Henikoff, S.; Schuster, S. M. *J. Biol. Chem.* **1992**, 267, 144.
- (61) Nakatsu, T.; Kato, H.; Oda, J. i. *Nat. Struct. Biol.* **1998**, 5, 15.
- (62) Krahn, J. M.; Kim, J. H.; Burns, M. R.; Parry, R. J.; Zalkin, H.; Smith, J. L. *Biochemistry* **1997**, 36, 11061.
- (63) Muchmore, C. R. A.; Krahn, J. M.; Kim, J. H.; Zalkin, H.; Smith, J. L. *Protein Sci.* **1998**, 7, 39.
- (64) Chen, S.; Tomchick, D. R.; Wolle, D.; Hu, P.; Smith, J. L.; Switzer, R. L.; Zalkin, H. *Biochemistry* **1997**, 36, 10718.
- (65) Kim, J. H.; Krahn, J. M.; Tomchick, D. R.; Smith, J. L.; Zalkin, H. *J. Biol. Chem.* **1996**, 271, 15549.
- (66) Smith, J. L.; Zaluzec, E. J.; Wery, J. P.; Niu, L.; Switzer, R. L.; Zalkin, H.; Satow, Y. *Science (Washington, DC, U. S.)* **1994**, 264, 1427.
- (67) Tesmer, J. J. G.; Klem, T. J.; Deras, M. L.; Davisson, V. J.; Smith, J. L. *Nat. Struct. Biol.* **1996**, 3, 74.
- (68) Karlberg, T.; Collins, R.; van den Berg, S.; Flores, A.; Hammarstroem, M.; Hoegbom, M.; Holmberg Schiavone, L.; Uppenberg, J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, D64, 279.
- (69) Lemke, C. T.; Howell, P. L. *Structure (Cambridge, MA, U. S.)* **2001**, 9, 1153.
- (70) Mougous, J. D.; Lee, D. H.; Hubbard, S. C.; Schelle, M. W.; Vocadlo, D. J.; Berger, J. M.; Bertozzi, C. R. *Mol. Cell* **2006**, 21, 109.
- (71) Sun, M.; Leyh, T. S. *Biochemistry* **2006**, 45, 11304.
- (72) Miller, M. T.; Bachmann, B. O.; Townsend, C. A.; Rosenzweig, A. C. *Nat. Struct. Biol.* **2001**, 8, 684.
- (73) Bachmann, B. O.; Townsend, C. A. *Biochemistry* **2000**, 39, 11187.
- (74) McNaughton, H. J.; Thirkettle, J. E.; Zhang, Z.; Schofield, C. J.; Jensen, S. E.; Barton, B.; Greaves, P. *Chemical Communications (Cambridge)* **1998**, 2325.

- (75) Waterman, D. G.; Ortiz-Lombardia, M.; Fogg, M. J.; Koonin, E. V.; Antson, A. A. *J. Mol. Biol.* **2006**, *356*, 97.
- (76) Mueller, E. G.; Palenchar, P. M. *Protein Sci.* **1999**, *8*, 2424.
- (77) Denessiouk, K. A.; Johnson, M. S. *Proteins: Struct., Funct., Genet.* **2000**, *38*, 310.
- (78) Ding, Y. Experimental and Theoretical Characterization of Asparagine Synthetase Inhibitors., University of Florida, 2002.
- (79) Hurd, R. E.; John, B. K. *J. Magn. Reson.* **1991**, *91*, 648.
- (80) Schnizer, H. G.; Boehlein, S. K.; Stewart, J. D.; Richards, N. G. J.; Schuster, S. M. *Methods Enzymol.* **2002**, *354*, 260.
- (81) Schnizer, H. G.; Boehlein, S. K.; Stewart, J. D.; Richards, N. G. J.; Schuster, S. M. *Biochemistry* **1999**, *38*, 3677.
- (82) Miles, B. W.; Raushel, F. M. *Biochemistry* **2000**, *39*, 5051.
- (83) Myers, R. S.; Amaro, R. E.; Luthey-Schulten, Z. A.; Davisson, V. J. *Biochemistry* **2005**, *44*, 11974.
- (84) Yount, R. G. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 1.
- (85) Mouilleron, S.; Golinelli-Pimpaneau, B. *Curr. Opin. Struct. Biol.* **2007**, *17*, 653.
- (86) Mouilleron, S.; Badet-Denisot, M.-A.; Golinelli-Pimpaneau, B. *J. Biol. Chem.* **2006**, *281*, 4404.
- (87) van den Heuvel, R. H. H.; Svergun, D. I.; Petoukhov, M. V.; Coda, A.; Curti, B.; Ravasio, S.; Vanoni, M. A.; Mattevi, A. *J. Mol. Biol.* **2003**, *330*, 113.
- (88) Boehlein, S. K.; Rosa-Rodriguez, J. G.; Schuster, S. M.; Richards, N. G. J. *J. Am. Chem. Soc.* **1997**, *119*, 5785.
- (89) MacKerel, A. D., Jr.; Brooks Iii, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; John Wiley & Sons: Chichester, 1998; Vol. 1; pp 271.
- (90) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (91) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208.
- (92) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459.

