

5-22-2006

## Thermodynamics of Buried Water in Protein Cavities and Revised Algorithms for Introducing Polarizability to Molecular Dynamics Simulations

Lisa Renee Olano  
*University of New Orleans*

Follow this and additional works at: <https://scholarworks.uno.edu/td>

---

### Recommended Citation

Olano, Lisa Renee, "Thermodynamics of Buried Water in Protein Cavities and Revised Algorithms for Introducing Polarizability to Molecular Dynamics Simulations" (2006). *University of New Orleans Theses and Dissertations*. 387.

<https://scholarworks.uno.edu/td/387>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact [scholarworks@uno.edu](mailto:scholarworks@uno.edu).

THERMODYNAMICS OF BURIED WATER IN PROTEIN CAVITIES AND REVISED  
ALGORITHMS FOR INTRODUCING POLARIZABILITY TO MOLECULAR  
DYNAMICS SIMULATIONS

A Dissertation

Submitted to the Graduate Faculty of the  
University of New Orleans  
in partial fulfillment of the  
requirements for the Degree of

Doctor of Philosophy  
in  
Chemistry

by

Lisa Renee Olano

B.S. Emory University, 1998

M.Ph. The George Washington University, 2001

May 2006

## Dedication

This dissertation is dedicated to my family who always knew I would find something I loved to do despite some detours on the way, and to my friends, graduate school would have been a far less enjoyable experience if you all had not been there.

## Acknowledgments

I would like to thank my research adviser Steven Rick for allowing me the chance to explore a different side of chemistry. I would also like to thank Maureen Shuh for suggesting I would enjoy theoretical chemistry.

I express my gratitude to the entire chemistry faculty for always being willing to answer my questions. Especially the members of my committee, Edwin Stevens, Peter Politzer, Bruce Gibb, and Paul Hanson. Also to Dr. Jane Murray, who always offered support and helpful advise.

To my friends, especially Sandra Lochhead, Edie Banner, Elisha Josepha, Parisa Mahjoor, Patrick Sheridan, Margaret Miller, and Donna Neumann. Thank you for listening to me, whether you understood what I was talking about or not. Whenever I thought it was hopeless you were all there with coffee, lunch, a quick pep talk, and as a last resort chocolate.

I also need to acknowledge the support of my Aunt Barbara, who was always there when I needed her. She was a local lifeline during the course of my degree.

Things don't ever quite work out as planned, for me this manifested in the delay caused by hurricane Katrina. The loss of my possessions and the delay of my defense three days before it was scheduled was (to put it mildly) stressful. However, the situation would have been more difficult had I not had a great group to join at the Scripps Research Institute. I am also grateful for my roommates in California, Nadia and Melania, for giving me support and instant friendship.

# Contents

<b>Abstract</b>	<b>ix</b>
<b>1 Introduction and Background</b>	<b>1</b>
1.1 Proteins and Interior Waters . . . . .	1
1.2 Higher Pressure . . . . .	5
1.3 Simulations . . . . .	7
<b>2 Protein Simulations at 1 bar</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Barnase I76A Mutant . . . . .	19
2.3 Bovine Pancreatic Trypsin Inhibitor (BPTI) . . . . .	32
2.4 Lysozyme I106A Mutant . . . . .	41
2.5 Fluctuating Charge . . . . .	45
2.6 Summary . . . . .	47
<b>3 Protein Simulations at 3 kbar</b>	<b>55</b>
3.1 Introduction . . . . .	55
3.2 Barnase I76A Mutant: 1BRI . . . . .	56
3.3 Bovine Pancreatic Trypsin Inhibitor (BPTI) . . . . .	63
3.4 Lysozyme I106A Mutant . . . . .	73
3.5 Summary . . . . .	74
<b>4 Fluctuating Charge Algorithms</b>	<b>81</b>
4.1 Introduction . . . . .	81
4.2 Methods . . . . .	82
4.3 Fluctuating Charge . . . . .	87
4.4 Summary . . . . .	98
<b>5 Experimental</b>	<b>102</b>
5.1 Protein Simulations . . . . .	102
<b>6 Conclusions</b>	<b>113</b>
6.1 Protein Simulations . . . . .	113
<b>Vita</b>	<b>121</b>

# List of Tables

2.1	Barnase $\Delta G$ at values at 283, 298, 313K and Protein Entropy and Enthalpy Values Numbers in parenthesis represent 95% confidence limits. . . . .	19
2.2	Barnase hydrogen bonds that show large changes in binding at 1 atm . . . .	27
2.3	BPTI $\Delta G$ at values at 283, 298, 313K and Protein Entropy and Enthalpy Values Numbers in parenthesis represent 95% confidence limits. . . . .	32
2.4	BPTI hydrogen bonds that show large changes in binding at 1 atm . . . . .	40
2.5	Lysozyme $\Delta G$ at values at 283, 298, 313K. Protein Entropy and Enthalpy Values Numbers in parenthesis represent 95% confidence limits. . . . .	42
2.6	Percent of simulation hydrogen bond exists between buried water and surrounding cavity. . . . .	43
2.7	Lysozyme hydrogen bonds that show changes in binding at 1 atm greater then 10% . . . . .	43
3.1	Barnase thermodynamic data . . . . .	56
3.2	Barnase structure changes . . . . .	58
3.3	Barnase Hydrogen Bonds 1-21 . . . . .	59
3.4	Barnase Hydrogen Bonds 22-44 . . . . .	61
3.5	Barnase: Secondary Structure Details . . . . .	63
3.6	BPTI thermodynamic data . . . . .	63
3.7	BPTI: structural changes . . . . .	64
3.8	BPTI: Secondary Structure Details . . . . .	68
3.9	BPTI: hydrogen bonds 1-25 . . . . .	70
3.10	BPTI hydrogen bonds 25-50 . . . . .	71
3.11	BPTI hydrogen bonds 50-73 . . . . .	72
3.12	Lysozyme thermodynamic data . . . . .	73
3.13	Percent of simulation hydrogen bond exists between buried water and surrounding cavity. . . . .	73
3.14	Lysozyme structural changes . . . . .	74
3.15	Lysozyme: Secondary Structure Details . . . . .	77
4.1	Results for the TIP4P-FQ model for the potential energy per molecule, the average dipole moment, the root-mean-square of the dipole moment, and the translational diffusion constant as a function of the time step. Also shown is the average energy for the TIP4P-EW model. Using the Iterative method. .	87
4.2	Results for the TIP4P-FQ model using the extended Lagrangian method with a single charge mass, $m_q$ . . . . .	92

4.3	Results for the TIP4P-FQ model using the extended Lagrangian method with different charge masses for the two charge sites, M and H. . . . .	96
4.4	Results for the TIP4P-FQ model using the extended Lagrangian method with charge normal modes. . . . .	97

# List of Figures

1.1	Ribbon structure of three proteins in unit cell of 1BRI. . . . .	4
1.2	1BRI cavity structure . . . . .	5
1.3	Wild-type 1A2P structure versus 1BRI mutant . . . . .	6
1.4	2HEA asymmetrical unit from crystal structure :ribbon structure . . . . .	7
1.5	2HEA cavity structure . . . . .	8
1.6	5PTI assumed biological unit ribbon structure . . . . .	9
1.7	5PTI cavity structure . . . . .	10
1.8	The Lennard-Jones interactions . . . . .	11
2.1	The water molecule W274 in the interior of the I76A mutant of barnase, from the 1BRI crystal structure. <sup>2</sup> The figure was made using MOLSCRIPT. <sup>2</sup> . . .	20
2.2	Hydrogen bond distance between the H of the cavity water (WTP) and the carbonyl oxygen of a) Phe 7 and b) Tyr 97. . . . .	21
2.3	Mean square deviation of the full protein backbone ( $\diamond$ ) and protein backbone atoms within 6 Å of the cavity water (+) from the reference structure of 1BRI. Errors reported are 95% confidence limits. . . . .	22
2.4	Wild-type structure versus mutant structure cavity with Ile 88 . . . . .	23
2.5	Mean square deviation of heavy atoms in Barnase from the reference structure of 1BRI. Errors reported are 95% confidence limits. . . . .	24
2.6	Distribution of hydrogen bond lengths between the H on Ala 11 and the O on Phe 7 in both the empty and hydrated simulations of 1BRI. . . . .	25
2.7	Percentage change in time hydrogen bonds exist between the empty and hydrated states of 1BRI. . . . .	26
2.8	Percentage change in time solvent is hydrogen bound between the empty and hydrated states of 1BRI. . . . .	28
2.9	Polyhedron defining volume in Barnase . . . . .	29
2.10	$\chi$ 2 versus time. . . . .	30
2.11	Distribution of $\chi$ 2 torsion angle in barnase simulations . . . . .	31
2.12	Distribution of $\chi$ 1 torsion angle in barnase simulations . . . . .	32
2.13	The water molecule W122 in the interior of BPTI, showing the hydrogen bonds to the protein backbone. The coordinates are from the 5PTI crystal structure. <sup>3</sup> The figure was made using MOLSCRIPT. <sup>2</sup> . . . . .	34
2.14	Hydrogen bond distance between the two protons of the cavity water (WTP) and the oxygen of Thr 11. . . . .	35



2.15	Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6 Å of the cavity water (+) from the reference structure of 5PTI. Errors reported are 95% confidence limits. . . . .	36
2.16	Distribution of hydrogen bond lengths between the H on Gly 36 and the O on Thr 11 in both the empty and hydrated simulations of BPTI. . . . .	37
2.17	Mean square deviation of heavy atoms in BPTI from the reference structure. Errors reported are 95% confidence limits. . . . .	38
2.18	Percentage change in time hydrogen bonds exist between the empty and hydrated states of BPTI. . . . .	39
2.19	Tetrahedron defining volume in BPTI . . . . .	40
2.20	The water molecule W445 in the interior of the I106A mutant of lysozyme, from the 2HEA crystal structure. <sup>15</sup> The figure was made using MOLSCRIPT. <sup>2</sup>	41
2.21	Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6 Å of the cavity water (+) from the reference structure of 2HEA. Errors reported are 95% confidence limits. . . . .	44
2.22	Percentage change in time hydrogen bonds exist between the empty and hydrated states of Lysozyme. . . . .	45
2.23	Polyhedron defining volume in Lysozyme . . . . .	46
2.24	Free energy of hydration as the Lennard Jones well depth is scaled by 20% .	46
2.25	$\Delta G_{hyd}$ versus available hydrogen bonds . . . . .	47
2.26	Entropy versus available hydrogen bonds . . . . .	48
2.27	$\Delta G_{hyd}$ versus available hydrogen bonds (Lysozyme = 0.5 hydrogen bonds) .	49
2.28	Entropy versus available hydrogen bonds (Lysozyme = 0.5 hydrogen bonds)	51
3.1	Barnase $\lambda$ versus flexibility . . . . .	57
3.2	Barnase: Hydrogen bond near cavity HAla11 $\cdots$ 3OPhe7 . . . . .	58
3.3	Barnase: hydrogen bond map 1bar vs 3kbar . . . . .	60
3.4	$\chi^2$ Ile 88 histogram . . . . .	62
3.5	BPTI: flexibility versus hydration . . . . .	65
3.6	BPTI hydrogen bond map 1 atm and 3 kbar . . . . .	66
3.7	Slope of Mean square deviation for heavy atoms in BPTI 3 kbar versus 1 bar	67
3.8	BPTI hydrogen bond near cavity HGly36 $\cdots$ 3OThr11 histogram . . . . .	69
3.9	Lysozyme flexibility versus extent hydration . . . . .	75
3.10	Lysozyme hydrogen bond map 1 atm and 3 kbar . . . . .	76
4.1	Oxygen atom charge (m-site) versus time comparing the exact (dashed line) and the extended Lagrangian values (solid line). . . . .	89
4.2	Hydrogen atom charge versus time comparing the exact (dashed line) and the extended Lagrangian values (solid line). . . . .	90
4.3	Frequency spectrum of the charge velocity autocorrelation function using the original extended Lagrangian method (Method 1) for the hydrogen atom (dashed line) and M-site charges (solid line). The arrows show the frequencies of the gas-phase charge normal modes for the system. . . . .	91

4.4	Frequency spectrum of the charge velocity autocorrelation function with different charge masses for the hydrogen (dashed line) and M-site charges (solid line). . . . .	93
4.5	Frequency spectrum of the charge velocity autocorrelation function using charge normal modes for the hydrogen (dashed line) and M-site charges (solid line). . . . .	94
4.6	The average time dependence of the charge temperature with a single charge mass of $6.0 \times 10^{-5}$ (ps/e) <sup>2</sup> and $\delta t = 1.0$ fs (solid line), charge normal modes with a charge mass of $3.0 \times 10^{-5}$ (ps/e) <sup>2</sup> and $\delta t = 1.0$ fs (dotted line), and charge normal modes with a charge mass of $1.2 \times 10^{-4}$ (ps/e) <sup>2</sup> and $\delta t = 2.0$ fs (dashed line). . . . .	95
4.7	The oxygen-oxygen correlation function, comparing the results using a 1 fs time step with a single charge mass for both charge types equal to $6.0 \times 10^{-5}$ (ps/e) <sup>2</sup> (solid line) and using a 2 fs time step with the normal mode method for charge dynamics and a charge mass equal to $1.2 \times 10^{-4}$ (ps/e) <sup>2</sup> (dashed line). . . . .	98

## Abstract

Free energy calculations for the transfer of a water molecule from the pure liquid to an interior cavity site in a protein are presented. Three different protein cavities, in bovine pancreatic trypsin inhibitor (BPTI), the I106A mutant of lysozyme, and in the I76A mutant of barnase, represent very different environments for the water molecule, one which is polar, forming four water-protein hydrogen bonds, and two which are more hydrophobic, only forming one or two water-protein hydrogen bonds. The calculations give very different free energies for the different cavities, with only the polar BPTI cavity predicted to be hydrated. The corresponding entropies for the transfer to the interior cavities are calculated as well and show that the transfer to the polar cavity is significantly entropically unfavorable while the transfer to the non-polar cavity is entropically favorable. For all proteins an analysis of the fluctuations in the positions of the protein atoms shows that the addition of a water molecule makes the protein more flexible. This increased flexibility appears to be due to an increase length and weakened strength of protein-protein hydrogen bonds near the cavity.

Similar free energy studies are performed on the three proteins at high pressure, 3 kbar. As in the 1 atm studies BPTI is the only protein that should be hydrated at 3 kbar, however the  $\Delta G_{protein}$  changes appear to be not strongly dependent on the number of hydrogen bonds available. Changes in protein structure and flexibility are analyzed in an attempt to more fully understand the changes proteins undergo prior to pressure induced denaturation. These changes can help understand the forces at work in the last stages of protein folding. The role of interior water in this process is also analyzed.

Changes to the fluctuating charge algorithm that handles polarizability in molecular dynamics simulations were performed to allow for longer time steps. The fluctuating charge model treats partial charges as variables which are propagated using Lagrangian dynamics. A coordinate transform to normal mode charge variables is applied to the TIP4P-FQ model of water to decrease the coupling between the atomic and charge degrees of freedom.

# Chapter 1

## Introduction and Background

### 1.1 Proteins and Interior Waters

Historically proteins have been split into two classes based on the appearance of their quaternary structure, fibrous and globular. While fibrous proteins are a vital components of living systems with members including keratins and collagen, the more prevalent of the two types is the globular protein.<sup>21</sup> So named for its globule or spherical shape, the terminology predates techniques such as x-ray crystallography. The folding pathway of these proteins is believed to be largely guided by the hydrophobic effect. That is to say, the non-polar amino acid residues will aggregate in the central portion of the folded structure while the more polar side-chains favor the exterior sections of the tertiary structure where they are able to interact with the aqueous environment.<sup>6,21</sup> This folding process begins with the formation of the secondary structure,  $\alpha$  helices and  $\beta$  sheets, followed by the collapse of the hydrophobic groups and expulsion of the majority of solvent water from the interior.<sup>21</sup> This pathway results in well-packed interiors, however they are not without defects in the form of cavities. Cavities large enough to contain at least one water molecule account for approximately 1% of the total protein volume, or 1 cavity per 4 amino acid residues.<sup>7-10</sup> In fact proteins with cavities formed by mutation are still able to function,

unless the hydrophobicity of the protein core is altered.<sup>11</sup> There are proposed roles for naturally occurring cavities within a protein structure, which are to minimize strain caused by dense packing and increase protein flexibility.<sup>11</sup>

Water in proximity to proteins are typically divided into three categories.<sup>12</sup> The first is bulk water which surrounds the protein but does not have strong interactions with the residues. The second is surface water which does interact with the protein. The third type of water is found in deep clefts or protein cavities where it does not exchange rapidly with either of the first two types. The extent of hydration of protein cavities is a complex question to answer. Typical techniques used to determine the structure of proteins in solution are X-ray crystallography, nuclear magnetic resonance imaging (NMR), and neutron diffraction. A 1994 study of protein cavities imaged using these techniques indicated approximately 18% of the cavities large enough to contain a water molecule were hydrated.<sup>7</sup> That results in approximately 1 water molecule for every 27 amino acid residues was based on the study of 75 protein structures. The average number of hydrogen bonds, or polar contacts, between the buried water moieties and their surroundings determined in these early studies was between 3 and 3.23.<sup>7,9,13</sup> However the ability to detect water molecules is highly environment dependent, and will be more difficult to detect in cavities with 2 or less hydrogen bonds available. While the paper by Goodfellow *et al.* noted that 42% of the buried waters in their study had 2 or less polar contact with the amino acids bordering the site the majority of those were in contact with other waters in a interior water cluster. The difficulty in detecting waters in less enthalpically favored environments is due to their disorder, making detection by traditional x-ray techniques difficult while spectroscopic methods like NMR are able to detect water in a broader range of environments. While NMR is able to avoid some of the pitfalls of under-detection of interior water there are questions in the literature regarding over-detection.<sup>1</sup>

The presence of waters in hydrophilic cavities is easily grasped, as they can serve as hydrogen bond bridges between otherwise unsatisfied polar moieties thereby relieving

internal strain. However buried waters and interior cavities may serve more of a functional purpose.<sup>7,12,16,30</sup> While often referred to as packing defects, the presence of water in internal cavities and voids are usually conserved between protein homologs such as the serine proteases.<sup>7,17</sup> A 1998 study proposed that the movement of water through interior cavities in serine proteases is necessary for substrate binding.<sup>18</sup> A more recent study involving Cytochrome C Oxidase, a membrane protein that acts as a protein pump and has a role in ATP synthesis of aerobic cells, does not function without the presence of water in the catalytic cavity.<sup>30</sup>

Examples of proteins with hydrophobic inner cavities include two produced by point mutations. The isoleucine 76 to alanine mutant of barnase, and the isoleucine 106 to alanine mutant of human lysozyme. In each case the crystal structure indicates the presence of an interior water despite the surrounding cavity having two or less hydrogen bonds available. The detection of water in such a disordered environment is attributed to the small available volume in each cavity which limits the positional disorder of the oxygen. The resolution of both proteins in the crystal structure was under 2 angstroms with barnase measured at 1.9 Å and lysozyme at 1.8 Å.

There are 3 protein units within the unit cell of the barnase mutant which has P 3<sub>2</sub> symmetry (Figure 1.1). Of these units, A, B, and C, only one, protein C (in red), had sufficient electron density to identify the presence of a water molecule. There are no large changes between the three crystal structures, in fact the root-mean-squared deviations (RMSD) of the structures are negligible. The protein cavity is largely hydrophobic with only the polar oxygen of Phe 7 available to hydrogen bond (Figure 2.1). The mutation of the isoleucine residue into the smaller alanine resulted in the formation of an interior cavity. A second isoleucine, residue 88, rotated into the space created by this mutation (Figure 2.4). The wild-type Barnase is a non-specific ribonuclease produced by *Bacillus amyloliquefaciens* and is found in an extracellular environment.<sup>19</sup> This bacterial enzyme is toxic to the cell with the active site centered around His 102.<sup>20</sup>

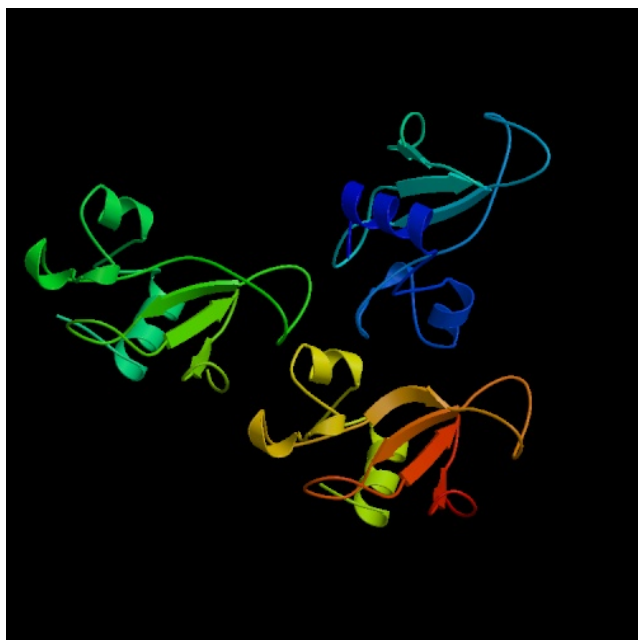


Figure 1.1: Ribbon structure of three proteins in unit cell of 1BRI.

The lysozyme is a mutant of human lysozyme which hydrolyzes 1,4- $\beta$ -linkages in chitodextrins. The largest of the proteins studied it has 130 residues and is extensively studied both in the human variation and those in this protein family but expressed in other organisms (Figure 1.4).<sup>3-5,15,16,24</sup> A 1993 report by Pepys *et al.*<sup>25</sup> even suggested a link between Lysozyme mutants and amyloid fibrils. The mutant of lysozyme formed when the Ile 106 is mutated into an Ala has two possible hydrogen bond donors indicated in the x-ray crystal structure (Figure 1.5).