- (93) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London, U. K.)* **1977**, *267*, 585.
- (94) Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335.
- (95) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
- (96) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
- (97) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
- (98) Vinter, J. G.; Davis, A.; Saunders, M. R. *J. Comput. Aided Mol. Des.* **1987**, *1*, 31.
- (99) Daura, X.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535.
- (100) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
- (101) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
- (102) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
- (103) Clark, M.; Cramer, R. D., III.; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982.
- (104) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (105) Schlenkrich, M.; Brickmann, J.; Mackerell, A. D., Jr.; Karplus, M. *Biological Membranes* **1996**, 31.
- (106) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
- (107) Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comput. Chem.* **1996**, *17*, 450.
- (108) Schmidt, J. M.; Brueschweiler, R.; Ernst, R. R.; Dunbrack, R. L., Jr.; Joseph, D.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 8747.
- (109) Feng, M.-H.; Philippopoulos, M.; MacKerell, A. D., Jr.; Lim, C. *J. Am. Chem. Soc.* **1996**, *118*, 11265.

- (110) Lazaridis, T.; Karplus, M. *Science (Washington, D. C.)* **1997**, 278, 1928.
- (111) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, 105, 1902.
- (112) Reiher, W. E., III. *Theoretical Studies of Hydrogen Bonding*, Harvard University, 1985.
- (113) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, 79, 926.
- (114) Gaussian 03, R. C., Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.
- (115) Pavelites, J. J.; Gao, J.; Bash, P. A.; Mackerell, A. D., Jr. *J. Comput. Chem.* **1997**, 18, 221.
- (116) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, 11, 361.
- (117) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, 100, 16502.
- (118) Vaiana, A. C.; Cournia, Z.; Costescu, I. B.; Smith, J. C. *Comput. Phys. Commun.* **2005**, 167, 34.
- (119) Fiser, A.; Do, R. K. G.; Sali, A. *Protein Sci.* **2000**, 9, 1753.
- (120) Fagerberg, T.; Cerottini, J.-C.; Michielin, O. *J. Mol. Biol.* **2006**, 356, 521.
- (121) Doll, K.; Schoen, J. C.; Jansen, M. *Journal of Physics: Conference Series* **2008**, 117, No pp given.
- (122) Fernandez-Lima, F. A.; Wei, H.; Gao, Y. Q.; Russell, D. H. *J. Phys. Chem. A* **2009**, 113, 8221.
- (123) Villani, V. *J. Mol. Struct.* **2004**, 672, 17.

- (124) Hixson, C. A.; Wheeler, R. A. *J. Chem. Theory Comput.* **2009**, *5*, 1883.
- (125) Collura, V.; Higo, J.; Garnier, J. *Protein Sci.* **1993**, *2*, 1502.
- (126) Rapp, C. S.; Friesner, R. A. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 173.
- (127) Rognan, D.; Zimmermann, N.; Jung, G.; Folkers, G. *Eur. J. Biochem.* **1992**, *208*, 101.
- (128) Ota, N.; Agard, D. A. *J. Mol. Biol.* **2001**, *314*, 607.
- (129) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1990**, *29*, 1847.
- (130) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377.
- (131) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
- (132) Černý, V. *Journal of Optimization Theory and Applications* **1985**, *45*, 41.
- (133) Leach, A. R. *Reviews in Computational Chemistry* **1991**, *2*, 1.
- (134) Geman, S.; Geman, D. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **1984**, *6*, 721.
- (135) Steinbach, P. J.; Brooks, B. R. *Chem. Phys. Lett.* **1994**, *226*, 447.
- (136) Gordon, H. L.; Kwan, W. K.; Gong, C.; Larrass, S.; Rothstein, S. M. *J. Chem. Phys.* **2003**, *118*, 1533.
- (137) Cheatham, T. E., III.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.
- (138) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- (139) Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. *J. Phys. Chem.* **1996**, *100*, 17011.
- (140) Rapaport, D. C. *The Art of Molecular Dynamics Simulation*; Cambridge University Press: Cambridge, 1995.
- (141) Cramer, C. J. *Essentials of Computational Chemistry*; John Wiley & Sons: New York, 2002.
- (142) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.