The bovine pancreatic trypsin inhibitor is one of the most studied proteins both for its small size and pharmaceutical applications (Figure 1.6).<sup>26</sup> The protein is marketed as a pharmaceutical for use in open heart surgery. There are four internal waters in this relatively small protein containing 58 residues of these water molecules, 3 are grouped together while 1 is isolated in a polar cavity with 4 possible hydrogen bonds (Figure 2.13).<sup>3</sup>

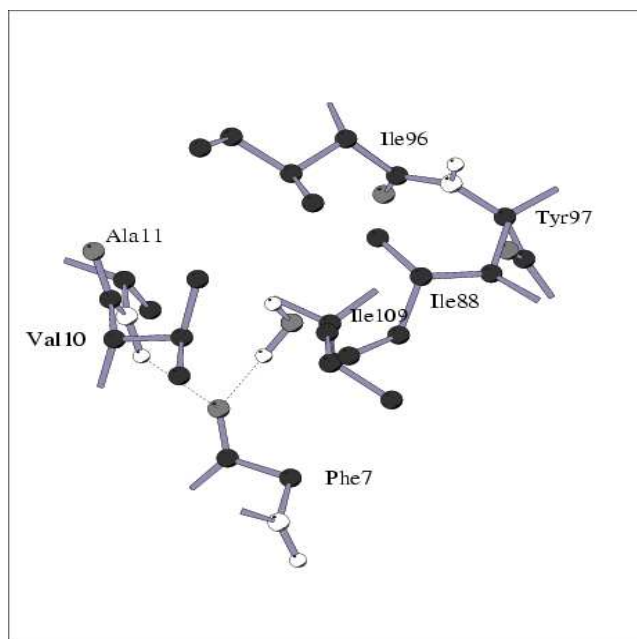


Figure 1.2: The water molecule W274 in the interior of the I76A mutant of barnase, from the 1BRI crystal structure.<sup>2</sup> The figure was made using MOLSCRIPT.<sup>2</sup>

## 1.2 Higher Pressure

The effect of increased pressure on globular protein structure and function is not completely intuitive. For instance while the change in volume upon denaturation is small at atmospheric pressure, with the unfolded protein occupying slightly more volume than its folded counterpart, the volume of the denatured protein is more compact at elevated pressure. This is contrary to the predicted effect using a simple hydrocarbon model.<sup>8,13–17</sup> The dynamics of proteins under pressure have been analyzed to try to understand conformational changes proteins undergo prior to denaturation.<sup>5–8</sup> Studies of the denaturation process at increased pressure allow the role of decreased total volume to be separated from other variables that contribute in thermal or chemical unfolding.<sup>6,9–12</sup> A thorough understanding of the forces involved in protein unfolding can be used to unravel the forces at work in the complicated process governing the transition from the molten globule stage to the final tertiary structure.<sup>12,18–21</sup>

These pressure induced changes have also raised questions regarding the role of



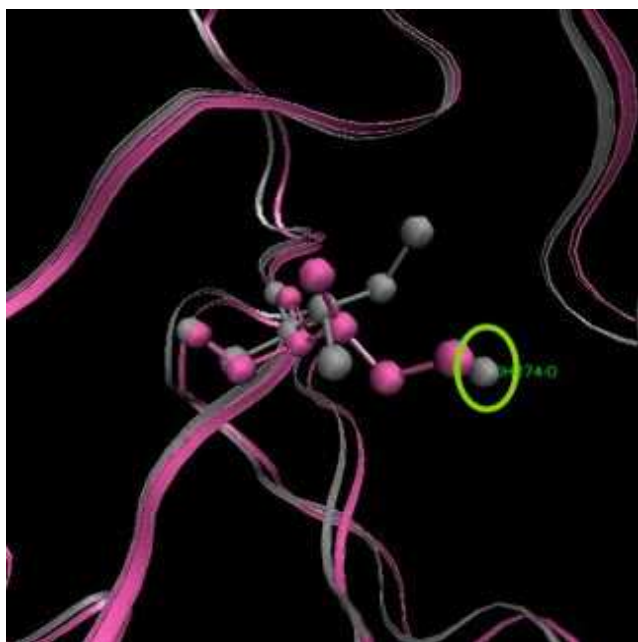


Figure 1.3: Wild-type structure versus mutant structure cavity with Ile 88. Wild-type structure is seen in purple, the mutant in silver. The circled moiety is the crystal water in the mutant structure. Proteins visualized using VMD software.

buried waters with earlier work indicating that internal water or the introduction of water into unoccupied cavities is a possible nucleation point for the process<sup>5,13,42,43</sup> Ideally X-Ray crystallography could be used to fully elucidate the nature of the structural changes, however acquiring structures at high pressure requires specialized equipment and problems with the instability of the crystal also hamper data collection.<sup>24,42</sup> Methods such as NMR have come to the forefront in the attempt to elucidate the nature of the structural changes at increased pressure.<sup>9,11,12,18,42,44</sup> While in depth studies of such data can be used to determine shifts in hydrogen bond lengths and angles the exact role of the internal water is difficult to study via such experiments. The theoretical studies performed have been hampered by difficulty characterizing the small changes in hydrogen bond lengths, which often falls within the calculated error.<sup>45</sup>

The role of buried waters in structural changes that occur at higher pressure can be explored through molecular dynamics studies where interactions between the protein and the internal water can be turned on and off. The process of calculating the free energy of

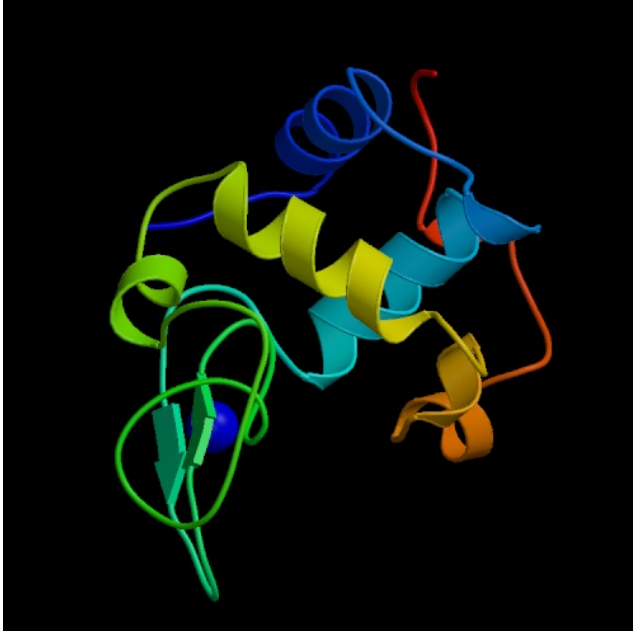


Figure 1.4: 2HEA asymmetrical unit from crystal structure :ribbon structure

hydration also affords the opportunity to observe the hydration dependence of hydrogen bond lengths, volume, and overall protein flexibility. These data can be combined to form both a more thorough understanding of the effects of pressure on protein dynamics and the contribution of internal waters in any changes that occur.

### 1.3 Simulations

The majority of the calculations presented in this work will have been performed without the addition of polarizability, that is the electron density does not respond to changes in the local environment. A typical potential for such a system is seen in Eq 1.1.

$$\begin{aligned}
 U = & \sum_{bonds} K_B(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 \\
 & + \sum_{dihedrals} \sum_n \frac{V_n}{2} (1 + \cos(n\phi - \gamma)) \\
 & + \sum_{nonbondedpairs} 4\epsilon_{ij} \left[ \frac{(\sigma_{ij})^1}{r_{ij}} - \frac{(\sigma_{ij})^6}{r_{ij}} \right] + \frac{q_i q_j}{r_{ij}}
 \end{aligned} \tag{1.1}$$

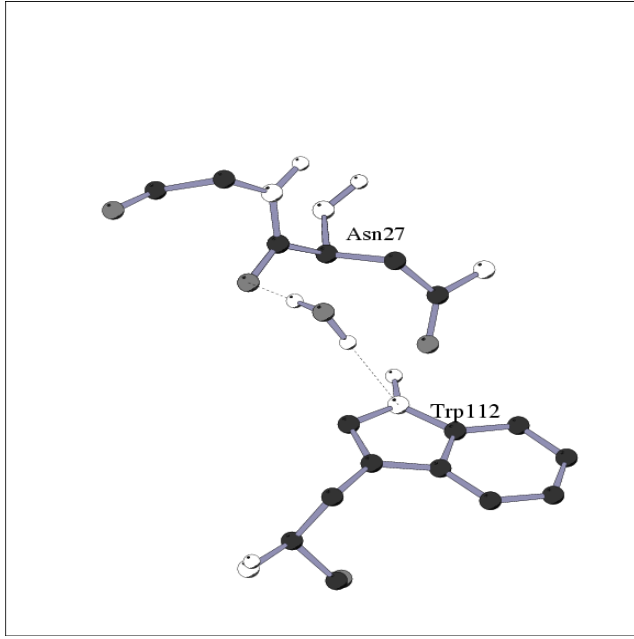


Figure 1.5: The water molecule W445 in the interior of the I106A mutant of lysozyme, from the 2HEA crystal structure.<sup>15</sup> The figure was made using MOLSCRIPT.<sup>2</sup>

The  $r$  is bond length,  $K$  is the force constant, the bond angle is  $\theta$ , the dihedral angle is  $\phi$ , and the barrier height is  $V_n$ . The equilibrium or average values are  $r_0$  for bond length,  $\theta_0$  and  $\gamma$  for equilibrium angles. The coulombic interaction is  $q_i q_j / r_{ij}$  and the intermolecular forces are described with a Lennard-Jones interaction,  $(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6$ . The nature of the repulsive component of the Lennard-Jones ( $r^{-12}$ ) and the long range attractive portion ( $r^{-6}$ ) can be seen in Figure 1.8. A rough polarizability is included in the parameterization of the charges for the coulombic interactions and the  $\sigma$  values for the Lennard-Jones interaction.<sup>1</sup>

Free energies can be obtained from one of two main methods, thermodynamic integration and free energy perturbation.<sup>12</sup> In thermodynamic integration the free energy change from the empty state,  $\lambda = 0$ , to the hydrated state,  $\lambda = 1$  is determined through the relationship seen in Eq 1.2, where the integral is approximated using the trapezoidal rule. Simulations at several values of  $\lambda$  are done, and averages  $\langle \dots \rangle_\lambda$  are calculated.

$$G_1 - G_0 = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda. \quad (1.2)$$

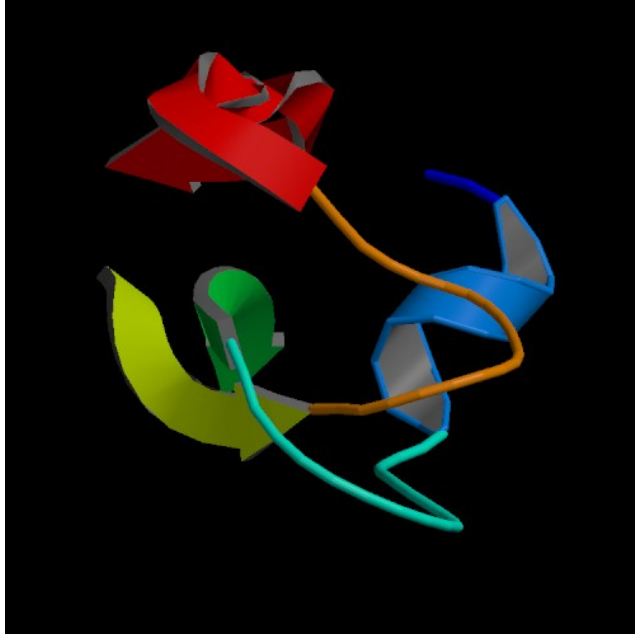


Figure 1.6: 5PTI assumed biological unit ribbon structure

For free energy perturbation the value of  $\Delta G$  is determined by

$$\Delta G = G_1 - G_0 = \sum_i G_{\lambda(i+1)} - G_{\lambda(i)} \quad (1.3)$$

$$= \sum_i (-RT \ln \langle e^{-\frac{[V_{\lambda(i+1)} - V_{\lambda(i)}]}{RT}} \rangle_{\lambda(i)}). \quad (1.4)$$

While both the forward  $\lambda \rightarrow \lambda + \delta\lambda$  and  $\lambda \rightarrow \lambda - \delta\lambda$  are calculated and should give the same average change in free energy, the magnitude of the errors are different and therefore the average value will be skewed.<sup>13</sup>

Many methods for treating polarizability exist (dipole polarizability, shell models, fluctuating charges). Dipoles introduce new interactions and so are more computationally expensive. Fluctuating charge models do not introduce new intermolecular interactions so are in principle less expensive. Applications to proteins have so far required the use of a smaller time step.<sup>1,2,22</sup>

An advantage of the addition of polarizability through electronegativity equalization is that it allows the addition of polarizability without requiring the new interactions. A

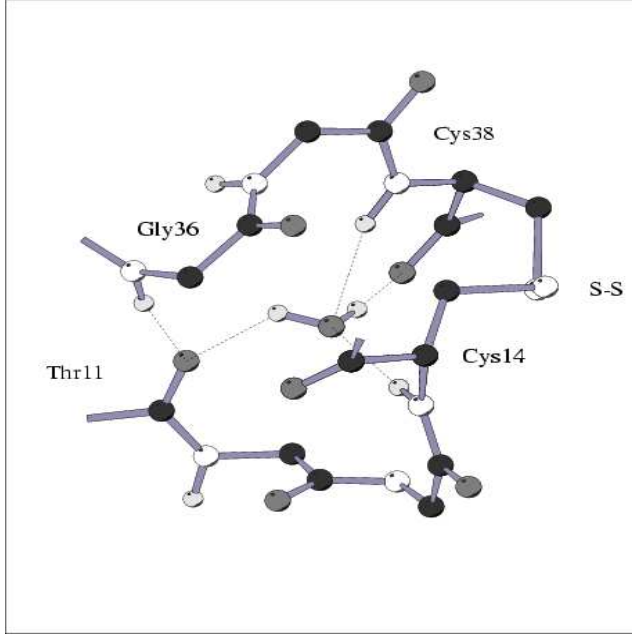


Figure 1.7: The water molecule W122 in the interior of BPTI, showing the hydrogen bonds to the protein backbone. The coordinates are from the 5PTI crystal structure.<sup>3</sup> The figure was made using MOLSCRIPT.<sup>2</sup>

Taylor series, where we ignore higher level terms, can be used to express the energy to create a charge,  $q$ , on an atom

$$U(q) = E^0 + \chi^0 q + 1/2 J q^2. \quad (1.5)$$

From this Taylor series the ionization potential, IP, is  $U(1) - U(0)$ , and the electron affinity, EA, is  $U(0) - U(-1)$ . The coefficients for the Taylor series can be determined using the following relationships  $IP = \chi^0 + 1/2 J$  and  $EA = \chi^0 - 1/2 J$ , therefore  $\chi^0 = 1/2(IP - EA)$  and  $J = IP - EA$ . The definition for  $\chi^0$  is Mulliken's definition for electronegativity, and  $1/2 J = \eta$  the hardness of an atom.<sup>51</sup> For molecules the Electronegativity Equilization model becomes

$$U(q) = \sum_{\alpha} (E_{0,\alpha} + \chi_{\alpha}^0 q_{\alpha} + 1/2 J_{\alpha\alpha} q_{\alpha}^2) + \sum_{\alpha} \sum_{\beta > \alpha} J_{\alpha\beta}(r_{\alpha\beta}) q_{\alpha} q_{\beta} \quad (1.6)$$

where  $J_{\alpha\beta}$  is dependent on atomic distances and approaches  $1/r_{\alpha\beta}$  at large distances.<sup>1</sup>

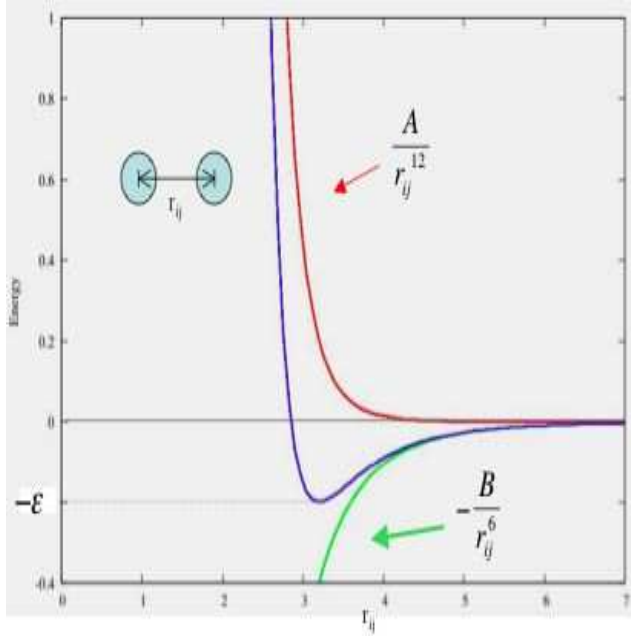


Figure 1.8: The Lennard-Jones interactions

With the charge constraints

$$\sum_{\alpha=1}^{N_q} \dot{q}_{i,\alpha} = 0 \quad \sum_{\alpha=1}^{N_q} \ddot{q}_{i,\alpha} = 0 \quad (1.7)$$

the undetermined multiplier,  $\lambda$  can be determined

$$U' = U + \lambda \sum q U(\mathbf{q}) = U(\mathbf{q}) - \lambda \left( \sum_{\alpha} q_{\alpha} - q_{TOT} \right). \quad (1.8)$$

In neutral molecules this gives

$$\left( \frac{\partial U}{\partial q_{\alpha}} \right) - \lambda = 0. \quad (1.9)$$

Since this is true for all atoms, and  $\partial U / \partial q$  is the electronegativity the minimization of energy is as simple as equalizing the electronegativities

$$\chi_{\alpha} = \left( \frac{\partial U}{\partial q_{\alpha}} \right) = \chi_{\alpha}^0 + J_{\alpha\alpha} q_{\alpha} + \sum_{\beta \neq \alpha} J_{\alpha\beta} r_{\alpha\beta} q_{\beta}. \quad (1.10)$$

To avoid the computational expense of the iterative method or matrix inversion the charges in the fluctuating charge method are propagated according to Newtonian mechanics with the extended Lagrangian.<sup>14</sup> Charges are given a fictitious mass,  $M_q$ , and the equations of motion are calculated.

$$M_q \ddot{q}_{i\alpha} = -\frac{\partial U}{\partial q_{i\alpha}} - \lambda_i \quad (1.11)$$

$$\lambda_i = -\frac{1}{N_{atoms}} \sum_{\alpha=1}^{N_{atoms}} \frac{\partial U}{\partial q_{i\alpha}} \quad (1.12)$$

$$(1.13)$$

The iterative method is used to determine the charges exactly at the beginning of a simulation, and at periodic intervals throughout the simulation to keep the charges cool.<sup>1</sup> The fluctuating charge method is less computationally expensive than the iterative method, but charge temperature requires a small time step for some applications. The change in algorithms we explore attempt to address this problem.

# Bibliography

- [1] Buckle, A. M.; Cramer, P.; Fersht, A. R. *Biochemistry* **1996**, *35*, 4298–4305.
- [2] Kraulis, P. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.
- [3] Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301–329.
- [4] Takano, K.; Funahashi, J.; Yamagata, Y.; Fujii, S.; Yutani, K. *J. Mol. Biol.* **1997**, *274*, 132–142.
- [5] Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry*. John Wiley and Sons, Inc., New York 2002.
- [6] Kauzmann, W. *Nature* **1987**, *325*, 763–764.
- [7] Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 1224–1235.
- [8] Schoenborn, B. P.; Garcia, A.; Knott, R. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 105–119.
- [9] Rashin, A. A.; Iofin, M.; Honig, B. *Biochemistry* **1986**, *25*, 3619–3625.
- [10] Liang, J.; Dill, K. A. *Biophys. J.* **2001**, *81*, 751–766.
- [11] Akasako, A.; Haruki, M.; Oobatake, M.; Kanaya, S. *J. Biol. Chem.* **1997**, *272*, 18686.
- [12] Tarek, M.; Tobias, D. J. *Biophys. J.* **2000**, *79*, 3244–3257.
- [13] Edsall, J. T.; McKenzie, H. A. *Adv. Biophys* **1983**, *16*, 53–183.
- [14] Zhang, L.; Hermans, J. *Proteins: Struct. Funct. Gen.* **1996**, *24*, 433–438.



- [15] Tashiro, M.; Stuchebrukhov, A. A. *J. Phys. Chem. B* **2005**, *109*, 1015–1022.
- [16] Xu, J.; Baase, W. A.; Quillin, M. L.; Baldwin, E. P.; Matthews, B. W. *Protein Science* **2001**, *10*, 1067–1078.
- [17] Sreenivasan, U.; Axelsen, P. H. *Biochemistry* **1992**, *31*, 12785.
- [18] Meyer, E.; Cole, G.; Radhakrishnan, R.; Epp, O. *Acta Crystallogr B*. **1988**, *44*, 26–55.
- [19] Yanagawa, H.; Yoshida, K.; Torigoe, C.; Park, J.-S.; Sato, K. *J. Biol. Chem.* **1993**, *268*, 5861–5965.
- [20] Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878–8889.
- [21] Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.
- [22] Mann, G.; Hermans, J. *J. Mol. Bio.* **2000**, *302*, 979–989.
- [23] Refaee, M.; Tezuka, T.; Akasaka, K.; Williamson, M. P. *J. Mol. Biol.* **2003**, *327*, 857–865.
- [24] Akasaka, K.; Tezuka, T.; Yamada, H. *J. Mol. Biol* **1997**, *271*, 671–678.
- [25] Pepys, M. B.; Hawkins, P. N.; Booth, D. R.; Vigushin, D. M.; Tennent, G. A.; Soutar, A. K.; Totty, N.; Nguyen, O.; Blake, C.; Terry, C. J. *Nature* **1993**, *362*, 553–557.
- [26] Ascenzi, P.; Bocedi, A.; Bolognesi, M.; Spallarossa, A.; Coletta, M.; Cristofaro, R. D.; Menegatti, E. *Curr Protein Pept Sci* **2003**, *4*, 231–251.
- [27] Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. *Proc. Natl. Acad. Sci.* **1998**, *95*, 1552–1555.
- [28] Harpaz, Y.; Gerstein, M.; Chothia, C. *Structure* **1994**, *2*, 641–649.
- [29] Royer, C. A. *Methods in Enzymology* **1995**, *259*, 357–377.

- [30] Kitchen, D. B.; Reed, L. H.; Levy, R. M. *Biochemistry* **1992**, *31*, 10083–10093.
- [31] Payne, V. A.; Matubayasi, N.; Murphy, L. R.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101*, 2054–2060.
- [32] Hillson, N.; Onuchic, J. N.; Garcia, A. E. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14848–14853.
- [33] Wroblowski, B.; Diaz, J. F.; Heremans, K.; Engelborghs, Y. *PROTEINS: Struct. Funct. Gen.* **1996**, *25*, 446–455.
- [34] Paci, E.; Marchi, M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11609–11614.
- [35] Peng, X.; Jonas, J.; Silva, J. L. *Biochemistry* **1994**, *33*, 8323–8329.
- [36] Weber, G.; Drickamer, H. G. *Q. Rev. Biophys.* **1983**, *16*, 89–112.
- [37] Samarasinghe, S. D.; Campbell, D. M.; Jonas, A.; Jonas, J. *Biochemistry* **1992**, *31*, 7773–7778.
- [38] Jonas, J.; Ballard, L.; Nash, D. *Biophys. J.* **1998**, *75*, 445–452.
- [39] Zhang, J.; Peng, X.; Jonas, A.; Jonas, J. *Biochemistry* **1995**, *34*, 8631–8641.
- [40] Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem* **1990**, *59*, 631–660.
- [41] Buck, M.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1994**, *237*, 247–254.
- [42] Williamson, M. P.; Akasaka, K.; Refaee, M. *Protein Sci.* **2003**, *12*, 1971–1979.
- [43] Yamato, T.; Higo, J.; Seno, Y.; Go, N. *Proteins: Struct. Funct. Gen.* **1993**, *16*, 327–340.
- [44] Kamatari, Y. O.; Yamada, H.; Akasaka, K.; Jones, J. A.; Dobson, C. M.; Smith, L. J. *Eur. J. Biochem* **2001**, *268*, 1782–1793.