- (143) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (144) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (145) Ikeuchi, H.; Meyer, M. E.; Ding, Y.; Hiratake, J.; Richards, N. G. *J. Bioorg. Med. Chem., In Press, Corrected Proof*.
- (146) DeLano, W. L. The PyMOL Molecular Graphics System.; DeLano Scientific LLC: Palo Alto, CA, USA., 2008.
- (147) Gutierrez, J. A. Inhibition and Functional Characterization of Asparagine Synthetase, University of Florida, 2006.
- (148) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. *J. Med. Chem.* **2005**, *48*, 962.
- (149) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935.
- (150) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411.
- (151) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
- (152) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
- (153) Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488.
- (154) Abagyan, R.; Totrov, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375.
- (155) Welch, W.; Ruppert, J.; Jain, A. N. *Chem. Biol.* **1996**, *3*, 449.
- (156) McMartin, C.; Bohacek, R. S. *J. Comput. Aided Mol. Des.* **1997**, *11*, 333.
- (157) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. *J. Comput. Aided Mol. Des.* **1999**, *13*, 547.
- (158) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. *Biopolymers* **2003**, *68*, 76.
- (159) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graphics Modell.* **2003**, *21*, 289.
- (160) Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759.
- (161) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999**, *42*, 5100.

- (162) Stahl, M.; Rarey, M. *J. Med. Chem.* **2001**, *44*, 1035.
- (163) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., III. *J. Comput. Aided Mol. Des.* **2003**, *17*, 755.
- (164) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151.
- (165) 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. *J. Med. Chem.* **2004**, *47*, 1739.
- (166) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.
- (167) Tominaga, Y.; Jorgensen, W. L. *J. Med. Chem.* **2004**, *47*, 2534.
- (168) Hann, M. M.; Leach, A. R.; Harper, G. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856.
- (169) Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791.
- (170) Muegge, I. *Perspectives in Drug Discovery and Design* **2000**, *20*, 99.
- (171) Muegge, I. *J. Comput. Chem.* **2001**, *22*, 418.
- (172) Muegge, I. *J. Med. Chem.* **2006**, *49*, 5895.
- (173) Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337.
- (174) DeWitte, R. S.; Shakhnovich, E. I. *J. Am. Chem. Soc.* **1996**, *118*, 11733.
- (175) Purvis, G. D., III. Calculating a potential of mean force (PMF) score of a protein-ligand complex; Fujitsu America, I., Ed. USA, 2005; pp 27.
- (176) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
- (177) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720.
- (178) Brown, D. G.; Visse, R.; Sandhu, G.; Davies, A.; Rizkallah, P. J.; Melitz, C.; Summers, W. C.; Sanderson, M. R. *Nat. Struct. Biol.* **1995**, *2*, 876.
- (179) Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd ed.; Springer-Verlag: New York, 1999.
- (180) Solis, F. J.; Wets, R. J. B. *Mathematics of Operations Research* **1981**, *6*, 19.

- (181) Muegge, I. *Med. Chem. Res.* **1999**, *9*, 490.
- (182) Muegge, I.; Rarey, M. *Reviews in Computational Chemistry* **2001**, *17*, 1.
- (183) Ha, S.; Andreani, R.; Robbins, A.; Muegge, I. *J. Comput. Aided Mol. Des.* **2000**, *14*, 435.
- (184) Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. *J. Med. Chem.* **1999**, *42*, 2498.
- (185) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (186) Bowen, J. P.; Allinger, N. L. *Reviews in Computational Chemistry* **1991**, *2*, 81.
- (187) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (188) Fersht, A. R.; Shi, J. P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature (London, U. K.)* **1985**, *314*, 235.
- (189) Fersht, A. R. *Trends Biochem. Sci.* **1987**, *12*, 301.
- (190) Mills, J. E. J.; Dean, P. M. *J. Comput. Aided Mol. Des.* **1996**, *10*, 607.
- (191) Thornton, J. M.; MacArthur, M. W.; McDonald, I. K.; Jones, D. T.; Mitchell, J. B. O.; Nandi, L.; Price, S. L.; Zvelebil, M. J. J. M. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* **1993**, *345*, 113.
- (192) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 225.
- (193) Xing, L.; Hodgkin, E.; Liu, Q.; Sedlock, D. *J. Comput. Aided Mol. Des.* **2004**, *18*, 333.
- (194) Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287.
- (195) Perola, E.; Walters, W. P.; Charifson, P. S. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 235.
- (196) Jain, A. N. *J. Med. Chem.* **2004**, *47*, 947.
- (197) Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499.
- (198) Verdonk, M. L.; Taylor, R. D.; Chessari, G.; Murray, C. W. Illustration of Current Challenges in Molecular Docking. In *Structure-Based Drug Discovery*, 2007; pp 201.

- (199) Prota, A.; Vogt, J.; Pilger, B.; Perozzo, R.; Wurth, C.; Marquez, V. E.; Russ, P.; Schulz, G. E.; Folkers, G.; Scapozza, L. *Biochemistry* **2000**, *39*, 9597.
- (200) Vogt, J.; Perozzo, R.; Pautsch, A.; Prota, A.; Schelling, P.; Pilger, B.; Folkers, G.; Scapozza, L.; Schulz, G. E. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 545.
- (201) Russ, P.; Schelling, P.; Scapozza, L.; Folkers, G.; De Clercq, E.; Marquez, V. E. *J. Med. Chem.* **2003**, *46*, 5045.
- (202) Spadola, L.; Novellino, E.; Folkers, G.; Scapozza, L. *Eur. J. Med. Chem.* **2003**, *38*, 413.
- (203) Pilger, B. D.; Perozzo, R.; Alber, F.; Wurth, C.; Folkers, G.; Scapozza, L. *J. Biol. Chem.* **1999**, *274*, 31967.
- (204) Balzarini, J.; De Clercq, E.; Baumgartner, H.; Bodenteich, M.; Griengl, H. *Mol. Pharmacol.* **1990**, *37*, 395.
- (205) Schelling, P.; Folkers, G.; Scapozza, L. *Anal. Biochem.* **2001**, *295*, 82.
- (206) Wurth, C.; Kessler, U.; Vogt, J.; Schulz, G. E.; Folkers, G.; Scapozza, L. *Protein Sci.* **2001**, *10*, 63.
- (207) Bennett, M. S.; Wien, F.; Champness, J. N.; Batuwangala, T.; Rutherford, T.; Summers, W. C.; Sun, H.; Wright, G.; Sanderson, M. R. *FEBS Lett.* **1999**, *443*, 121.
- (208) Champness, J. N.; Bennett, M. S.; Wien, F.; Visse, R.; Summers, W. C.; Herdewijn, P.; De Clercq, E.; Ostrowski, T.; Jarvest, R. L.; Sanderson, M. R. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 350.
- (209) Koroniak, L.; Ciustea, M.; Gutierrez, J. A.; Richards, N. G. *J. Org. Lett.* **2003**, *5*, 2033.
- (210) Boehlein, S. K.; Nakatsu, T.; Hiratake, J.; Thirumoorthy, R.; Stewart, J. D.; Richards, N. G. J.; Schuster, S. M. *Biochemistry* **2001**, *40*, 11168.
- (211) Koizumi, M.; Hiratake, J.; Nakatsu, T.; Kato, H.; Oda, J. I. *J. Am. Chem. Soc.* **1999**, *121*, 5799.
- (212) Foote, J.; Lipscomb, W. N. *J. Biol. Chem.* **1981**, *256*, 11428.
- (213) Parmentier, L. E.; Weiss, P. M.; O'Leary, M. H.; Schachman, H. K.; Cleland, W. W. *Biochemistry* **1992**, *31*, 6577.
- (214) Boehlein, S. K.; Walworth, E. S.; Schuster, S. M. *Biochemistry* **1997**, *36*, 10168.
- (215) Boehlein, S. K.; Walworth, E. S.; Richards, N. G. J.; Schuster, S. M. *J. Biol. Chem.* **1997**, *272*, 12384.

BIOGRAPHICAL SKETCH

Robert N. Humkey was born in Lubbock, Texas, and grew up in the small town of Versailles, Kentucky. He spent a very active youth with his parents and his two younger brothers and younger sister. In 1998, Robert began his undergraduate studies in biochemistry at the University of San Diego. During his senior year, an internship with a local pharmaceutical company introduced Robert to a higher level of research-driven chemistry that inspired him to pursue a doctorate in computational biochemistry. In 2002, he began the pursuit of his doctoral degree at the University of Florida, under the guidance of Dr. Nigel G. J. Richards. Robert's graduate research was focused on using computational methods for drug discovery. While in pursuit of his doctorate, Robert also completed an MBA with concentrations in competitive strategy, marketing and finance from the University of Florida's Hough Graduate School of Business.