- [45] Li, H.; Yamada, H.; Akasaka, K. *Biochemistry* **1998**, *37*, 1167–1173.
- [46] Rick, S. W.; Stuart, S. J. Potentials and Algorithms for Incorporating Polarizability in Computer simulations. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Ed.; volume 18 John Wiley and Sons, Inc., New York 2002; pp 89–146.
- [47] Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. I.; Vincent, J. J.; Crowlet, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. Amber 6, University of California, San Francisco, 1999.
- [48] Lu, N.; Kofke, D. A. *J. Chem. Phys* **2001**, *114*, 7303–7311.
- [49] Olano, L. R.; Rick, S. W. *J. Am. Chem. Soc.* **2004**, *126*, 7991–8000.
- [50] Patel, S.; III, C. L. B. *J. Computational Chemistry* **2004**, *25*, 1–16.
- [51] Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.
- [52] Olano, L. R.; Rick, S. W. *J. Computational Chemistry* **2005**, *26*, 699–707.

# Chapter 2

## Protein Simulations at 1 bar

### 2.1 Introduction

Molecular dynamics simulations were performed on 3 proteins, the Bovine Pancreatic Trypsin Inhibitor (BPTI), the I76A mutant of Barnase, and the I106A mutant of Lysozyme. All of these proteins have x-ray crystal structures that indicate the presence of water in isolated cavities.<sup>2,3,15</sup> The three protein cavities involved have varying degrees of hydrophobicity indicated in the original crystal structures. The barnase and lysozyme mutants were created with sufficiently small cavity volumes to limit the disorder of the internal water and facilitate its observation with x-ray crystallography. The cavity within BPTI isolates the water in a rigid ice-like structure, while the more hydrophobic cavities in the barnase and lysozyme mutants have 1 and 2 available polar groups lining the cavity, respectively.

The studies were performed using the Amber 6.0 molecular dynamics programs. Calculations to determine the free energy, entropy, and enthalpy of hydration were performed using free energy perturbation. Interactions with the interior water were scaled in using a coupling term of  $\lambda$ .

$$E_{\lambda} = \lambda E_{hydrated} + (1 - \lambda) E_{empty} \quad (2.1)$$

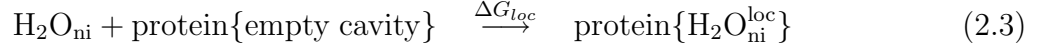
$$\Delta G = \sum_i G_{\lambda(i+1)} - G_{\lambda(i)} = \sum_i -RT \ln \langle e^{\frac{-E_{\lambda(i+1)} + E_{\lambda(i)}}{RT}} \rangle_{\lambda(i)} \quad (2.2)$$

The resulting coordinate trajectories were analyzed to determine changes in hydrogen bonding, flexibility, and volume. This data can not only be analyzed for the end points of a completely hydrated or empty cavity, but can also be determined at every value of  $\lambda$ . From this complete set of data the hydration dependence of properties can be calculated.

The disorder of water in hydrophobic protein cavities increases its B-value in x-ray crystallography which makes its identification in a crystal structure difficult.<sup>1,6-10</sup> Computational studies to determine free energy of hydration can be used to help understand the entropic and enthalpic costs of removing a water from the bulk solvent and placing it inside a variety of protein environments.<sup>15</sup> Another question regarding buried water which computational studies can help answer is how it effects protein flexibility and stability.<sup>7,8</sup> This can not only be studied at the two extremes, empty and hydrated, but also the dependence of flexibility on extend of hydration. The role of buried water in a variety of protein processes from enzyme catalysis to ligand binding has been accepted for some time, necessitating a thorough understanding of its effects and the ability to predict it's presence.<sup>13-15</sup>

## 2.2 Barnase I76A Mutant

The free energy for adding a water to the empty cavity (see Eq. 5.5)



is negative for barnase, however the hydration process is a competition between  $\Delta G_{protein}$  and the free energy of removing a water molecule to the pure liquid,  $\Delta G_{wat}$ . So while  $\Delta G_{protein}$  is negative for barnase the  $\Delta G_{wat}$  is sufficiently positive to make the overall process unfavorable, with a positive  $\Delta G_{hyd}$  (Table 2.1). Our calculated  $\Delta G_{hyd}$  for barnase is much larger than the  $-2.0$  kcal/mol estimate based on the stability of the barnase mutant, taken to have a filled cavity, relative to other Ile→Ala mutants with empty cavities.<sup>2</sup>

Table 2.1: Barnase  $\Delta G$  at values at 283, 298, 313K and Protein Entropy and Enthalpy Values Numbers in parenthesis represent 95% confidence limits.

	Protein	Water	Hydration
$\Delta G$ (kcal/mol) 283K	-1.2(2)	6.36(6)	5.1(2)
$\Delta G$ (kcal/mol) 298K	-1.5(1)	6.18(3)	4.7(1)
$\Delta G$ (kcal/mol) 313K	-1.3(2)	5.99(5)	4.7(2)
$\Delta S$ (cal/mol/K)	4(9)	12(3)	16(9)
$-T\Delta S$ (kcal/mol)	-1(3)	-3.7(8)	-5(3)
$\Delta H$ (kcal/mol)	0(3)	9.85(8)	10(3)

From the temperature dependences of the free energies, the entropy found from

$$\Delta S = -\left(\frac{\partial \Delta G}{\partial T}\right)_{P,N} \quad (2.5)$$

$$\Delta S \approx -\left(\frac{\Delta G(T + \Delta T) - \Delta G(T - \Delta T)}{2\Delta T}\right) \quad (2.6)$$

is seen in Table 2.1. The error associated with the calculation of entropy is significantly larger than other thermodynamic data. Thus, longer simulation times are required to lower the total error of the free energy measurements which in turn decreases the error bars on  $\Delta S$ .<sup>19</sup> The entropic contribution to  $\Delta G_{wat}$  is unfavorable and is in good agreement with the experimental value of -12.24 cal/mol/K.<sup>17</sup> The entropy of hydration for the barnase cavity is positive ( $\Delta S_{hyd}=16\pm9$  cal/mol/K), indicating that while the process has a positive free energy, it is entropically favored. The enthalpy change can be found from  $\Delta S$  and  $\Delta G_{hyd}$  that is calculated using the linear fit of the temperature data. This method gave a  $\Delta G_{hyd}$  of  $4.7\pm1$  kcal/mol for barnase, and therefore the enthalpic cost of placing hydrating the protein is  $10\pm3$  kcal/mol.

The x-ray crystal structure of the barnase mutant shows a water in an environment with a single hydrogen bond donor, identified at the carbonyl oxygen position of phenylalanine 7 (see Figure 2.1). In an analysis of the trajectory of the hydrated protein

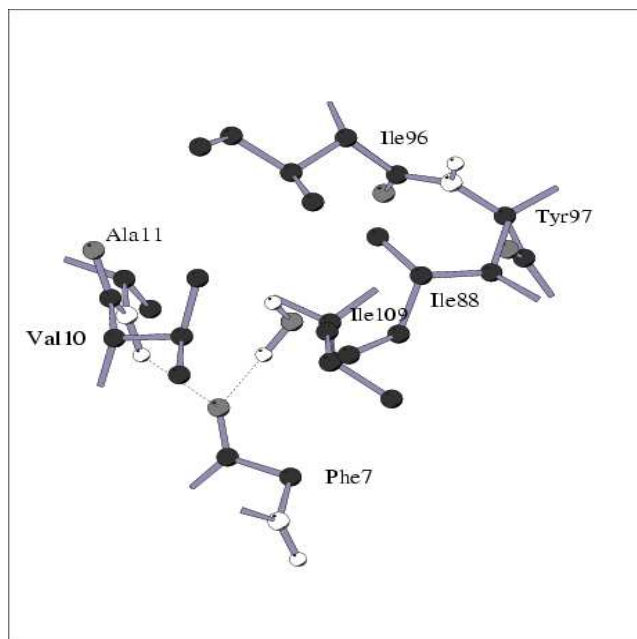


Figure 2.1: The water molecule W274 in the interior of the I76A mutant of barnase, from the 1BRI crystal structure.<sup>2</sup> The figure was made using MOLSCRIPT.<sup>2</sup>

with,  $\lambda = 1$  and  $\lambda = 0.95$ , a shift in the water position was noted, this shift allowed the water to hydrogen bond to the carbonyl oxygen position of tyrosine 97. This atom is far

enough from Phe 7 that the water molecule cannot simultaneously form hydrogen bonds to both oxygen atoms, resulting in the water exchanging between the two positions during the course of the simulation. When the distance between both protons and the two possible hydrogen bond donors are observed any distance that is less than 2.25 Å can be assumed to be a real hydrogen bond. Using this method the water in this cavity is very dynamic

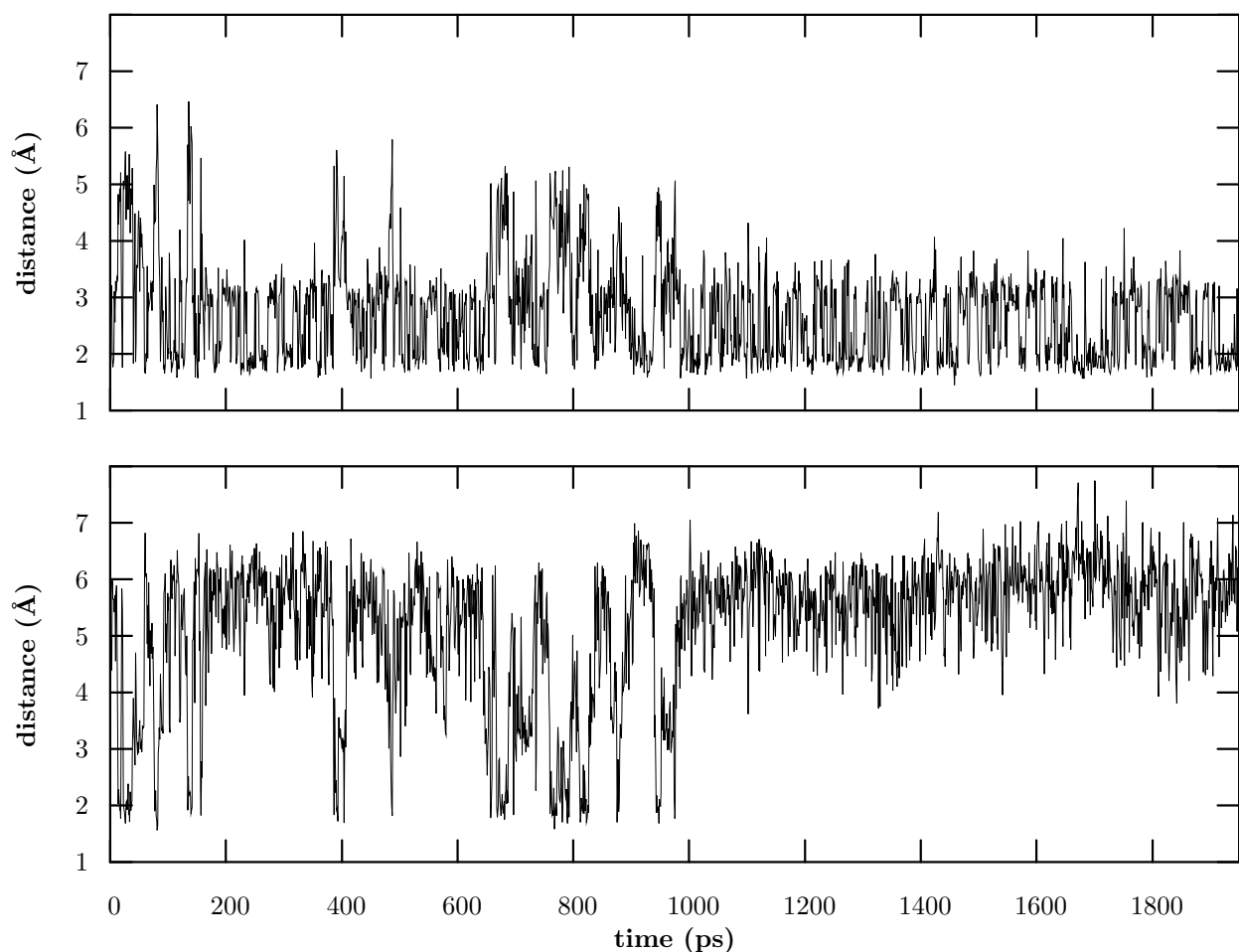


Figure 2.2: Hydrogen bond distance between the H of the cavity water (WTP) and the carbonyl oxygen of a) Phe 7 and b) Tyr 97.

with a  $C_2$  flip occurring to allow both protons to share the single available site of hydrogen bonding. While bound at the O on Phe 7 the single hydrogen bond is split between the two water proteins with an average flipping time of 4.6 ps (Figure 2.2 a). The water molecule occupies the site by the Tyr 97 13.1% of the simulation time at  $\lambda = 0.95$ . In the



secondary binding site, as in the first binding site, only one hydrogen bond is formed. The protons flip in this site about once every 14 ps (Figure 2.2 b). This longer residency time can be explained in part by the sterics of the cavity by Tyr 97, where the water must partially exit the cavity before the flip can occur.

Flexibility can be determined through the use of the mean square fluctuation in the

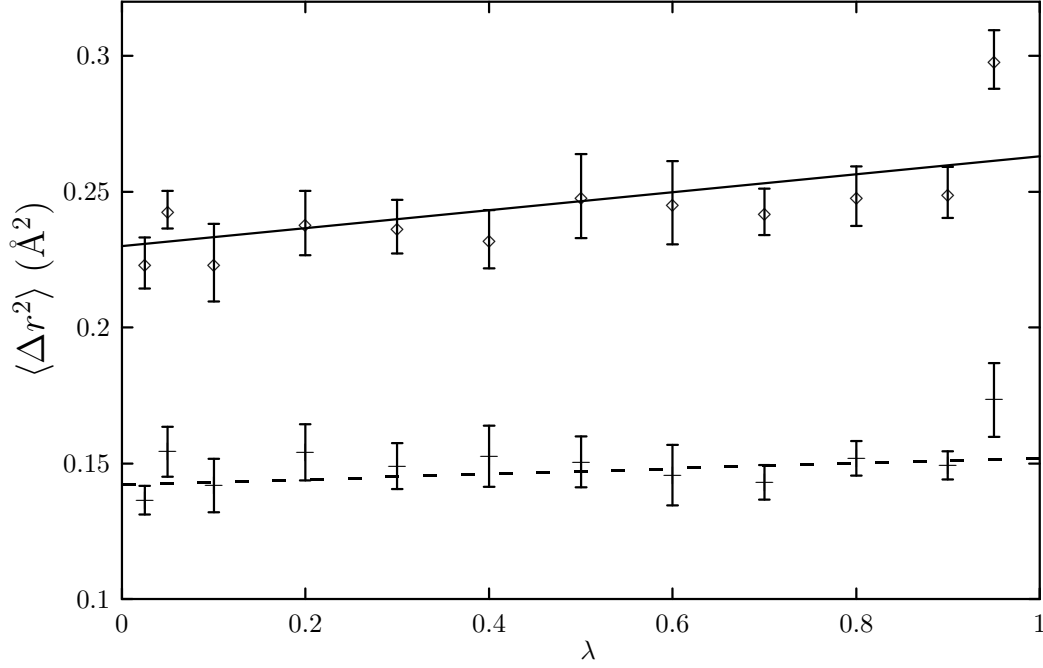


Figure 2.3: Mean square deviation of the full protein backbone ( $\diamond$ ) and protein backbone atoms within 6 Å of the cavity water ( $+$ ) from the reference structure of 1BRI. Errors reported are 95% confidence limits.

atomic positions,  $\langle \Delta r_i^2 \rangle$ . This is calculated from

$$\langle \Delta r_i^2 \rangle = \frac{1}{N} \sum_{j=1}^N (\mathbf{r}_i(t_j) - \bar{\mathbf{r}}_i)^2 \quad (2.7)$$

where  $\mathbf{r}_i(t_j)$  is the position of atom  $i$  at time  $t_j$  and  $\bar{\mathbf{r}}_i$  is the average position of atom  $i$  over the  $N$  stored time values. To eliminate motion due to translations and rotations of the protein, each configuration is rotated and translated onto a reference structure (the original crystal structure) prior to calculating  $\langle \Delta r_i^2 \rangle$  from Eq. 2.7 This is similar to calculating thermal B-factors in x-ray crystallography and allows the mobility to each atom to be

measured. A larger  $\langle \Delta r_i^2 \rangle$  value indicates that an atom is able to move to a greater extent during the course of the calculation. In general, the backbone heavy atoms show the smallest  $\langle \Delta r_i^2 \rangle$  and these are the values which can be calculated with the most precision. For this reason, we choose to focus on these atoms to examine the differences in  $\langle \Delta r_i^2 \rangle$  with and without the interior water molecule. The flexibility trends are constant whether just the backbone is analyzed or heavy atoms on side chains are included. The average  $\langle \Delta r_i^2 \rangle$

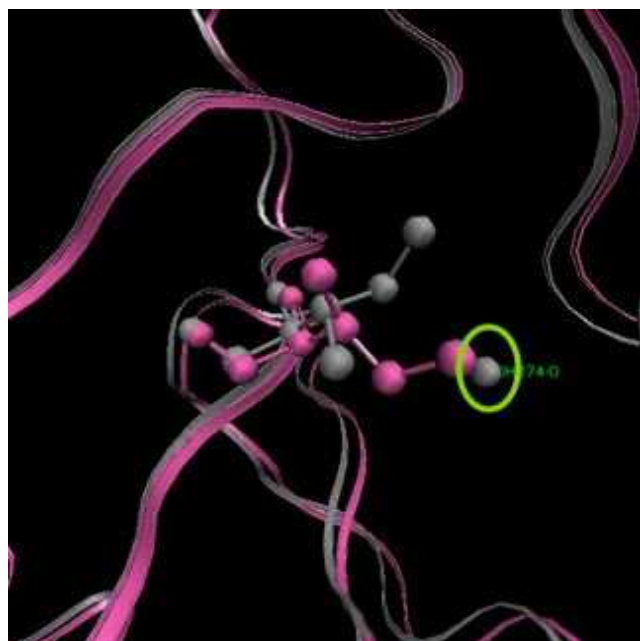


Figure 2.4: Wild-type structure versus mutant structure cavity with Ile 88

for the full protein was calculated as was the value for those atoms that fall within 6 Å of the protein cavity (Figure 2.3). The slope of the linear fit increases of 0.033 Å<sup>2</sup> and 0.10 Å<sup>2</sup> for the entire protein and close residues, respectively. One atom which decreases its value mean square fluctuations when the water is present is Ile 88 C<sup>δ</sup>. This atom borders the cavity and the rotation of Ile 88 from its position in the wild-type structure is what creates the cavity (Figure 2.4).<sup>2</sup> While other atoms show an increase in  $\langle \Delta r_i^2 \rangle$ , for this atom  $\langle \Delta r_i^2 \rangle$  decreases from 0.43±0.03 Å<sup>2</sup> to 0.23±0.02 Å<sup>2</sup>. The increased mobility of this atom in the absence of water may be due to the space created by the loss of the water. The polar atom bordering the barnase cavity (Phe 7 O) appears to get more flexible as the water is added

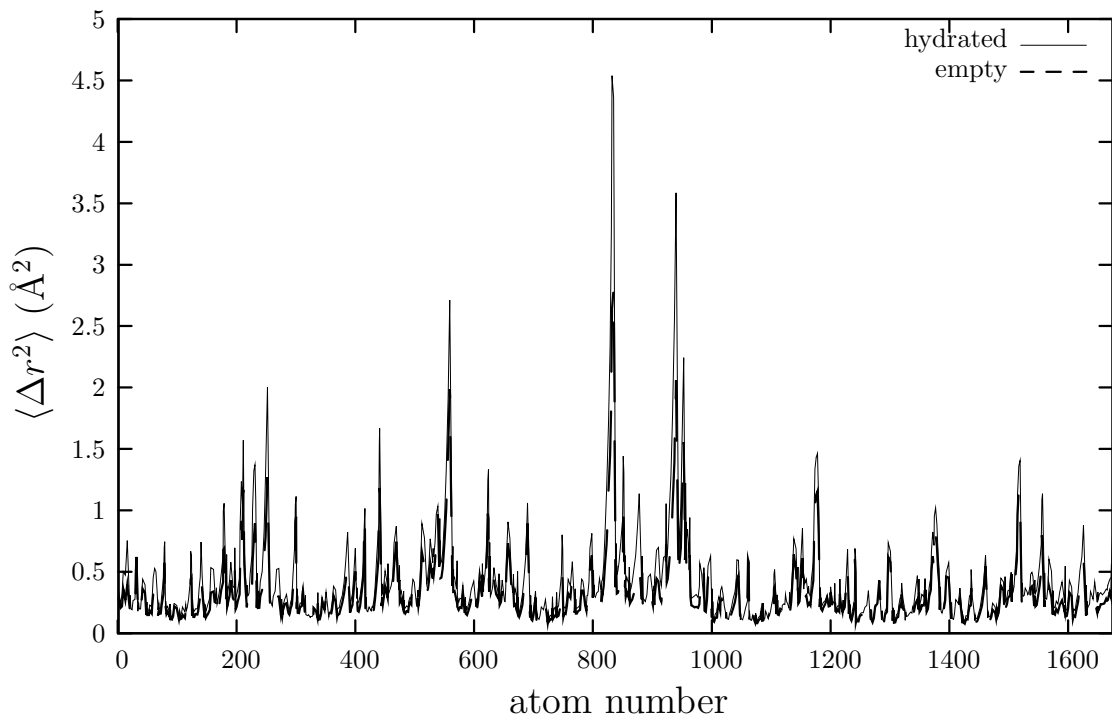


Figure 2.5: Mean square deviation of heavy atoms in Barnase from the reference structure of 1BRI. Errors reported are 95% confidence limits.

(Figure 2.5).

The increase in protein flexibility seen on hydration may be understood at least in part by looking at the hydrogen bonds formed near the cavity. The carbonyl oxygen of Phe 7 has one hydrogen bond partner other than the water (Figure 2.1). This hydrogen bond formed to the proton of Ala 11, is shorter when the cavity is empty with a length of 2.06 Å than for the hydrated state with a length of 2.18 Å (Figure 2.6). The decrease in strength of this hydrogen bond, as judged by bond length, might help explain the neutral  $\Delta H_{protein}$  determined for this protein. The formation of the water–protein bond is at the cost of the weakening of a protein–protein hydrogen bond.

While no hydrogen bonds are gained or lost within 6 Å of the protein cavity, although one does get weaker, there are changes in overall hydrogen bonding elsewhere in the protein. Calculations were performed which determined the percent of the simulation

hydrogen bonds existed. The cutoff of a hydrogen bond was a Donor $\cdots$ H–Acceptor angle of 90 degrees and donor–acceptor distance of 4 Å. Each possible combination of donor and acceptor atoms in the crystal structure was assigned a hydrogen bond number within the Carnal program. The calculation to determine which of these donor and acceptor atoms fell within the cutoff was performed for every picosecond of simulation time. From this a clear image of what hydrogen bonds are formed and lost during the simulation can be seen when the difference of these hydrogen bond calculations is plotted. By taking the difference of the two states both the hydrogen bonds that are weak and those that are of the same strength in both states will be found near the baseline. This leaves any hydrogen bond that is found when the protein cavity is empty with have a strong positive value, while any hydrogen bond that is formed upon hydration will have a large negative value (Figure 2.7). While there are noticeable changes in the hydrogen environment between the empty and

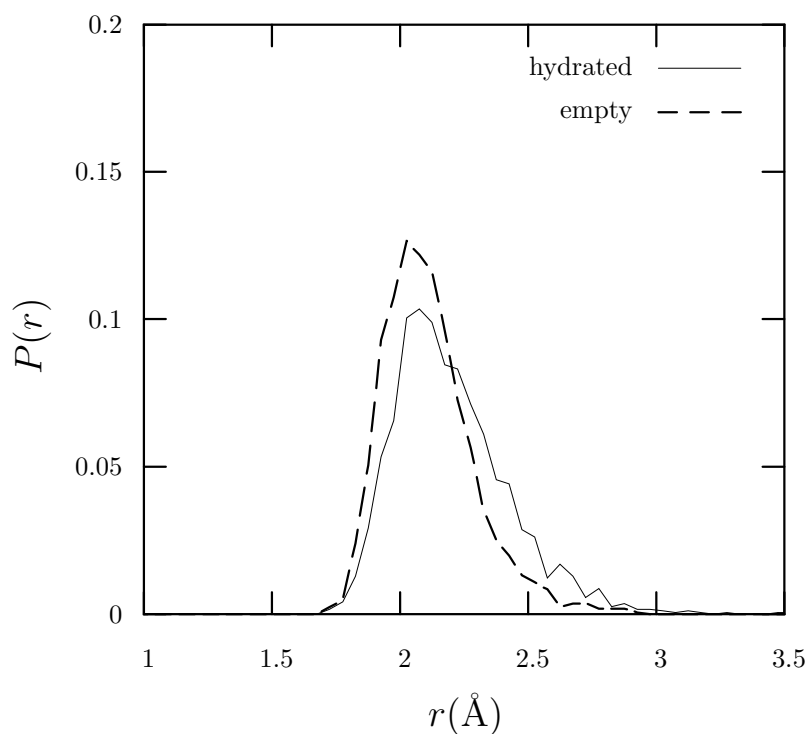


Figure 2.6: Distribution of hydrogen bond lengths between the H on Ala 11 and the O on Phe 7 in both the empty and hydrated simulations of 1BRI.

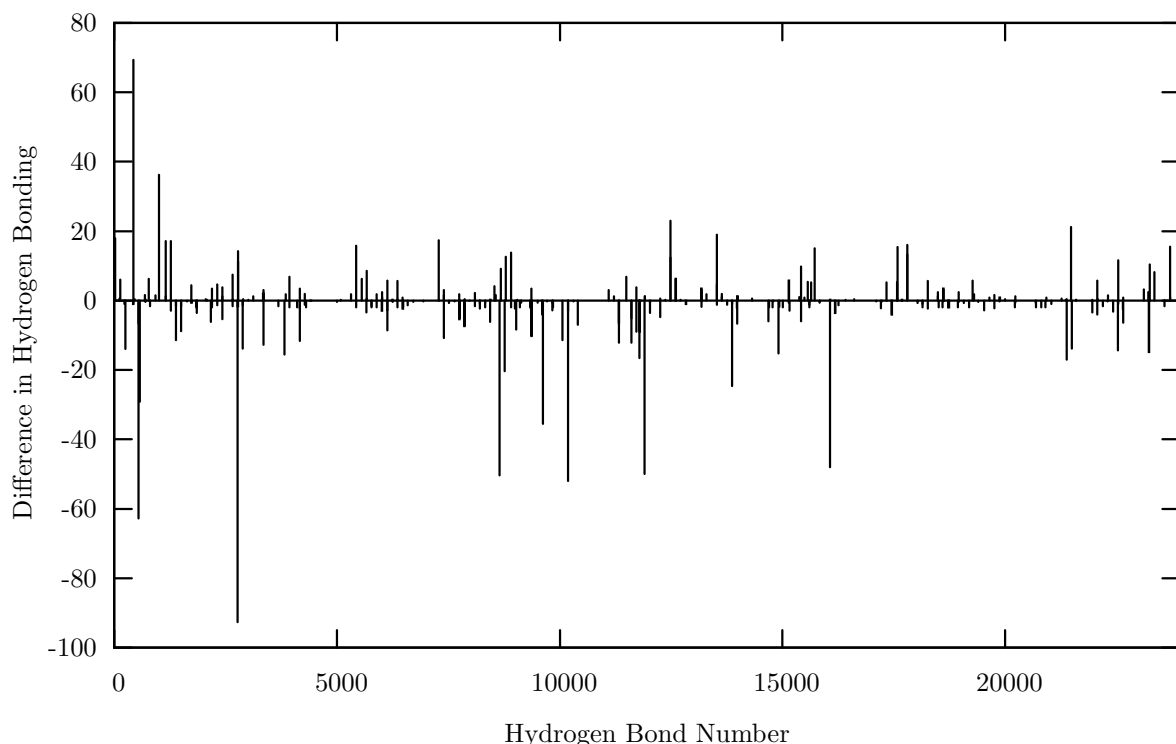


Figure 2.7: Percentage change in time hydrogen bonds exist between the empty and hydrated states of 1BRI.

hydrated states the largest shift is seen in the hydrogen bond between the  $H_{\delta 1}$  of histidine 18 and the O of leucine 14 (hydrogen bond # 2772). This bond falls within the cutoff 92.6% of the simulation for the hydrated state, but never falls within the cutoff when the cavity is empty. There are 5 hydrogen bonds that have over a 40% change in percent simulation time for the hydrated state, but none have that large a change for the empty protein (see Table 2.2). Some of the changes that occur are within residues, that is a different atom in a particular molecule is now participating in the hydrogen bond. There are some secondary structure units that are affected as a result of these hydrogen bond shifts. The hydrogen bond formed between His 18 and Leu 14 in the hydrated structure involved the helix 1. There is a bond formed between  $\beta$ -turns 1 and 2, two bonds formed between  $\beta$ -turns 3 and 4 and one in the transition between  $\beta$ -turns 6 and 7. Secondary

Table 2.2: Barnase hydrogen bonds that show large changes in binding at 1 atm

Hydrogen Bond	Bond #	% Simulation Time	
		Empty	Hydrated
(HIS 18 ND1)---(HIS 18 HD1)...(LEU 14 O)	2772	0	92.6
(ASN 58 ND2)---(ASN 58 HD21)...(LEU 63 O)	10186	36.4	88.4
(LYS 49 NZ)---(LYS 49 HZ2)...(ASP 22 O)	8646	4	54.4
(LEU 63 N)---(LEU 63 H)...(ASN 58 OD1)	11905	41.6	91.6
(SER 80 OG)---(SER 80 HG)...(THR 79 O)	16068	23	71

structure determinations were done using the wiring diagram and assignments from the structure summary in the pdb data bank.<sup>23</sup> While these changes appear to be correlated to the removal of water from the system, they could also be due to changes in the orientation of the terminal amino acids. There was also some shift in solvent exposure (Figure 2.8). The atoms that are solvent exposed in the hydrated structure do not appear to make any hydrogen bonds when the protein cavity is empty. The average length of the hydrogen bonds in the system were calculated for the hydrated and empty proteins, and the shift was minimal and within error. The hydrated proteins bonds are  $2.41 \pm 3$  Å in length while the empty proteins hydrogen bonds are  $2.3 \pm 3$  in length.

The addition of a water molecule may also change the size of the cavity. Cavity volumes can be found by rolling a probe sphere around the van der Waals surface as implemented in the Molecular Surface Package<sup>19</sup> or VOIDOO<sup>20</sup> programs. However, we found that our results, especially for this protein were very sensitive to the size of the probe sphere radius and consistent results using the same probe radius for a sequence of structures even along a single trajectory could not be used. A simple method for estimating the volume is to define the cavity as an irregular polyhedron with vertices

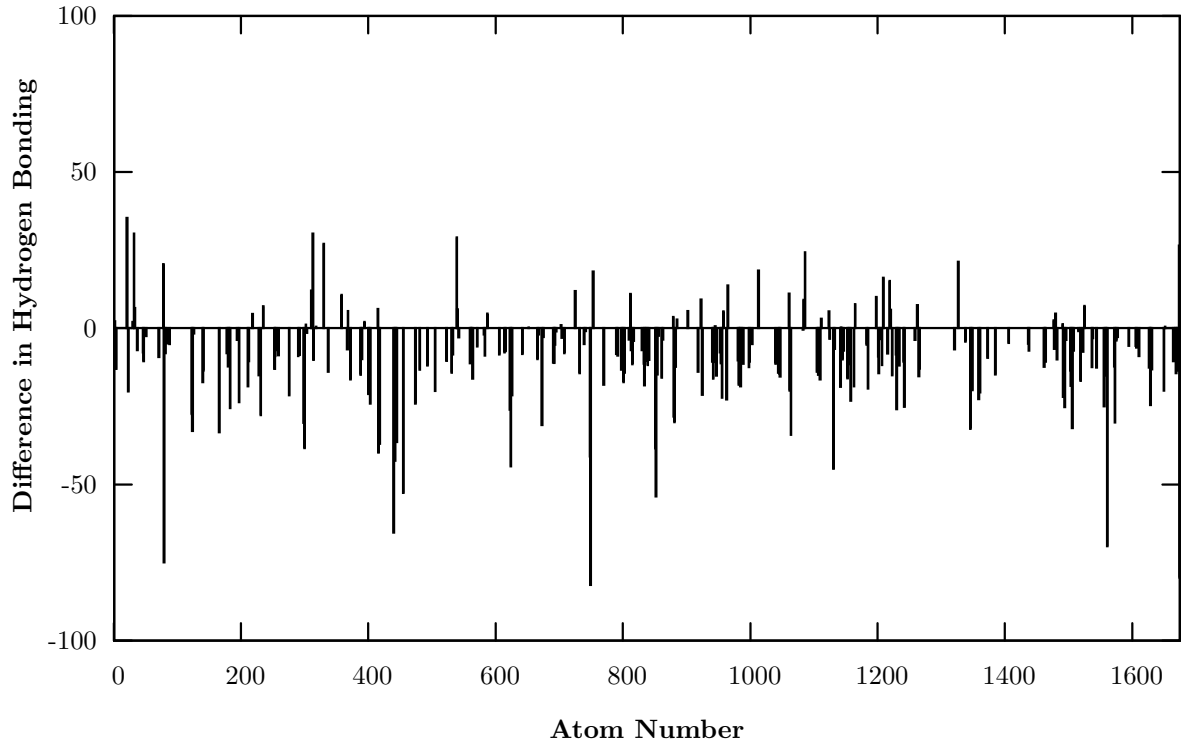


Figure 2.8: Percentage change in time solvent is hydrogen bound between the empty and hydrated states of 1BRI.

defined by atoms on the edges. The volume of the tetrahedron is found from

$$V = \frac{\begin{vmatrix} x_4 & y_4 & z_4 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_1 & y_1 & z_1 & 1 \end{vmatrix}}{6} \quad (2.8)$$

where  $x_i$ ,  $y_i$ , and  $z_i$  are the coordinates of the four atoms<sup>21</sup> For barnase, the polyhedron is defined by 6 points, making two tetrahedron which share the same base. The base was composed of  $C_{\gamma 1}$  Ile 88,  $C_{\gamma 2}$  Ile 96, and  $C_{\delta 1}$  Ile 109. The first tetrahedron used O Phe 7 as its vertex and the second used O Tyr 97 (Figure 2.9). Using this definition the original 3 protein chains from the crystal structure were analyzed. The volume of the cavity as

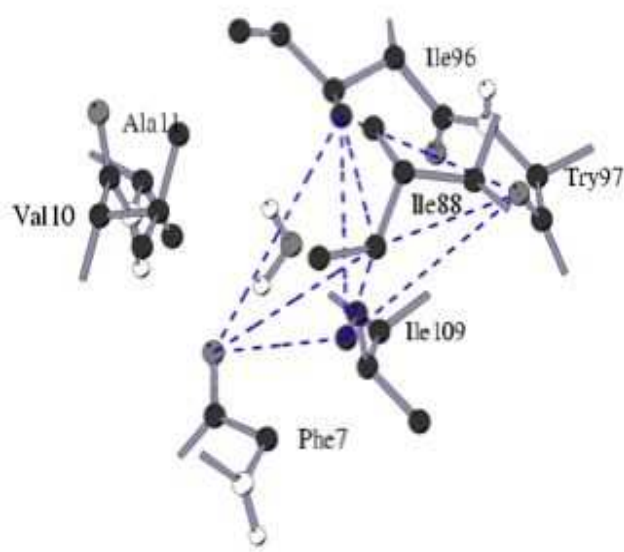


Figure 2.9: Structure of cavity in I76A mutant of barnase showing the polyhedron for the calculation of volume

defined in protein C is  $23.5 \text{ \AA}^3$  and falls between the volumes for A, and B,  $22.6 \text{ \AA}^3$  and  $24.1 \text{ \AA}^3$ , respectively. The similar cavity volumes agrees with the similarity of the three structures as measured by the root-mean-square deviation (rmsd) overlap of the residues within  $6 \text{ \AA}$  of the interior water. The atoms in this region on protein C differed from those on A by  $0.17 \text{ \AA}$  and those on B by  $0.13 \text{ \AA}$ . Determination of cavity volumes for a snapshot taken every ps for the 1700 ( $\lambda = 0$ ) and 2000 ps ( $\lambda = 1$ ) simulations was performed using the simple polyhedron method. The differences between the two cavities are small with a total average volume of  $26 \pm 3 \text{ \AA}^3$  for the hydrated cavity versus  $24 \pm 3 \text{ \AA}^3$  for the empty cavity. The difference between the volumes of the hydrated and empty cavity is smaller than the fluctuations of the cavity. The volumes are about the same as those calculated from the crystal structures. For the tetrahedron defined by the three base atoms and O Phe 7, which is the site the water molecule occupies the majority of the time, the volume difference is  $13 \pm 2 \text{ \AA}^3$  for the hydrated protein versus  $12 \pm 2 \text{ \AA}^3$  for the non-hydrated protein. This is more than half the volume of the polyhedron made up of the two tetrahedra, indicating that this tetrahedron is the larger of the two. The empty cavity is



slightly tighter both in volume and its fluctuations, as given by the standard deviation.

Also of interest is the change in the torsion angle of the isoleucine 88 as the water is removed. As mentioned earlier the rotation of this isoleucine is critical to the formation of the cavity in which the water is found in the x-ray crystal structure. The two torsions of interest are those formed by  $C - C_\alpha - C_\beta - C_{\gamma 1}$  ( $\chi 1$ ) and  $C_\alpha - C_\beta - C_{\gamma 1} - C_{\delta 1}$  ( $\chi 2$ ). In the

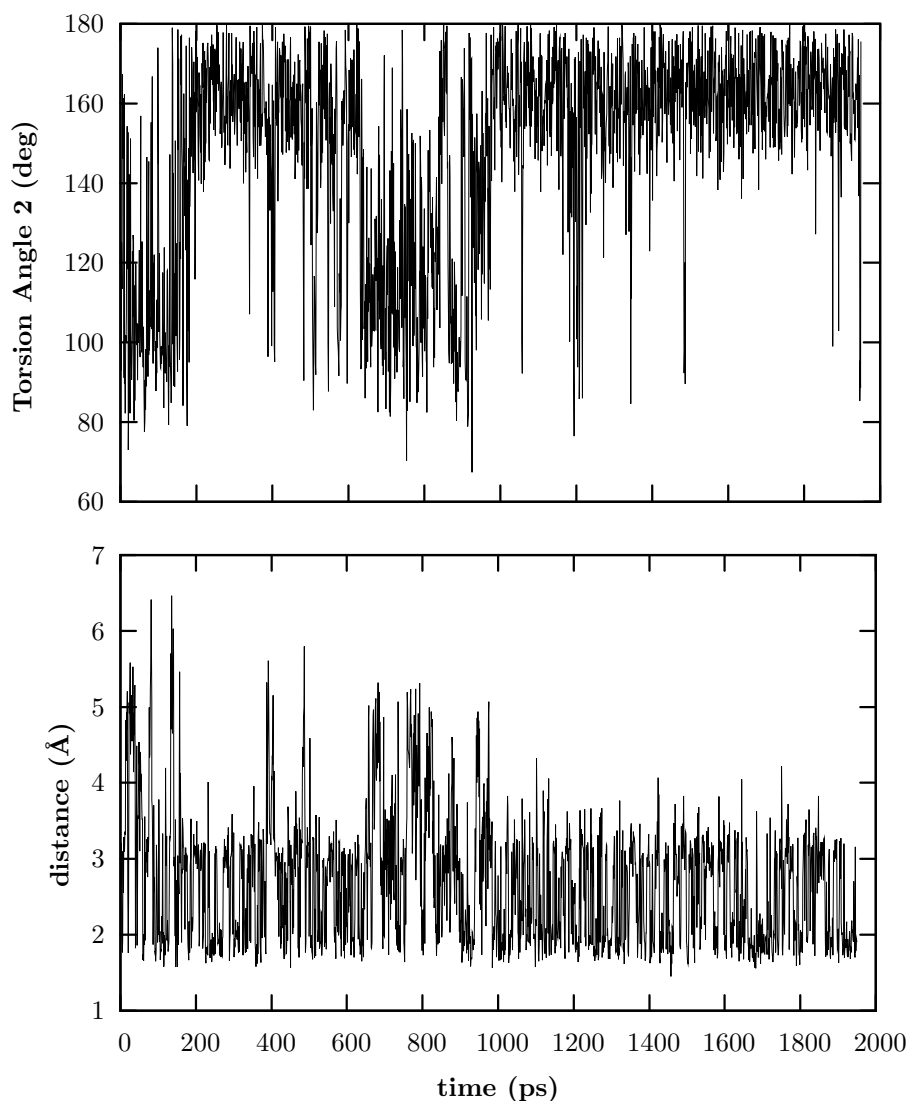


Figure 2.10:  $\chi 2$  versus time.

wild-type structure the average torsion angle across the three proteins found in the unit cell

are 87 degrees for  $\chi_1$  and 168 degrees for  $\chi_2$ . In the mutant structure, once again averaged over the three proteins in the unit cell,  $\chi_1$  is 174 degrees, while  $\chi_2$  is 147 degrees. If the torsions for the trajectory of the hydrated protein, in particular  $\chi_2$ , are analyzed with that of the distance to the O of Phe 7 a trend emerges (Figure 2.10). If the cavity is completely empty,  $\lambda=0$ , the Ile rotates to partially fill the cavity some percentage of the time. The Ile also rotates to occupy the empty cavity when the water moves to be near the Tyr 97. This rotation into the cavity does not align with the wild-type position of the Ile 88. Instead of the Ile orienting such that the CD1 occupies the cavity, it rotates the CG1 and CG2 into the space. When the cavity is empty the torsion angle,  $\chi_2$ , is similar to that of the Ile in the wild-type protein (Figure 2.11). The distribution of torsion angle,  $\chi_1$ , for the empty

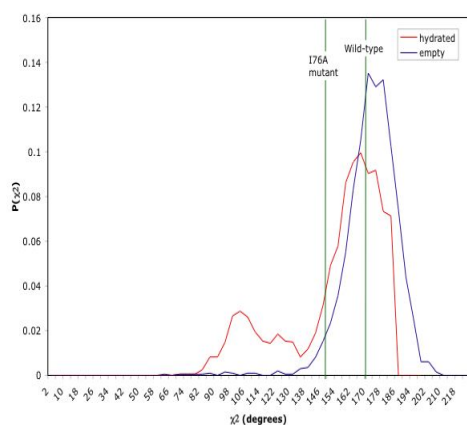


Figure 2.11: Distribution of  $\chi_2$  torsion angle in barnase simulations

cavity indicates a shift away from the angle found in the mutant crystal structure.

However, the new torsion angle is slightly lower than that of the wild-type (Figure ??).

This behavior is only present when the water does not occupy the cavity. Even at  $\lambda = 0.025$  the torsion angle reverts back to that seen in the x-ray crystal structure of 1BRI.

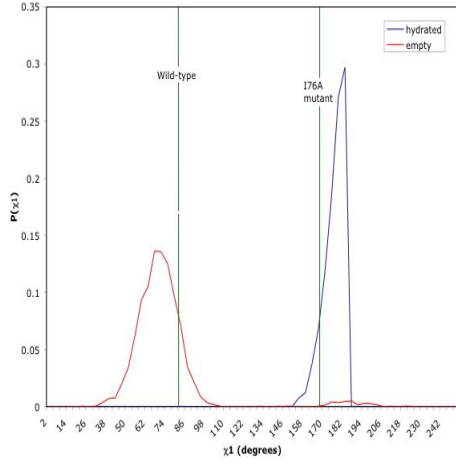


Figure 2.12: Distribution of  $\chi_1$  torsion angle in barnase simulations

## 2.3 Bovine Pancreatic Trypsin Inhibitor (BPTI)

For BPTI our calculations find that  $\Delta G_{protein}$  is greater in magnitude than  $\Delta G_{wat}$  and for this cavity  $\Delta G_{hyd}$  is negative. Our calculations then indicate that the polar cavity in BPTI is hydrated with a significantly negative  $\Delta G_{hyd}$  (Table 2.3).

Table 2.3: BPTI  $\Delta G$  at values at 283, 298, 313K and Protein Entropy and Enthalpy Values. Numbers in parenthesis represent 95% confidence limits.

	Protein	Water	Hydration
$\Delta G$ (kcal/mol) 283K	-11.0(2)	6.36(6)	-4.6(2)
$\Delta G$ (kcal/mol) 298K	-10.9(1)	6.18(3)	-4.7(1)
$\Delta G$ (kcal/mol) 313K	-10.3(2)	5.99(5)	-4.3(2)
$\Delta S$ (cal/mol/K)	-24(9)	12(3)	-12(9)
$-T\Delta S$ (kcal/mol)	7(3)	-3.7(8)	3(3)
$\Delta H$ (kcal/mol)	-19(3)	9.85(8)	-9(3)

Again the temperature dependence of the free energies can be used to calculate entropy. In the case of BPTI we have a large body of literature with which to compare out

data. The unfavorable nature of the entropic component of free energy is expected as the transition is to a well ordered state. There is gas phase mass spectroscopy data on BPTI, where the  $\Delta S$  for the transfer of a water molecule from the vapor to the protein at 0° C is  $-62 \pm 5$  cal/mol/K,<sup>11</sup> giving an entropy change of  $-33 \pm 5$  cal/mol/K for the transfer from the liquid phase to gas phase BPTI.<sup>23,24</sup> This deviation from our calculated value maybe due to the difference in the vibration modes in the two phases, gas phase and solution. It may be binding at a different site in the gas phase experiments.

The study of Fischer *et al.* estimated the  $\Delta S$  from the vibrational normal modes with solvent modeled using a distance dependent dielectric constant. This gives a  $\Delta S$  of -13.4 cal/mol/K for the vibrational entropy difference between hydrated and empty BPTI. If the rotational entropy of gas-phase water is subtracted ( $\Delta S_{rot} = 10.6$  cal/mol/K) then a value of  $\Delta S$  equal to -24 cal/mol/K is found.<sup>9</sup> This entropy change is for a process comparable to Eq. 5.5: a translationally restrained but rotationally free water plus an unhydrated BPTI going to hydrated BPTI. The Fischer, *et al.* estimate of -24 cal/mol/K contains a part due to the ordering of the water molecule in the protein environment and also a contribution from a change in the low frequency vibrational modes, indicating a more flexible protein.<sup>9</sup>

The enthalpy changes are found using  $\Delta H = \Delta G + T\Delta S$ . For BPTI,  $\Delta H_{protein}$  is  $-19 \pm 3$  kcal/mol, which is about the enthalpy change expected upon forming four hydrogen bonds. The measured value for gas phase BPTI is  $-21.3 \pm 1.0$  kcal/mol<sup>11</sup> and a calculated value for gas phase BPTI is -19.8 kcal/mol,<sup>7</sup> indicating that  $\Delta H_{protein}$ , unlike  $\Delta S_{protein}$ , is similar for the gas and liquid phases. The binding site of the water in these experiments is not known, so comparisons with our calculated values should be considered as speculative. The calculated enthalpy change for removing a water molecule from the pure liquid is  $9.85 \pm 0.08$  kcal/mol, which is close to the experimental value of 9.974 kcal/mol.<sup>17</sup> Since the removal of a water molecule from the liquid eliminates about two hydrogen bonds, this value of  $\Delta H_{water}$  is consistent with an enthalpic cost of 5 kcal/mol per hydrogen bond. The

overall enthalpy change for hydrating the cavity in BPTI,  $\Delta H_{hyd}$ , is exothermic.

The isolated water in the BPTI interior makes four hydrogen bonds to the protein. Of these four atoms (the carbonyl oxygen on residues Thr 11 and Cys 38 and the amide hydrogens on residues Cys 14 and Cys 38), only one, the carbonyl oxygen on Thr 11, makes a hydrogen bond to another atom in the protein (Figure 2.13). Our simulations reveal that the four hydrogen bond partners of the interior water do not form any new

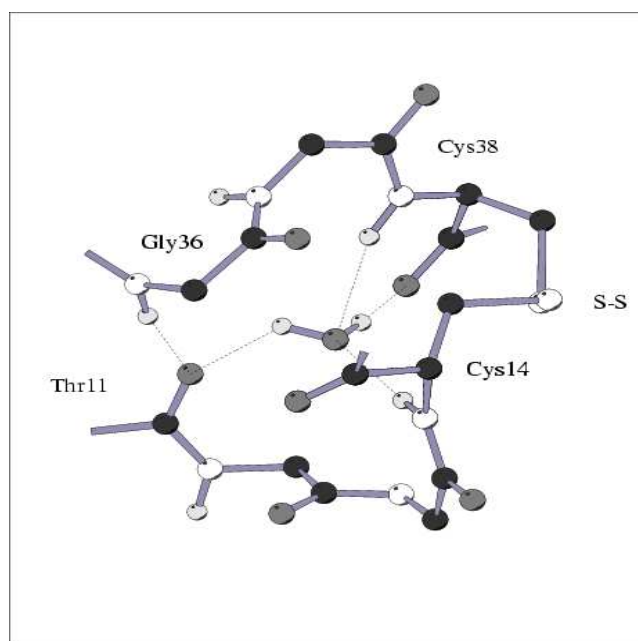


Figure 2.13: The water molecule W122 in the interior of BPTI, showing the hydrogen bonds to the protein backbone. The coordinates are from the 5PTI crystal structure.<sup>3</sup> The figure was made using MOLSCRIPT.<sup>2</sup>

hydrogen bonds as the water molecule is removed. Therefore, the addition of the water to the BPTI cavity creates four new hydrogen bonds. In the pure liquid a water has (about) four hydrogen bonds, but when the water is removed, the solvent can rearrange and remake two hydrogen bond, so only two are lost. The creation of four new hydrogen bonds in the BPTI cavity, as opposed to two lost in the liquid, and the fact that these hydrogen bonds are with atoms that are, for the most part, not involved in other hydrogen bonds, explains the large enthalpy change,  $\Delta H_{hyd}$ . Analysis of the trajectories at  $\lambda = 1$ , indicate that the water protons hydrogen bonded to the O on Thr 11 and the O on Cys 38 do not flip during

the 2500 ps of our simulations, (Figure 2.14), this is agreement with the literature which indicates the time scale for hydrogen exchange in BPTI will be on the order of 45 ns.<sup>26</sup> Both of the water protons are within 2.8 Å of their respective hydrogen bond donors for more than 99.5% of the simulation length. The water O-Cys 38 amide H distance is within 2.8 Å 98.2% of the time. The water O-Cys 14 amide H distance is within 2.8 Å only 82.5 % of the time, indicating that this hydrogen bond is the least stable of the four hydrogen bonds.

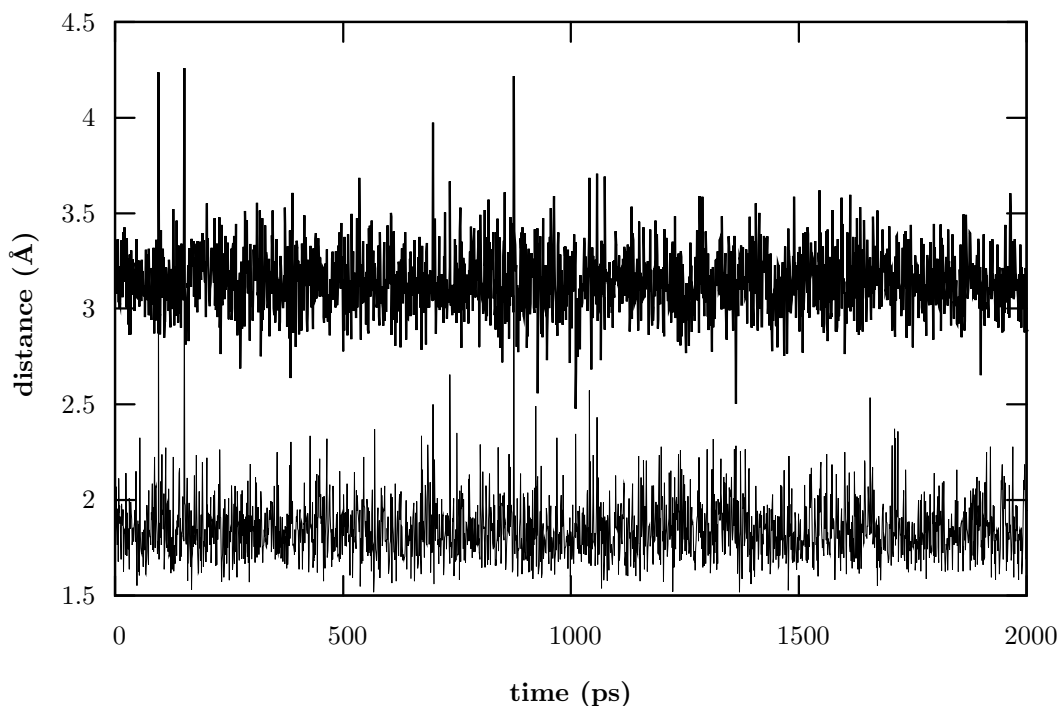


Figure 2.14: Hydrogen bond distance between the two protons of the cavity water (WTP) and the oxygen of Thr 11.

A look at the  $\langle \Delta r^2 \rangle$  as a function of the parameter  $\lambda$  ( $\lambda=0$  is the empty and  $\lambda=1$  is the hydrated state) shows that its value increases as the water-protein interactions are scaled in (Figure 2.15). These results suggest that the protein gets more flexible when interior cavities are occupied. The change in  $\langle \Delta r^2 \rangle$  is larger for BPTI than it was for barnase, perhaps because the water interacts more strongly with BPTI than it does with barnase. The values of  $\langle \Delta r^2 \rangle$  are smaller for the atoms close to the water, indicating that

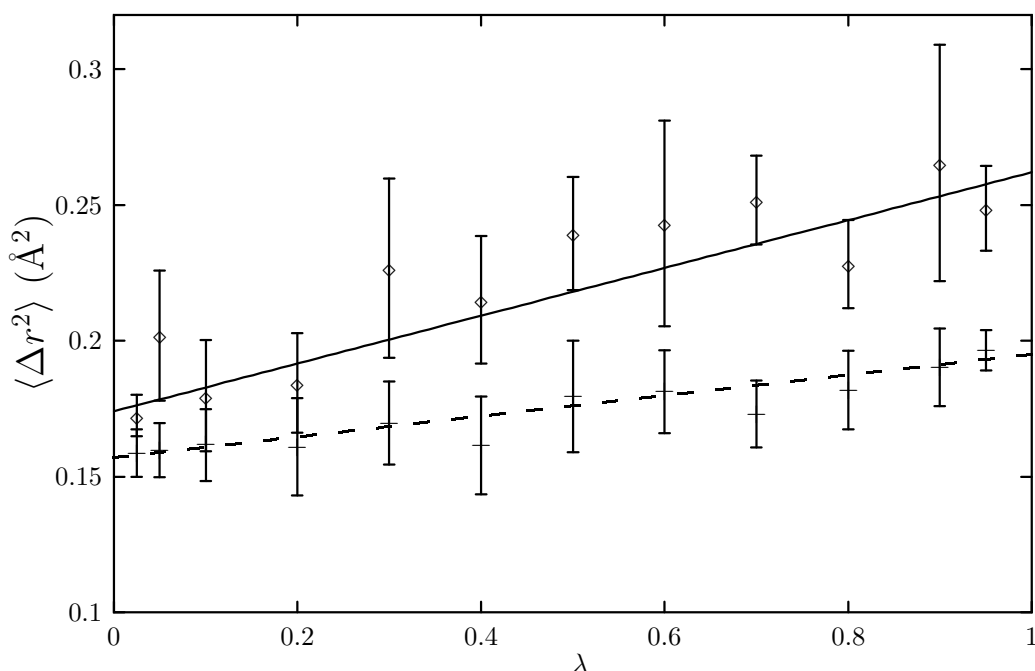


Figure 2.15: Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6  $\text{\AA}$  of the cavity water (+) from the reference structure of 5PTI. Errors reported are 95% confidence limits.

this is a more rigid than average region of the proteins, and also the mean square deviation of the heavy backbone atoms close to the water changes *less* than the the average of all atoms. The  $\langle \Delta r^2 \rangle$  for all backbone atoms increases by  $0.088 \text{ \AA}^2$  and for the close atoms it increases by  $0.038 \text{ \AA}^2$  with hydration.

The increase in protein flexibility seen on hydration may be understood at least in part by looking at the hydrogen bonds formed near the cavity. Of the four protein atoms which form hydrogen bonds to the water in the BPTI cavity, only one of these atoms (Thr 11 O) forms a second hydrogen bond with another protein atom. As the water molecule is removed from the cavity, the bond between the O on Thr 11 and the H on Gly 36 becomes slightly tighter with the average length decreasing to  $1.92 \text{ \AA}$  from  $2.05 \text{ \AA}$  (Figure 2.16). It appears as we saw in the case of the barnase cavity that the hydrogen bonds between protein atoms at the surface of the cavity get longer due to the presence of the buried water. Although the local hydrogen bond stretching helps to explain the short range effects

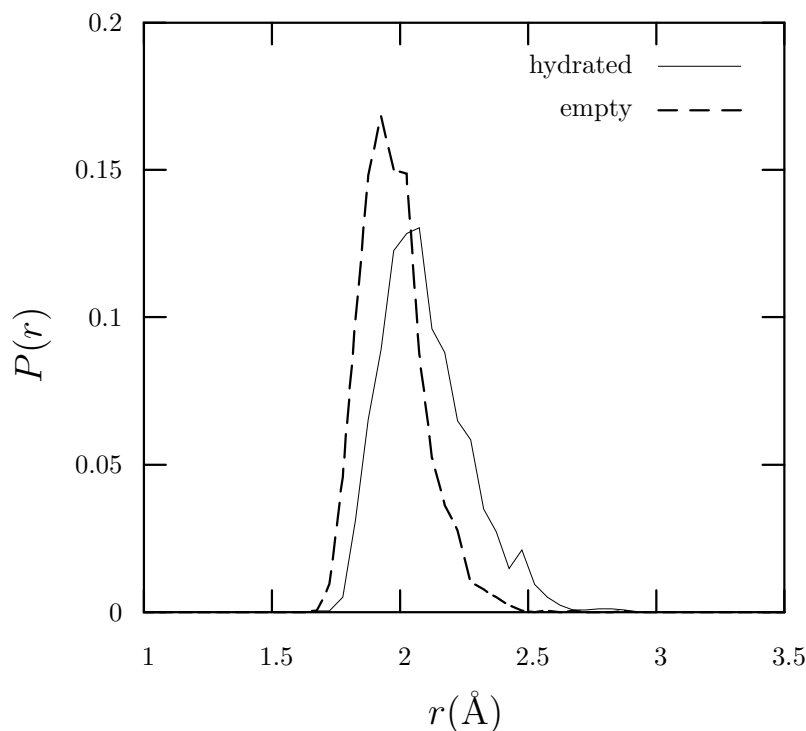


Figure 2.16: Distribution of hydrogen bond lengths between the H on Gly 36 and the O on Thr 11 in both the empty and hydrated simulations of BPTI.

of removing the solvent water the changes are broader in scope. A plot of the  $\langle \Delta r^2 \rangle$  values for the hydrated and empty cavities shows that the change in hydrogen bonding is almost global. Only atoms NH1 of Arg 17, C $_{\delta}$  and C $_{\gamma}$  of Lys 15, O $_{\delta 2}$  of Asp 3, C $_{\delta 2}$  of Leu 6, and C $_{\gamma 2}$  of Thr 32 are more flexible when the cavity is empty and these differences are very slight, less than 0.2 Å<sup>2</sup> (Figure 2.17). Of these only three are in residues with distinct secondary structure roles<sup>27</sup> Asp 3 and Leu 6 are part of the G(3,10) helix secondary structure and Thr 32 is part of an extended  $\beta$ -sheet. There are 60 atoms with a  $\langle \Delta r^2 \rangle$  more than 0.2 Å<sup>2</sup> more flexible when the protein is hydrated. While the largest changes occur at the C terminus, the heavy atoms of Arg 39 and Lys 46 all experience changes greater than 1.0 Å<sup>2</sup>.

While there are no hydrogen bonds formed between the amino acids bordering the cavity there are changes in overall hydrogen bonding. Hydrogen bond numbers were



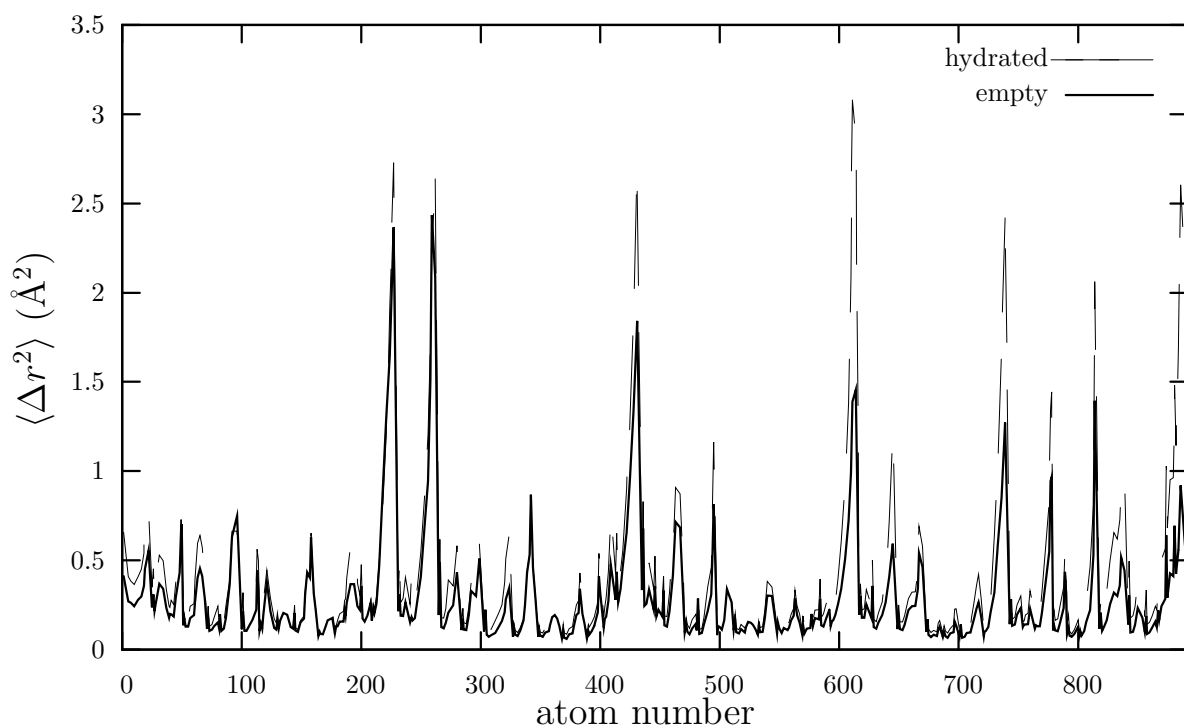


Figure 2.17: Mean square deviation of heavy atoms in BPTI from the reference structure. Errors reported are 95% confidence limits.

assigned to every possible hydrogen bond in the protein system. The percentage of the simulation each bond fell within the  $50^\circ$  angular cutoff and the 4 Å donor to heavy atom acceptor distance were calculated. In order to view only those hydrogen bonds experiencing changes in residency time, the total percent residencies for the hydrated bonds were subtracted from those of the empty state (Figure 2.18). Using this method both weaker hydrogen bonds and strong bonds that don't exhibit large scale changes will be found near the baseline. A bond that has over a 40% difference in hydrogen bond residency time is considered to be formed or lost during hydration. There are two bonds that appear to be present when the protein cavity is empty, but not when its hydrated, and two that are in the hydrated protein but not the empty (see Table 2.4) Three of the hydrogen bonds lost/formed upon hydration involve a bond between helix 1 and the amino acids bordering the structure. The secondary structure in question here is the G(3,10)

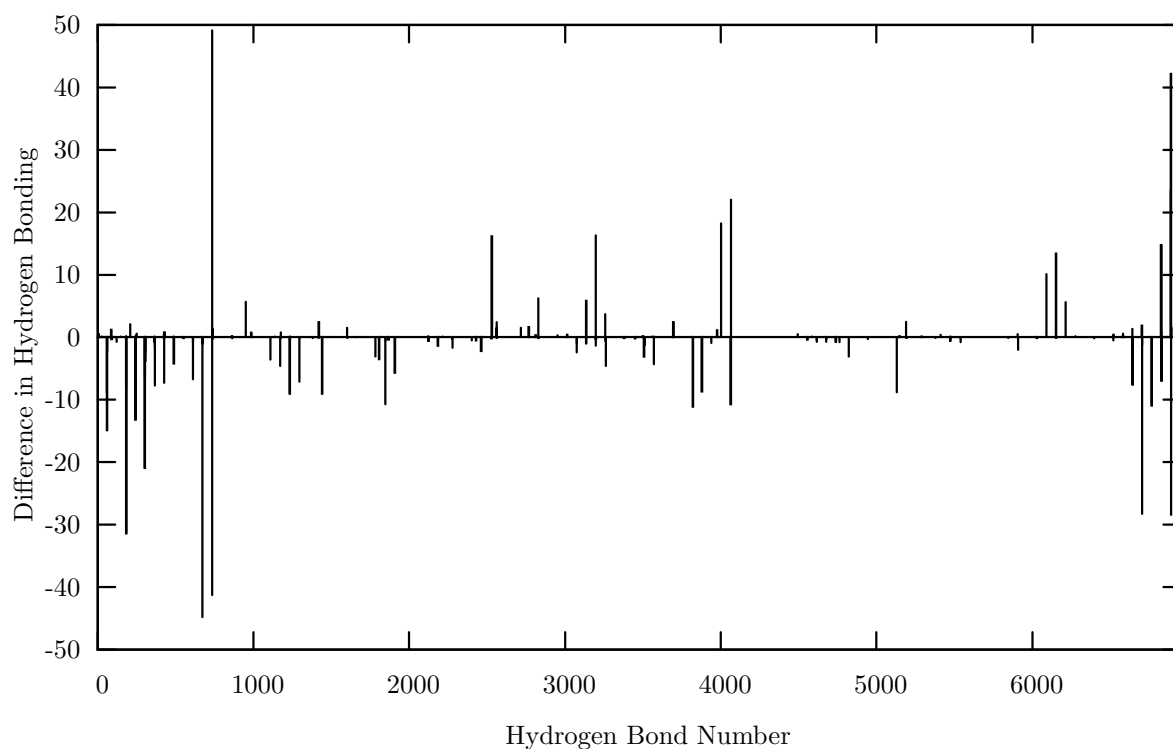


Figure 2.18: Percentage change in time hydrogen bonds exist between the empty and hydrated states of BPTI.

helix.<sup>23</sup> It is unknown what ramifications such a change in bonding would have on the stability of the structure. The only other secondary structures involved near the C-terminus and these bonds were weak in both states.

A treatment for an estimate of the cavity volumes similar to that employed for the barnase mutant was for BPTI. A single irregular tetrahedra was constructed using vertices that are the sites involved in hydrogen bonding: O Thr 11, O Cys 38, N Cys 14, and N Cys 38. As the water is in a stable hydrogen bond network between these 4 atoms, using them as a definition of the volume the water can occupy is not unreasonable (Figure 2.19). The volumes were again calculated for snapshots taken over a trajectory and averaged. The hydrated cavity had a volume of  $8.78(0.69) \text{ \AA}^3$  while the empty cavity had a slightly larger average volume of  $9.27(0.80) \text{ \AA}^3$ . The hydrated cavity does appear to be slightly smaller, perhaps due to the attractive forces between the water and the protein. The cavity size in

Table 2.4: BPTI hydrogen bonds that show large changes in binding at 1 atm

Hydrogen Bond	Bond #	% Simulation Time	
		Empty	Hydrated
(GLU 7 N)···(GLU 7 HD1)···(PHE 4 O)	736	75.2	26.1
(GLY 57 N)···(GLY 57 HD21)···(ARG 53 O)	6889	42.3	0.1
(LEU 6 N)···(LEU 6 HZ2)···(PRO 2 O)	673	8.4	53.2
(GLU 7 N)···(GLU 7 H)···(ASP 3 O)	735	16.7	51.8

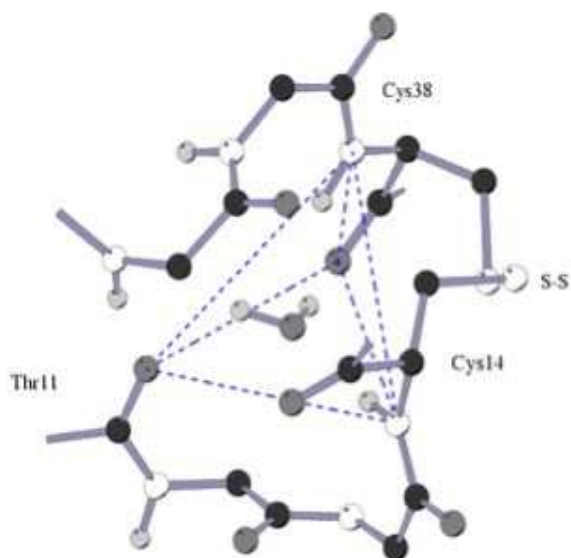


Figure 2.19: Structure of cavity in BPTI showing the polyhedron for the calculation of volume

the crystal structure is  $8.92 \text{ \AA}^3$ . For comparison, the volume of the tetrahedron made up of nearest neighbors in an ice crystal is  $10.8 \text{ \AA}^3$  (using an oxygen-oxygen nearest neighbor distance of  $2.760 \text{ \AA}$ ).<sup>28</sup> Comparisons between the cavities are complicated by the fact that different atom types are used to define them (either carbon, nitrogen, or oxygen atoms) and the atoms will have their own excluded volumes, leaving a different amount of free volume for the water molecule.

## 2.4 Lysozyme I106A Mutant

The x-ray crystal structure of the lysozyme mutant shows the interior water interacting with the surrounding amino acids via 2 hydrogen bonds, one to the oxygen of asparagine 27 and the second through the proton of the epsilon nitrogen of tryptophan 112 (Figure 2.20). As this is an intermediate hydrogen bond environment the values of  $\Delta G_{hyd}$  should

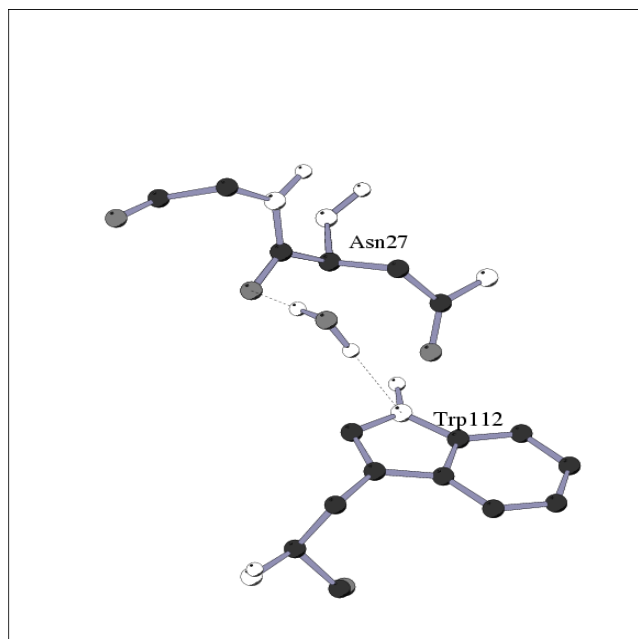


Figure 2.20: The water molecule W445 in the interior of the I106A mutant of lysozyme, from the 2HEA crystal structure.<sup>15</sup> The figure was made using MOLSCRIPT.<sup>2</sup>

fall between those of the two previous proteins studied. Our calculations show that while the free energy of hydration is more favorable than that calculated for the barnase mutant, the protein should still remain unhydrated, as  $\Delta G_{hyd}$  is positive (Table 2.5).

Table 2.5: Lysozyme  $\Delta G$  at values at 283, 298, 313K. Protein Entropy and Enthalpy Values Numbers in parenthesis represent 95% confidence limits.

	Protein	Water	Hydration
$\Delta G$ (kcal/mol) 283K	-4.0(2)	6.36(6)	2.4(2)
$\Delta G$ (kcal/mol) 298K	-3.5(1)	6.18(3)	2.7(1)
$\Delta G$ (kcal/mol) 313K	-3.4(3)	5.99(5)	2.6(3)
$\Delta S$ (cal/mol/K)	20(10)	12(3)	32(10)
$-T\Delta S$ (kcal/mol)	-6(3)	-3.7(8)	-10(3)
$\Delta H$ (kcal/mol)	2(3)	9.85(8)	12(3)

What is interesting about the lysozyme mutant, is that this overall increase in free energy of hydration may not be do to the enthalpic benefit of the additional hydrogen bond. Rather, it might be do to a large increase in the entropic component to free energy. The entropic component,  $\Delta S_{hyd}$ , is  $32 \pm 10$  cal/mol/K twice that calculated for the barnase mutant. This is not in agreement with the more ordered state imposed by two hydrogen bonds versus one. The process is still under enthalpic control with a  $\Delta H_{hyd}$  of  $12 \pm 3$  kcal/mol, and is therefore endothermic. Also of interest is this value for the enthalpy as it is close to the value seen in the barnase mutant if not slightly more endothermic. This data combines to make the two hydrogen bond model of this system unlikely.

If the coordinate trajectories of the hydrated states are analyzed for binding between the protons of the water and their hypothetical hydrogen bond partners, it becomes clear that they are not forming stable bonds within this system. The water is bound to the O of Asn 27 for just under 30% of the simulation and to the Trp 112 only 9% of the trajectory (see Table 2.6). A third hydrogen bond is formed to the delta oxygen of Asn 27 and exists approximately 12% of the simulation.

Table 2.6: Percent of simulation hydrogen bond exists between buried water and surrounding cavity.

Donor	Acceptor	% Simulation Time
H WAT	O Asn 27	29.8
O WAT	HE1 Trp 112	9
H WAT	OD1 Asn 27	12.2
H WAT	O Ala 106	3.4
H WAT	O Trp 28	0.2

This lack of stable hydrogen bonds is the likely cause of the high entropy value, and as there are fewer bonds than predicted the enthalpy value is not unexpected.

The flexibility of the protein does increase with the addition of a water into the protein cavity although this shift is slight. The slope of the increase is  $0.03 \pm 0.01 \text{ \AA}^2$  both for those atoms within 6  $\text{\AA}$  and for the full protein. If only the protein backbone is considered the dependence on hydration is even more slight (Figure 2.21). Unlike the barnase mutant and BPTI it is difficult to determine a hydrogen bond in the region of the cavity that alters as a result of the hydration. This problem is due at least in part to the lack of water interactions with the cavity wall. The bonds throughout the system reflect this lack of hydration dependence (Figure ??). The changes in the duration of hydrogen bonds between the two states rarely exceeded 15% of the simulation time, also these hydrogen bonds existed for only small sections of the simulation (Table ??).

Table 2.7: Lysozyme hydrogen bonds that show changes in binding at 1 atm greater then 10%

Hydrogen Bond	Bond #	% Simulation Time	
		Empty	Hydrated
(ARG 62 NE) $-\cdot-$ (ARG 62 HE) $\cdot\cdot\cdot$ (PRO 71 O)	19218	13	0
(ALA 76 N) $-\cdot-$ (ALA 76 H) $\cdot\cdot\cdot$ (TRP 64 O)	22690	13.6	1.2
(ARG 62 NH2) $-\cdot-$ (ARG 62 HH22) $\cdot\cdot\cdot$ (ASP 49 O)	18758	13.7	1.7
(ARG 62 ND1) $-\cdot-$ (ARG 62 HH11) $\cdot\cdot\cdot$ (PRO 71 O)	18928	12	0
(TRP 109 NE1) $-\cdot-$ (TRP 109 HE1) $\cdot\cdot\cdot$ (GLN 58 O)	31238	16.1	4.1
(ASN 75 N) $-\cdot-$ (ASN 75 H) $\cdot\cdot\cdot$ (ARG 62 O)	22543	0	13.5
(ALA 76 N) $-\cdot-$ (ALA 76 H) $\cdot\cdot\cdot$ (TYR 63 O)	22689	3.1	19.1

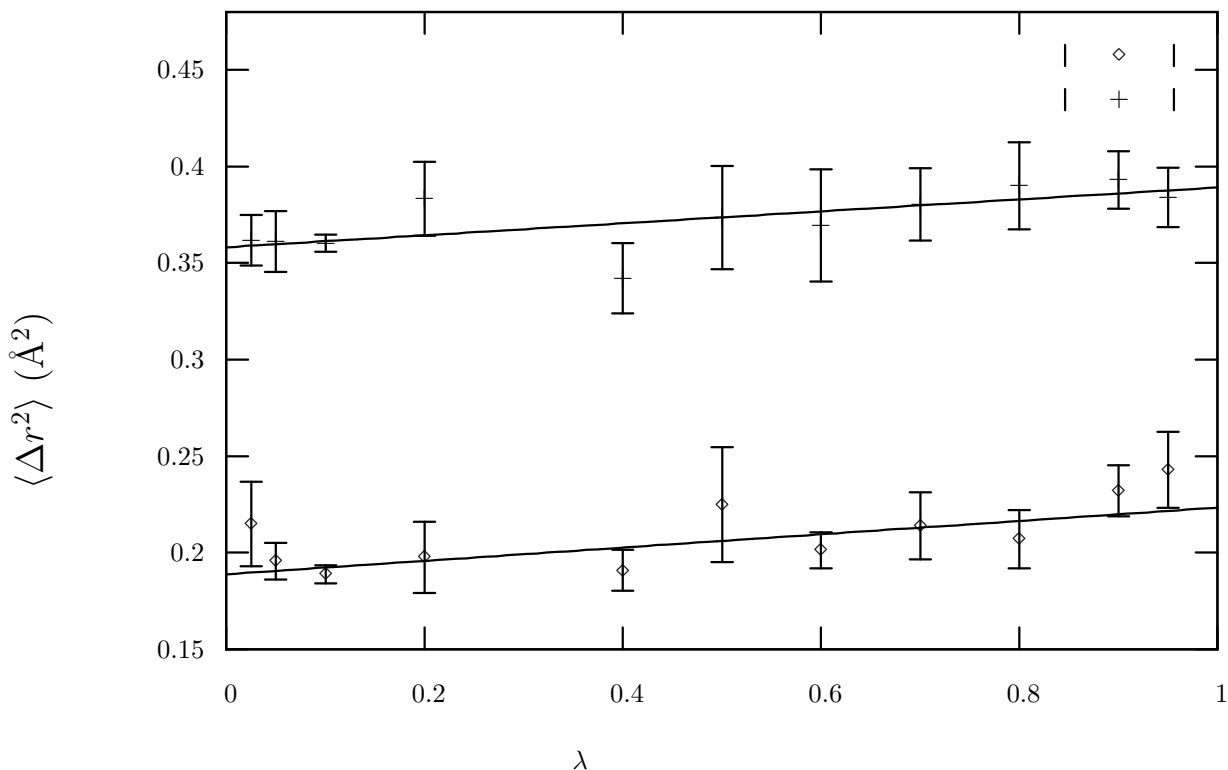


Figure 2.21: Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6 Å of the cavity water ( $+$ ) from the reference structure of 2HEA. Errors reported are 95% confidence limits.

Finding a volume to define the space the water occupies in the lysozyme mutant was difficult given its positional disorder. Rather than attempt a simple irregular tetrahedron, two fused tetrahedra were used to define the space, in a similar manner to the barnase mutant. These vertices were defined as  $C_\beta$  Leu 31,  $N_{\epsilon 1}$  Trp 112,  $C_{\gamma 2}$  Ile 23, O Asn 27, and CB Ala 106 (Figure 2.23). The result for the size of the cavity indicates that shifts in the overall protein structure are occurring as a result of the addition of a water. The cavity volume in the hydrated protein is  $30 \pm 4$  Å<sup>3</sup> while the cavity in the empty protein appears slightly larger at  $36 \pm 8$  Å<sup>3</sup>. While these values have overlapping error bars, magnitude of these error bars themselves indicate that changes are occurring inside the protein cavity. Unfortunately as with the other two proteins the radius of gyration data

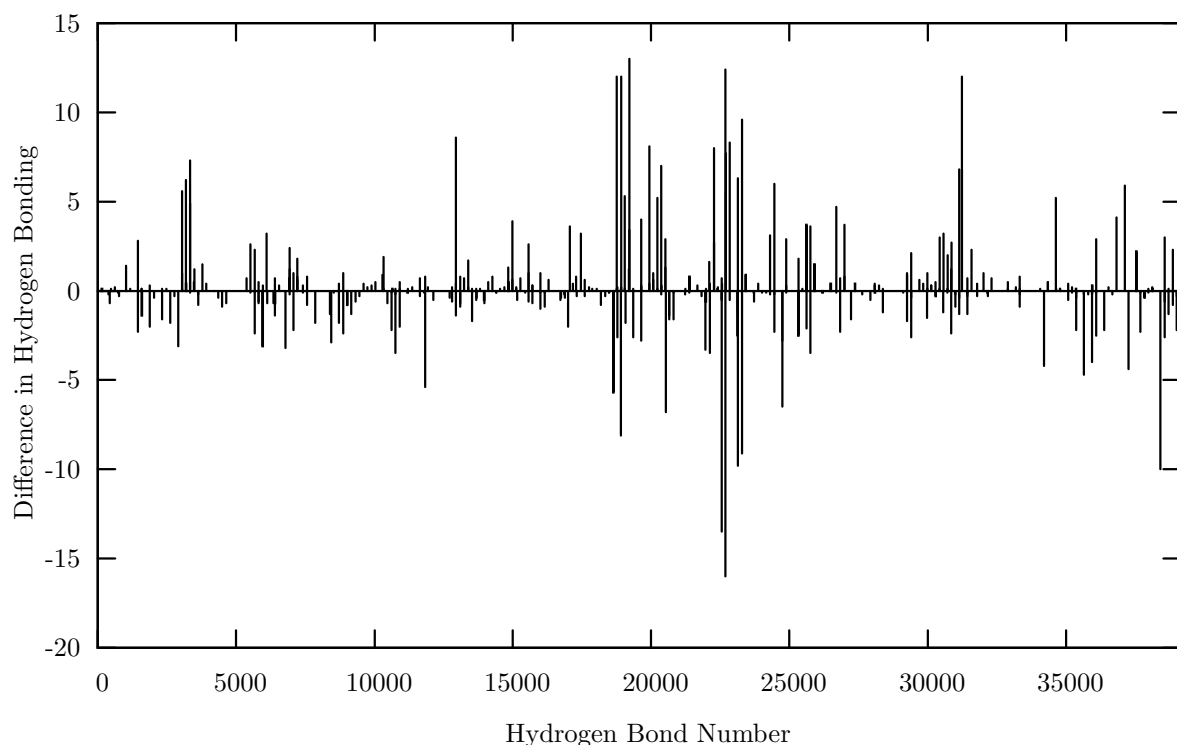


Figure 2.22: Percentage change in time hydrogen bonds exist between the empty and hydrated states of Lysozyme.

does not provide meaningful data as the values are statistically identical.

## 2.5 Fluctuating Charge

The simulations performed for the three proteins were done without including polarizability. There is data that suggests that the interior of proteins are in fact more polarizable than molecules on their own. If the potential energy minima for the interaction of the TIP3P water model used for these simulations and methane is calculated it is equal to -0.28 kcal/mol. This result is higher than the value calculated using *ab initio*, -0.71 kcal/mol.<sup>5</sup> This weak water-methane interaction indicates that the non-polarizable model underestimates the interaction between water and non-polar residues. One manner used to determine the effect polarizability might have on the  $\Delta G_{hyd}$  measurements is to scale the



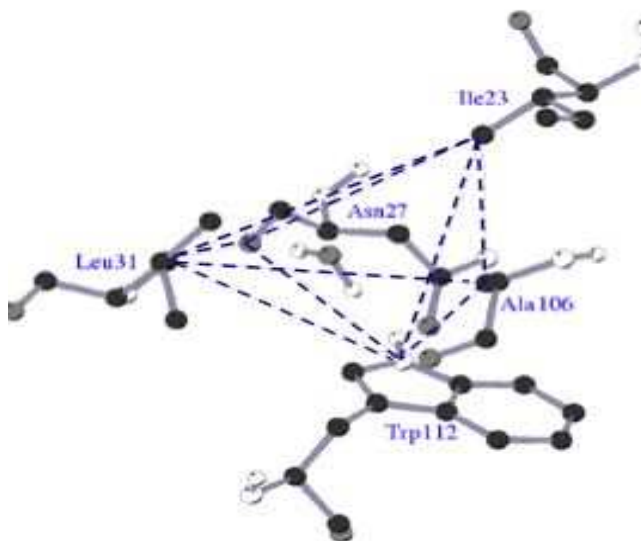


Figure 2.23: Structure of cavity in I106A mutant of Lysozyme showing the polyhedron for the calculation of volume

Lennard-Jones well depth using  $\epsilon_{ij} = \epsilon_{ij}^0(1 + \lambda)$  by 20%. These studies, performed on barnase, show that even smaller changes can have a marked effect on the enthalpy and therefore the free energy of hydration (Figure 2.24)

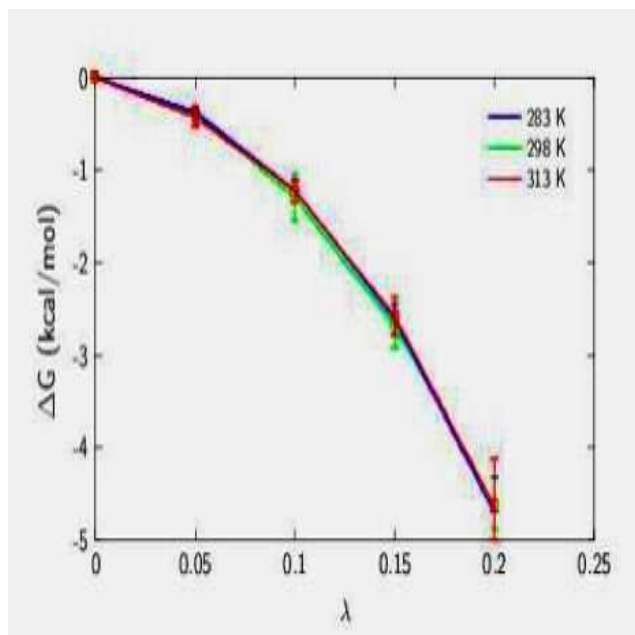


Figure 2.24: Free energy of hydration as the Lennard Jones well depth is scaled by 20%

## 2.6 Summary

The cavity in BPTI is an exothermic process and the free energy of hydration is negative.

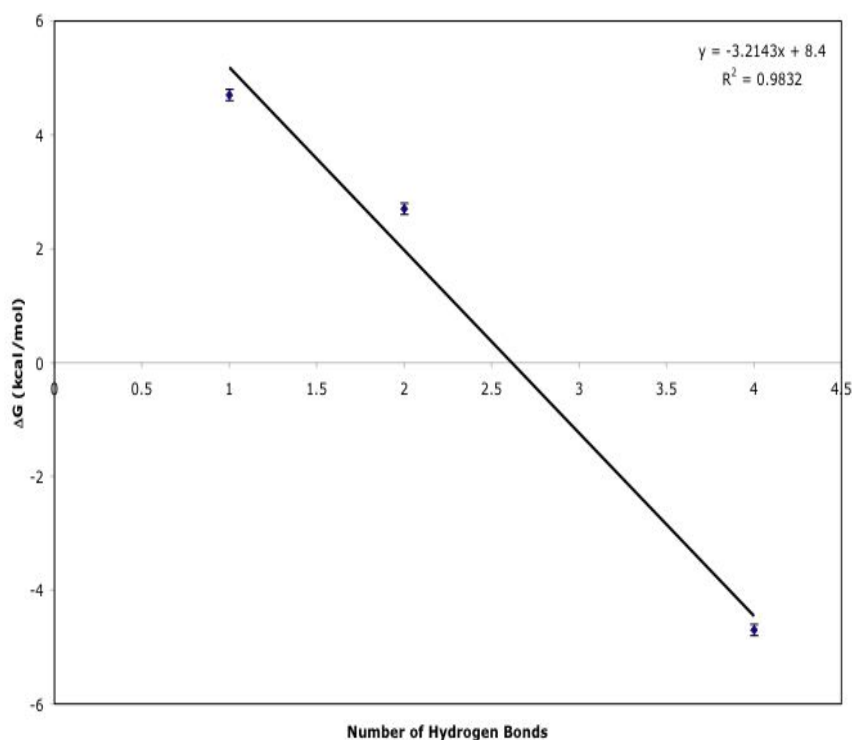


Figure 2.25:  $\Delta G_{hyd}$  versus available hydrogen bonds

However, despite the x-ray crystal structures that indicate hydration, the cavities in the barnase and lysozyme mutants are sufficiently endothermic to make the free energy of hydration positive. In the case of barnase the  $\Delta G_{hyd}$  is of sufficient magnitude that changing the protein potential in any way is unlikely to change the hydration state. The large positive value for barnase  $\Delta G_{hyd}$  combined with the torsion data for Ile 88 creates

doubt as to the presence of the water indicated in the crystal structure. The hydration of the cavity in lysozyme is somewhat less clear cut with a  $\Delta G_{hyd}$  of less than 3 kcal/mol. The lack of hydrogen bonding in this cavity despite their availability might be resolved through changing the protein potential. However, while this would decrease the enthalpy, it would also have decreased the entropy. This limits the effect polarizability has on the  $\Delta G_{hyd}$ .

There is a superficial correlation between the  $\Delta G_{hyd}$  and the number of hydrogen bonds indicated in the x-ray crystal structure (Figure 2.25) however, even without the lysozyme hydrogen bond data, the lack of correlation between the number of bonds and the  $\Delta S$  would have indicated that hydrogen bonding was not sufficient to describe the

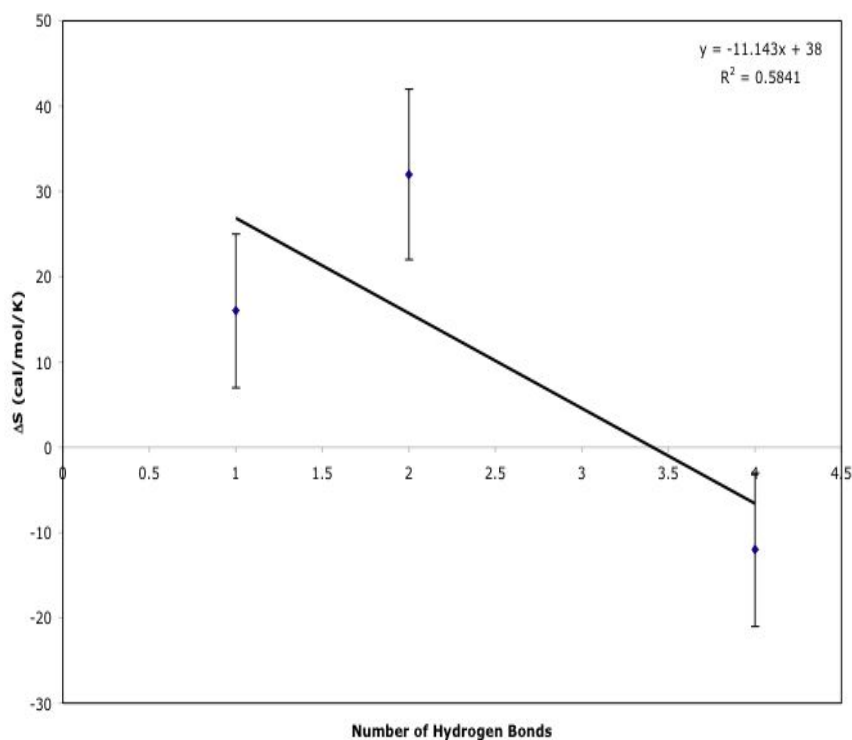


Figure 2.26: Entropy versus available hydrogen bonds

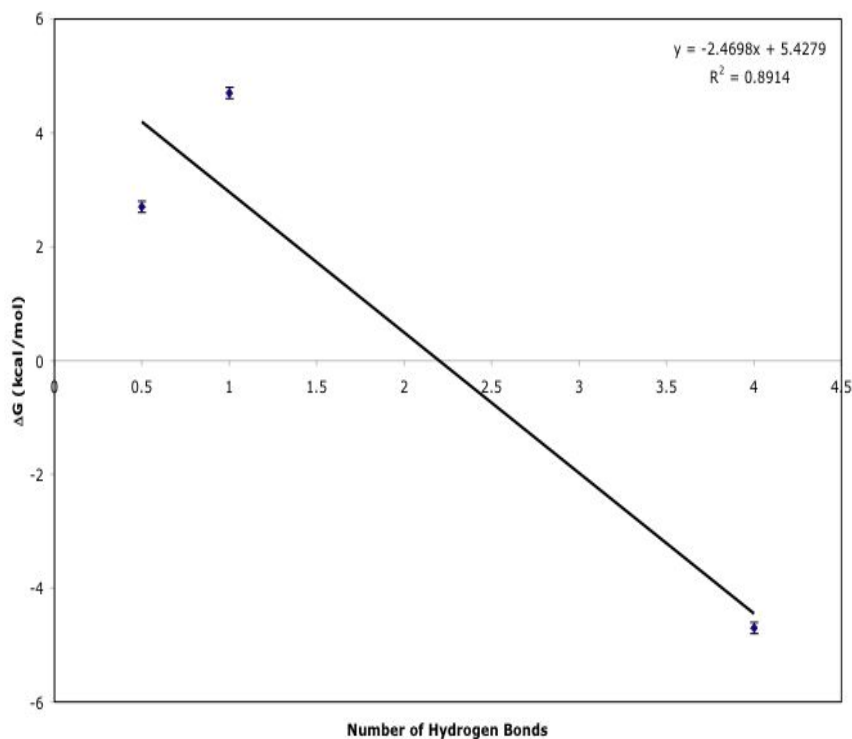


Figure 2.27:  $\Delta G_{hyd}$  versus available hydrogen bonds (Lysozyme = 0.5 hydrogen bonds) behavior of the cavity (Figure 2.26). The data is much better correlated with the number of hydrogen bonds when lysozyme is treated as having 0.5 hydrogen bonds instead of 2, which is consistent with our simulations (Figure 2.27 and 2.28).

The role of entropy in determining hydration should not be ignored. This helps to explain why the hydrophobic cavities such as in Cytochrome C Oxidase are hydrated with more than one water molecule.<sup>30</sup> In that enzyme the extent of hydration has been determined by studies of activity, in that the enzyme does not function without the water present. Multiple waters in a hydrophobic environment can hydrogen bond to each other, lowering  $\Delta H$ , while their disorder in the system will allow them to retain a large  $\Delta S$ .

Our flexibility results indicate that even in lysozyme, with its marginally interacting water, the flexibility of the protein increases with hydration. This change is greatest for BPTI, in agreement with the Smith *et al.*<sup>9</sup> While the flexibility does increase with hydration the change, is not isolated to the region surrounding the cavity. Rather, in all three cases any change observed was for the full protein. There were local areas of where hydrogen bonds were formed or lost upon hydration, but there was no large scale changes in protein structure.

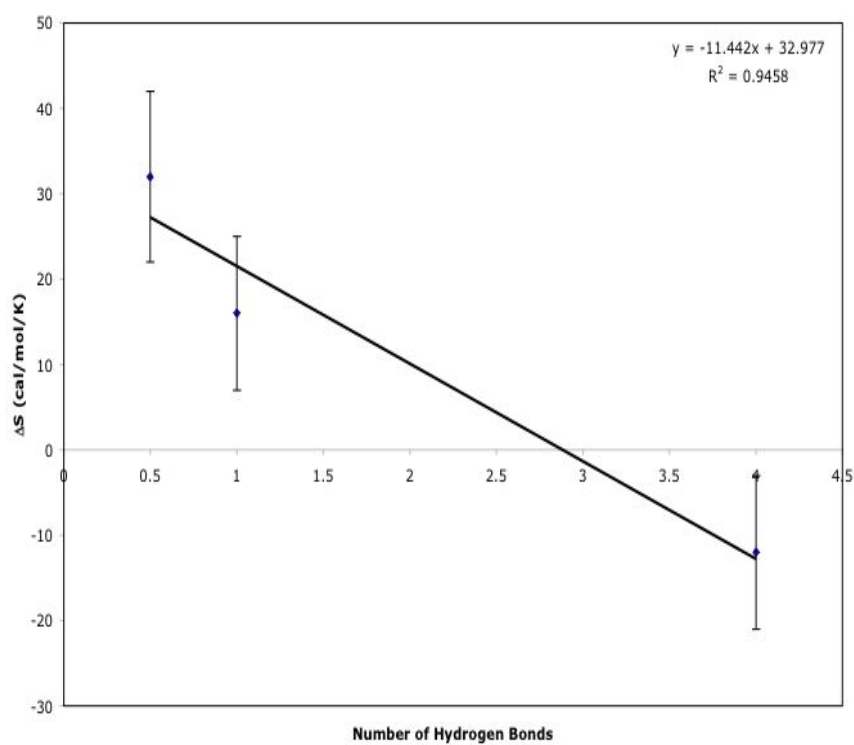


Figure 2.28: Entropy versus available hydrogen bonds (Lysozyme = 0.5 hydrogen bonds)

# Bibliography

- [1] Buckle, A. M.; Cramer, P.; Fersht, A. R. *Biochemistry* **1996**, *35*, 4298–4305.
- [2] Kraulis, P. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.
- [3] Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301–329.
- [4] Takano, K.; Funahashi, J.; Yamagata, Y.; Fujii, S.; Yutani, K. *J. Mol. Biol.* **1997**, *274*, 132–142.
- [5] Zhang, L.; Hermans, J. *Proteins: Struct. Funct. Gen.* **1996**, *24*, 433–438.
- [6] Otting, G.; Liepinsh, E.; Halle, B.; Frey, U. *Nat. Struct. Biol.* **1997**, *4*, 396–404.
- [7] Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Gronenborn, A. M.; Clore, G. M. *Science* **1995**, *267*, 1813–1817.
- [8] Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Gronenborn, A. M.; Clore, G. M. *Science* **1995**, *270*, 1848–1849.
- [9] Yu, B.; Blaber, M.; Gronenberg, A. M.; Clore, G. M.; Caspar, D. L. D. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 103–108.
- [10] Matthews, B. W.; Morton, A. G.; Dahlquist, F. W. *Science* **1995**, *270*, 1847–1848.
- [11] Fischer, S.; Verma, C. S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9613–9615.
- [12] Mao, Y.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 2950–2951.

- [13] Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneidern, J.; Kent, S. B. H. *Science* **1989**, *245*, 616–621.
- [14] Royer, Jr., W. E.; Pardanani, A.; Gibson, Q. H.; Peterson, E. S.; Friedman, J. M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14526–14531.
- [15] Abashkin, Y. G.; Burt, S. K.; Collins, J. R.; Cachau, R. E.; Russo, N.; Erickson, J. W. Density functional modelling of ligand-metal interactions and enzymatic activities in metalloenzymes. In *Metal-ligand interactions - structure and reactivity*; Russo, N.; Salahub, D., Ed.; Kluwer Academic 1995; pp 1–22.
- [16] Smith, D. E.; Haymet, A. D. J. *J. Chem. Phys.* **1993**, *98*, 6445–6454.
- [17] Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016–2027.
- [18] Laskowski, R. A.; Chistyakov, V. V.; Thornton, J. M. *Nucleic Acids Res.* **2005**, *33*, D266–D268.
- [19] Connolly, M. L. *Science* **1983**, *221*, 709–713.
- [20] Kleywegt, G. J.; Jones, T. A. *Acta Cryst. D* **1994**, *50*, 178–185.
- [21] Kopcsak, P. J. *Matrix Analysis and Analytical Geometry: a new approach*. Chemistry Publishing Company, Inc, New York 1968.
- [22] Woenckhaus, J.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 9586–9587.
- [23] Entropy of vaporization calculated from the method of Kuntz and Kauzmann is -29.4 cal/mol/K at 0° C.
- [24] Kuntz, Jr., I. D.; Kauzmann, W. *Adv. Protein Chem.* **1974**, *28*, 239–345.
- [25] Smith, S.; Smith, J. C.; Verma, C. S. *J. Phys. Chem. B* **2001**, *105*, 8050–8055.



- [26] Fischer, S.; Verma, C. S.; Hubbard, R. E. *J. Phys. Chem. B* **1998**, *102*, 1797–1805.
- [27] Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- [28] Eisenberg, D.; Kauzmann, W. *The Structure and Properties of Water*. Oxford University Press, New York and Oxford 1969.
- [29] Szczęśniak, M. M.; Chałasiński, G.; Cybulski, S. M.; Cieplak, P. *J. Chem. Phys.* **1993**, *98*, 3078.
- [30] Tashiro, M.; Stuchebrukhov, A. A. *J. Phys. Chem. B* **2005**, *109*, 1015–1022.

# Chapter 3

## Protein Simulations at 3 kbar

### 3.1 Introduction

The effect of increased pressure on globular protein structure and function has been studied to try to understand conformational changes proteins undergo prior to denaturation.<sup>5–8</sup> Studies of the denaturation process at increased pressure allow the role of decreased total volume to be separated from other variables that contribute in thermal or chemical unfolding.<sup>6,9–12</sup> For instance while the change in volume upon denaturation is small at atmospheric pressure, with the unfolded protein occupying slightly more volume than its folded counterpart, the volume of the denatured protein is more compact at elevated pressure.<sup>8,13–17</sup> A thorough understanding of the forces involved in protein unfolding can be used to unravel the forces at work in the complicated process governing the transition from the molten globule stage to the final tertiary structure.<sup>12,18–21</sup>

Simulations were performed at 3 kbar for 500–1000 ps using free energy perturbation. The three proteins studied were the Bovine Pancreatic Trypsin Inhibitor (BPTI), the I106A mutant of Human Lysozyme, and the I76A mutant of Barnase. Each protein has a cavity with an isolated water, however the environments vary in the number of available hydrogen bonds from 1 to 4. The number of bonds is only one of the factors

along with volume, mobility of neighboring side chains, and others which determine occupancy in the cavity. Protein calculations were performed with TIP3P water using the Gibbs program from the Amber 6 molecular dynamics package with the non-polarizable Cornell 94 force fields.<sup>12</sup> Analysis of the trajectories was performed using the Carnal program from the Amber package as well as the groups own programs written in Perl and Fortran.

## 3.2 Barnase I76A Mutant: 1BRI

When the free energy of hydration for barnase at 3 kbar is compared to the value at 1 bar a shift towards a more favorable process has occurred (see Table 3.1), but this change is

Table 3.1: Barnase  $\Delta G$  values at 1 bar and 3 kbar. Numbers in parenthesis represent 95% confidence limits.

Pressure	$\Delta G_{protein}$ (kcal/mol)	$\Delta G_{wat}$ (kcal/mol)	$\Delta G_{hyd}$ (kcal/mol)
1bar	-1.5(1)	6.18(4)	4.7(1)
3kbar	-0.70(8)	5.08(5)	4.38(9)

only about 0.4 kcal/mol. The change in  $\Delta G_{hyd}$  is caused by the decreased cost of creating a non-interacting water in the bulk, as the protein portion of the  $\Delta G$  calculation has shifted in the opposite direction by 0.8 kcal/mol. The overall  $\Delta G_{hyd}$  is still greater than 4 kcal/mol indicating the cavity should not be hydrated. Our previous studies involving water in the barnase cavity at 1 bar indicated it flipped roughly across the  $C_2$  axis approximately once every 4.6 ps during the simulation to allow the protons to share the single hydrogen bond donor in the cavity. The protons in the 3 kbar simulations have a longer residency time, with the flip occurring approximately every 5.7 ps. The overall disorder of the water is much smaller at 3 kbar as judged by the root-mean-square

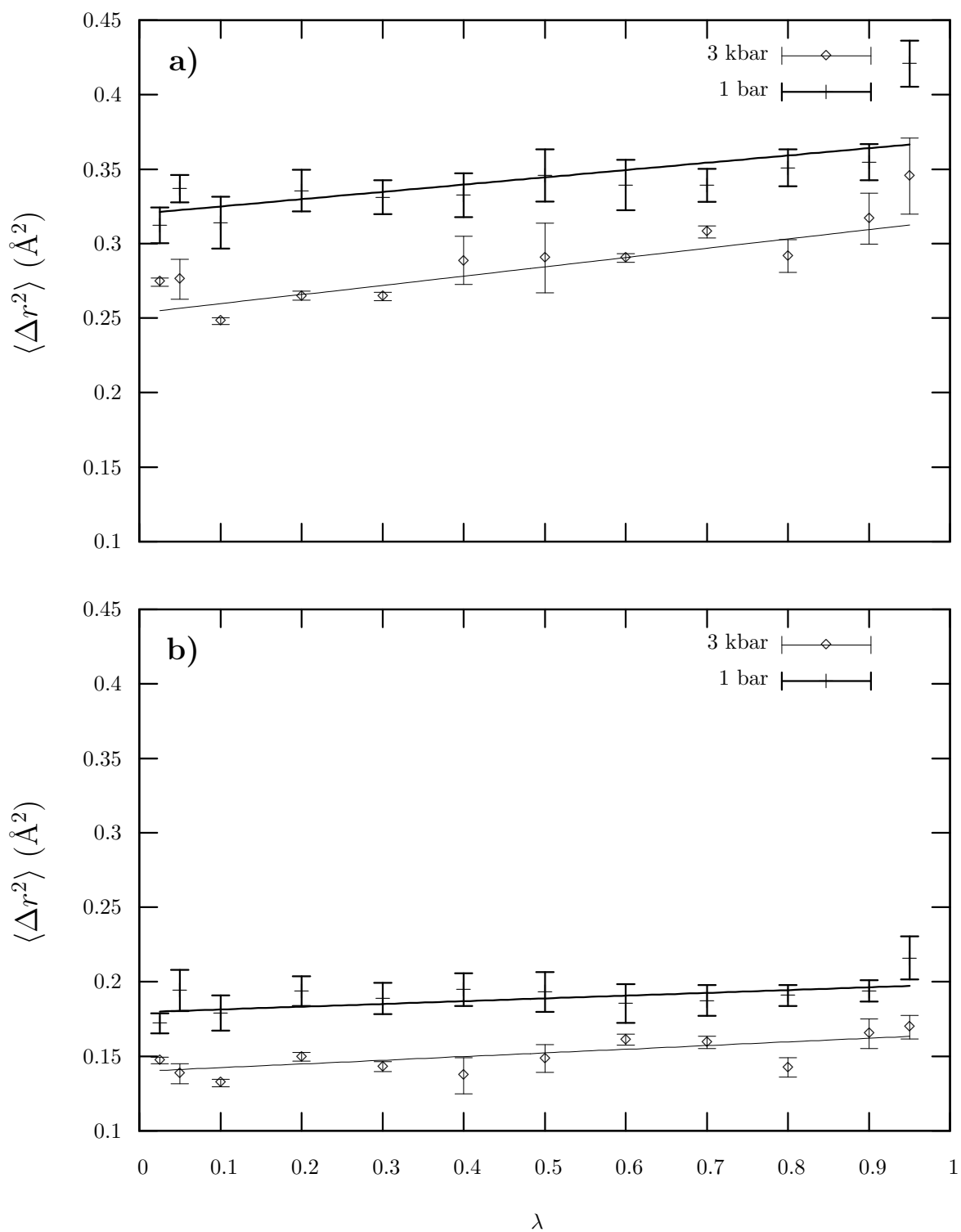


Figure 3.1: Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6 Å of the cavity water (+) from the average structure of Barnase. Errors reported are 95% confidence limits.

Table 3.2: Barnase: Volume of Cavity, Disorder of Water, and Protein Radius of Gyration. Numbers in parenthesis represent 95% confidence limits.

Pressure	Volume Å <sup>3</sup>		$\langle \delta r^2 \rangle^{0.5}$	Radius of Gyration Å	
	Empty	Hydrated		Empty	Hydrated
1 bar	26(3)	29(4)	0.8(6)	13.66(4)	13.63(6)
3 kbar	27(3)	28(3)	0.4(1)	13.75(4)	13.73(4)

deviation of the water oxygen position, which is smaller because the water does not shift from binding to the Phe 7 to Tyr 97 at 3 kbar. A rough volume of the cavity is calculated for barnase using two tetrahedron which share the same base. For the volume calculation the base is composed of C<sub>γ1</sub> Ile 88, C<sub>γ2</sub> Ile 96, and C<sub>δ1</sub> Ile 109, and the vertices are O Phe 7 and O Tyr 97 (Table 3.2). The increase binding times for the water may be explained by a slightly smaller cavity volume which restricts the freedom of movement. This would in turn

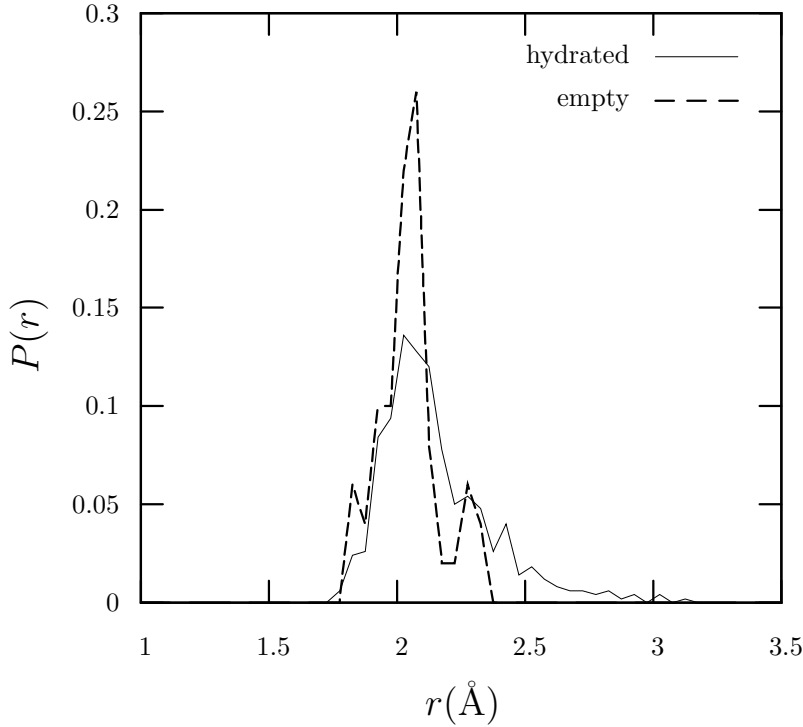


Figure 3.2: Distribution of hydrogen bond lengths between the H on Ala 11 and the O on Phe 7 in both the empty and hydrated simulations of 1BRI.

effect the entropic contribution to free energy, perhaps explaining the increase in  $\Delta G_{protein}$ .

Analysis of the  $\langle \Delta r_i^2 \rangle$  data indicate that the barnase mutant is slightly more flexible when hydrated (Figure 3.1). The dependence on extent of hydration is virtually identical to that observed at 1 bar. However, the overall flexibility of the system is lower than that observed at lower pressure. This can also be seen when hydrogen bonds throughout the

Table 3.3: Average hydrogen bond lengths for bonds 1–21 in the protein barnase. Numbers in parenthesis represent 95% confidence limits.

Bond	Hydrated		Empty	
	3 kbar	1 bar	3 kbar	1 bar
HVal10...OThr6	2.04(0.17)	2.09(0.23)	2.11(0.27)	2.09(0.18)
HAsp12...OAsp8	2.16(0.23)	2.18(0.27)	2.14(0.21)	2.21(0.27)
HLeu14...OVal10	1.94(0.12)	1.98(0.15)	1.96(0.12)	1.96(0.12)
HThr16...OAsp12	2.13(0.32)	2.39(0.42)	2.08(0.32)	2.56(0.52)
HTyr17...OTyr13	2.05(0.20)	1.98(0.18)	2.05(0.17)	1.97(0.15)
HD24Asn23...OIlle1	2.05(0.20)	2.83(1.04)	2.03(0.17)	3.73(0.92)
HIle25...OSer50	2.10(0.17)	2.01(0.15)	2.30(0.18)	1.97(0.13)
HAla30...OThr26	1.98(0.14)	2.05(0.17)	1.96(0.12)	2.02(0.15)
HGln31...OLys27	2.07(0.21)	2.09(0.26)	2.01(0.12)	2.04(0.20)
HGly34...OGln31	2.45(0.33)	2.52(0.41)	2.42(0.33)	2.37(0.33)
HTrp35...OAla30	2.04(0.16)	2.05(0.18)	2.05(0.16)	2.04(0.17)
HVal36...OD4Asn38	1.89(0.12)	1.94(0.15)	1.91(0.12)	1.92(0.13)
HLys39...OVal36	2.28(0.35)	2.18(0.28)	2.12(0.18)	2.11(0.21)
HVal45...OAsn41	2.78(0.36)	2.57(0.38)	2.86(0.39)	2.55(0.34)
HAla46...OLeu42	1.91(0.12)	1.97(0.16)	1.88(0.10)	1.95(0.14)
HLys49...OAla46	2.34(0.29)	2.51(0.33)	2.20(0.22)	2.47(0.27)
HSer50...OAsn23	2.08(0.18)	2.11(0.22)	2.08(0.17)	2.10(0.21)
HGly52...OIle25	1.94(0.14)	1.97(0.16)	1.90(0.11)	1.93(0.13)
HGly53...OAla73	1.90(0.13)	2.01(0.21)	1.92(0.12)	1.93(0.14)
HPhe56...OTrp71	1.97(0.16)	2.00(0.17)	1.90(0.11)	1.98(0.15)
HLeu63...OD4Asn55	2.52(0.30)	2.70(0.48)	2.60(0.26)	2.93(0.59)

protein are examined. Histograms showing the average hydrogen bond lengths and the distribution of these lengths were constructed for the 44 hydrogen bonds found in barnase with an average proton–donor distance of less than 3.0 Å. The average standard deviation of the bond lengths was less broad for the simulation at 3 kbar, 0.20 hydrated and 0.28 empty, versus the 0.32 hydrated and 0.39 empty observed at 1 bar. The overall length of

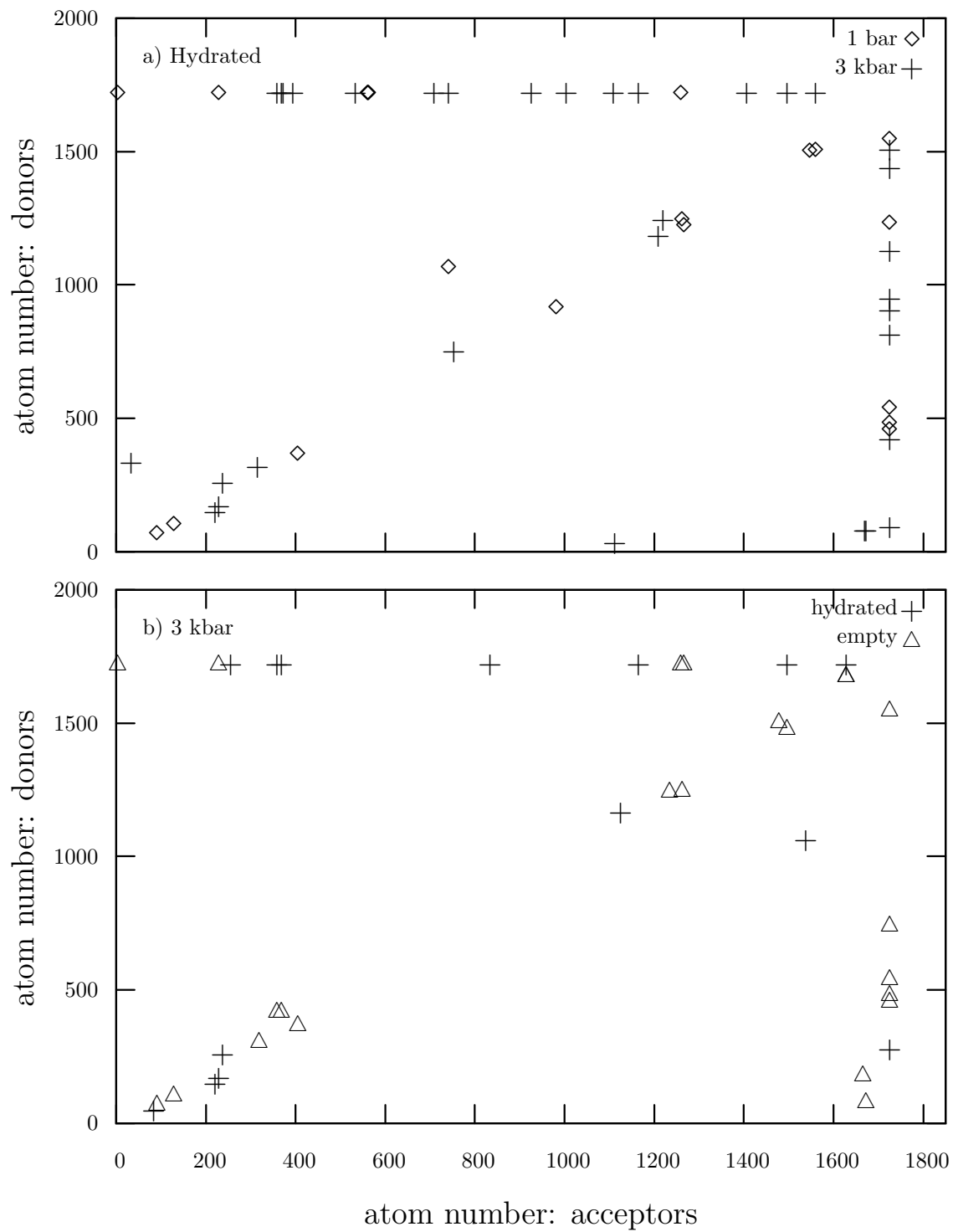


Figure 3.3: Hydrogen bonds for Barnase with at least 70% residency time at a) 1 bar( $\diamond$ ) and 3 kbar( $+$ ) and b) 3 kbar hydrated( $+$ ) and empty( $\triangle$ ). Bonds that exist at both pressures are removed for clarity. All solvent water is given an atom number of 1720.

Table 3.4: Average hydrogen bond lengths for bonds 22–44 in the barnase protein. Numbers in parenthesis represent 95% confidence limits.

Bond	Hydrated		Empty	
	3 kbar	1 bar	3 kbar	1 bar
HH24Thr69...OD1Asp90	2.75(0.31)	3.01(0.34)	2.83(0.28)	2.86(0.31)
HArg72...OTyr90	1.98(0.13)	1.94(0.14)	1.96(0.09)	1.96(0.13)
HGlu73...OAsp54	2.03(0.18)	2.06(0.18)	1.97(0.17)	2.14(0.19)
HAsp75...OIle51	2.10(0.20)	2.07(0.17)	2.03(0.14)	2.05(0.16)
HAsn77...OAsn5	2.51(0.53)	2.12(0.29)	2.78(0.57)	1.91(0.13)
HEArg83...OD5Asp72	1.84(0.10)	1.87(0.14)	1.83(0.09)	1.92(0.14)
HH24Arg83...OD1Ala72	1.80(0.10)	1.78(0.10)	1.80(0.10)	1.76(0.09)
HArg87...OThr99	2.09(0.21)	2.03(0.18)	2.04(0.15)	1.99(0.14)
HEArg87...OHTyr103	3.37(0.85)	3.11(0.68)	2.99(0.35)	2.90(0.54)
HH14Arg87...OAsp83	2.86(1.06)	1.94(0.15)	1.97(0.19)	1.91(0.14)
HIle88...OAla74	1.85(0.11)	1.89(0.12)	1.86(0.10)	1.87(0.10)
HLeu89...OTyr97	1.95(0.14)	1.98(0.15)	1.92(0.12)	1.94(0.13)
HTyr90...OArg72	2.09(0.16)	2.08(0.16)	2.05(0.14)	2.09(0.15)
HLeu95...OGSer91	2.01(0.19)	2.08(0.18)	2.01(0.15)	2.06(0.16)
HTyr97...OLEu89	1.91(0.11)	1.91(0.12)	1.90(0.10)	1.89(0.10)
HLys98...OThr107	1.98(0.14)	1.96(0.15)	1.99(0.11)	1.94(0.13)
HThr99...OArg87	2.03(0.15)	2.14(0.18)	2.14(0.16)	2.15(0.18)
HThr105...OD4Asp98	3.15(1.05)	2.08(0.33)	2.01(0.15)	2.05(0.28)
HZ4Lys108...OArg107	2.91(0.84)	3.31(0.94)	3.15(0.18)	3.91(0.90)
H109...OIle96	2.18(0.27)	2.20(0.35)	2.24(0.27)	2.09(0.21)
HH14Arg110...OIle106	1.96(0.17)	2.78(1.62)	1.88(0.15)	1.95(0.17)
HH25Arg110...OD1Asp5	2.12(0.41)	2.66(0.87)	2.46(0.52)	2.99(0.62)
HAla11...OPhe7	2.10(0.22)	2.19(0.37)	2.00(0.12)	2.06(0.18)

the hydrogen bonds did not appear pressure dependent when the protein cavity was empty with an average length of 2.33 Å in both cases. However, there was a shift observed for the hydrated cavity with 2.41 Å and 2.27 Å average lengths calculated for 1 bar and 3 kbar, respectively (see Table 3.3 and Table 3.4). Most notable is the shift of the proton on Ala 11 and oxygen of Phe 7 of barnase with the bond length increasing from  $2.1 \pm 2$  Å to  $2.2 \pm 4$  Å upon hydration (Figure 3.2). The hydrogen bond between H Ala 11 and O Phe 11 are longer at 3 kbar then they were during the simulation at 1 bar where they increased from  $2.1 \pm 0.2$  Å to  $2.2 \pm 0.4$  Å upon hydration.

When all of the contributing hydrogen bonds in barnase are analyzed the changes in



the tertiary protein structure become evident. These changes in hydrogen bonds are indicated by large shifts in the percentage of simulation time a particular hydrogen bond exists within the cutoff conditions. When comparing the hydrated simulations at both pressures the hydrogen bonds binding the helices and  $\beta$ -sheets do not change to any large extent. However, the hydrogen bonds comprising the turns and bends linking the secondary structure units do shift (Figure 3.3 a). It should be noted none of the amino acids with atoms located within 6 Å of the buried water gain or lose hydrogen bonds during the simulation. When interactions with the water are scaled out more pervasive structural changes occur with bonds to different  $\beta$ -sheets shifting slightly (Figure 3.3 b).

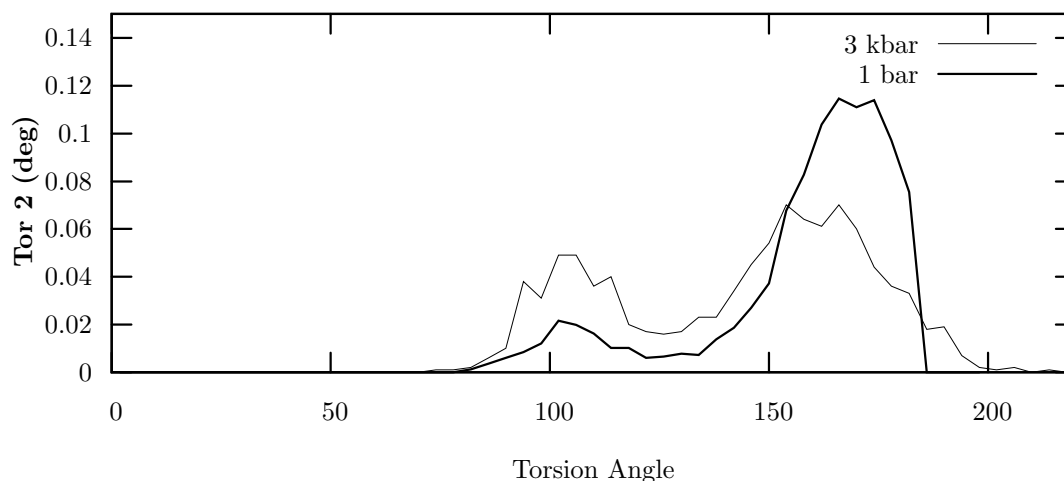


Figure 3.4: Distribution of  $\chi_2$  angles for Ile 88 at  $\lambda = 0.025$ .

For clarity the secondary structure details can be found in Table 3.5. These assignments are according to PDBsum<sup>23</sup> and there is some overlap in transitions. Residues that are not assigned are not part of any specified secondary structure.

The  $C_\alpha - C_\beta - C_{\gamma_1} - C_{\delta_1}$  torsion angle, identified as  $\chi_2$ , does show some pressure dependence. The probability distribution of the torsion angle in the hydrated states at 3 kbar and 1 bar are virtually identical. However, the proteins when filled with even a marginally interacting water,  $\lambda = 0.025$ , show some divergence (Figure 3.4). The Ile 88 at higher pressure has a greater density away from the mutant torsion angle (168°).

Table 3.5: Barnase: Secondary Structure Details

Structure Type	Residue Numbers	Atom Numbers
helix 1 (alpha)	Phe 7–Tyr 17	50–218
beta turn 1	Pro 21–Tyr 24	227–337
beta strand 1	Tyr 24–Ile 25	317–356
helix 2 (alpha)	Lys 27–Ala 32	371–455
helix 3 (alpha)	Ala 37–Lys 39	522–564
helix 4 (alpha)	Leu 42–Val 45	586–642
beta turn 2	Ala 46–Lys 49	652–695
beta strand 2	Ser 50–Ile 55	696–770
beta turn 3	Asn 58–Gly 61	802–861
beta turn 4	Gly 61–Pro 64	855–916
beta turn 5	Lys 66–Arg 69	924–987
beta strand 3	Trp 71–Asp 75	1002–1086
beta turn 6	Asp 75–Tyr 78	1075–1131
beta strand 4	Arg 87–Ser 91	1307–1338
beta turn 7	Ser 91–Trp 94	1328–1385
beta strand 5	Ile 96–Thr 99	1405–1480
beta turn 8	Thr 99–His 102	1467–1523
beta turn 9	Asp 101–Gln 104	1495–1561
beta turn 10	Tyr 103–Phe 106	1524–1595
beta strand 6	Thr 107–Arg 110	1596–1675

### 3.3 Bovine Pancreatic Trypsin Inhibitor (BPTI)

The  $\Delta G_{hyd}$  of BPTI as the pressure is increased to 3 kbar from 1 bar increases by 0.3 kcal/mol (Table 3.6). This shift does not effect the overall energetics of the process as the hydration of the protein should be spontaneous. While the  $\Delta G_{water}$  is less positive at higher pressure, the  $\Delta G_{protein}$  increases by more than 1 kcal/mol. A determination on the

Table 3.6: BPTI  $\Delta G$  values at 1 bar and 3 kbar. Numbers in parenthesis represent 95% confidence limits.

Pressure	$\Delta G_{protein}$ (kcal/mol)	$\Delta G_{wat}$ (kcal/mol)	$\Delta G_{hyd}$ (kcal/mol)
1bar	-10.9(1)	6.18(4)	-4.7(1)
3kbar	-9.5(1)	5.08(5)	-4.4(1)

change in the cavity can be discerned by calculation the volume of the tetrahedron defined by O Thr 11, O Cys 38, N Cys 14, and N Cys 38 (Table 3.7). The cavity volumes do not appear to shift between the simulations at 3 kbar and those at 1 bar. The root-mean-square deviation of the water oxygen from the average position also changes very little between the two simulations, there may be a slight decrease at higher pressure but this change alone should not be enough to lead to a 1.4 kcal/mol increase in  $\Delta G_{protein}$ . There is a shift in the trend for the radius of gyration. At 1 bar the protein with an empty cavity had a larger radius of gyration than the hydrated protein, the reverse is true at 3 kbar. The mean squared fluctuations of BPTI at 3 kbar indicate a flexibility that is independent of extent of hydration. The protein with the empty cavity is more flexible at high pressure with a total mean square deviation of 0.34(1) Å<sup>2</sup> at 3 kbar versus 0.27(1) Å<sup>2</sup> at 1 bar. The hydrated simulations of BPTI display the opposite trend with  $\langle \delta r^2 \rangle$  higher at 1 bar 0.38(2) Å<sup>2</sup> versus 0.34 Å<sup>2</sup> at 3 kbar (Figure 3.5). The changes effect of hydrogen bonds on flexibility changes can be seen in the loss and gain of hydrogen bonds both between the two pressures and the extent of hydration While there is a slight shift in hydrogen bonding of the hydrated protein at both pressures (Figure 3.6 a), the protein shows a significant increase in binding to solvent water when the interior cavity is empty (Figure 3.6 b).

Table 3.7: BPTI: Volume of Cavity, Disorder of Water, and Protein Radius of Gyration. Numbers in parenthesis represent 95% confidence limits.

Pressure	Volume Å <sup>3</sup>		$\langle \delta r^2 \rangle^{0.5}$	Radius of Gyration Å	
	Empty	Hydrated		Empty	Hydrated
1 bar	9.3(7)	8.8(8)	0.18(3)	11.44(6)	11.40(6)
3 kbar	9.3(7)	8.8(6)	0.16(2)	11.37(4)	11.40(4)

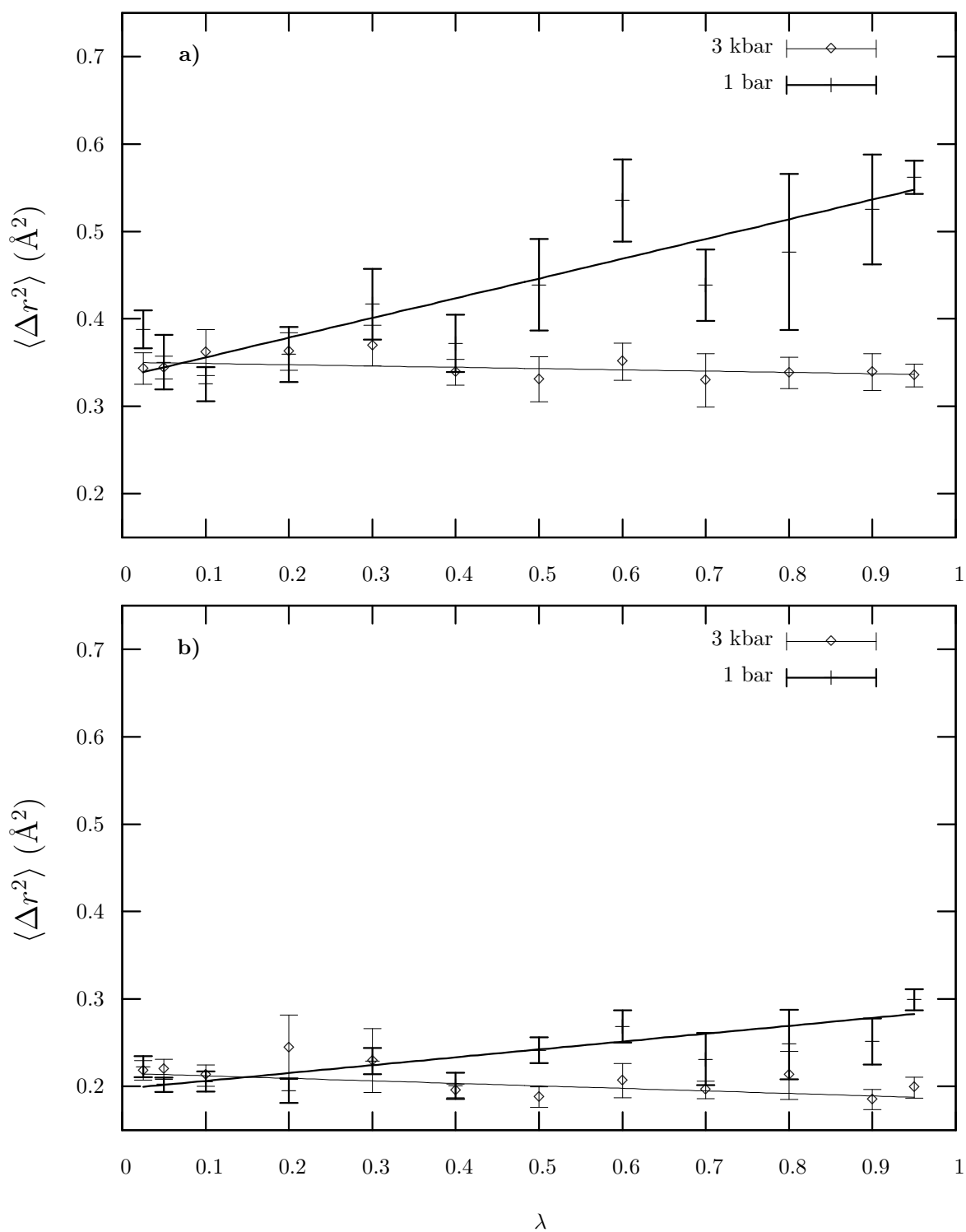


Figure 3.5: Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6  $\text{\AA}$  of the cavity water (+) from the average structure of BPTI. Errors reported are 95% confidence limits.

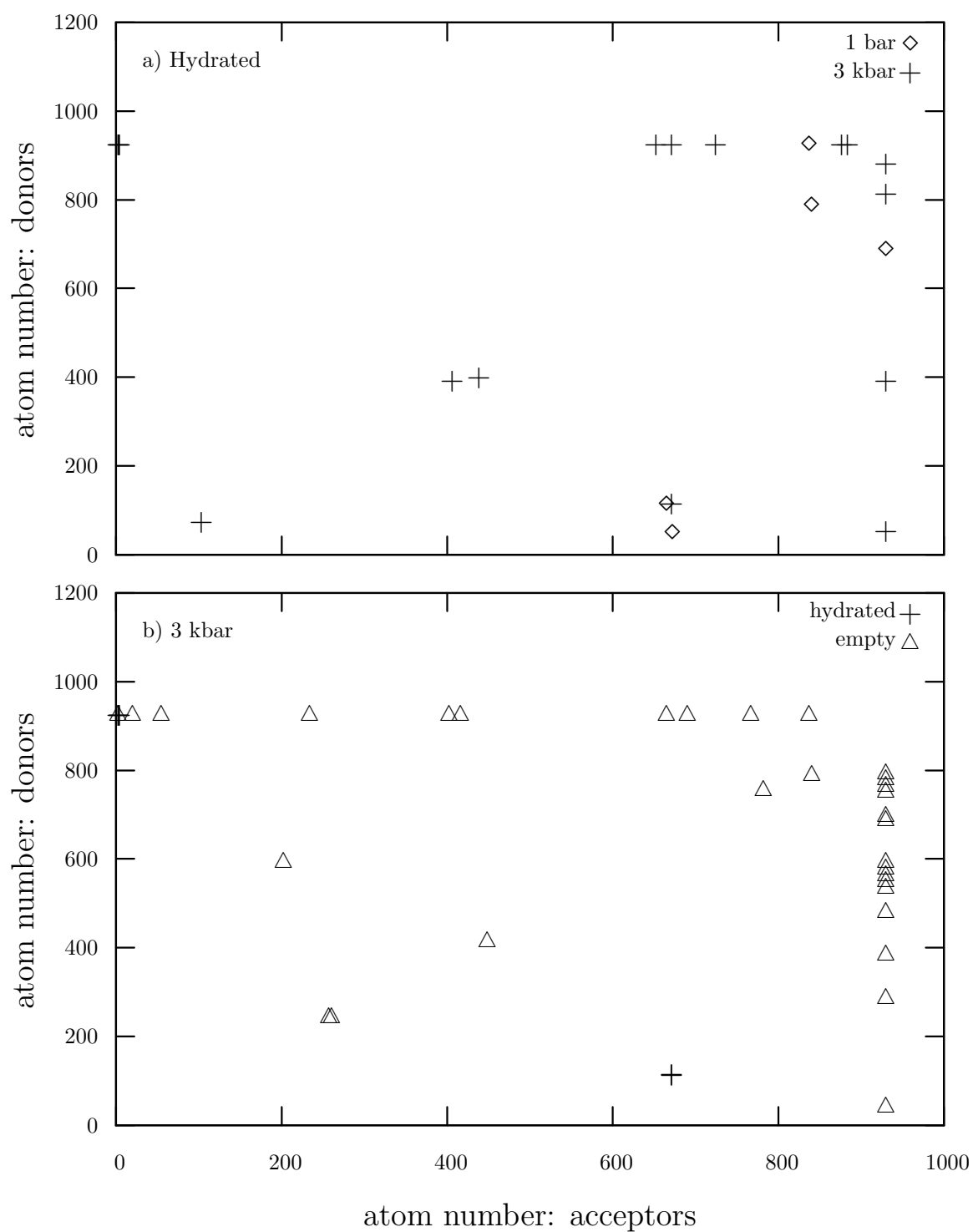


Figure 3.6: Hydrogen bonds for BPTI with at least 70% residency time at a) 1 bar( $\diamond$ ) and 3 kbar( $+$ ) and b) 3 kbar hydrated( $+$ ) and empty( $\triangle$ ). Bonds that exist at both pressures are removed for clarity. All solvent water is given an atom number of 925.

Analysis of this trend by secondary structure type indicated that this shift was not isolated to one structure motif but was instead almost global with helices, turns and sheets all effected. When the hydration dependence of flexibility is further broken down to the individual atom numbers it becomes clear that, with the exception of Arg 17, especially atom numbers 258, 259 and 262, and to a lesser extent Leu 29 and Met 52, atom numbers 463 and 814, there is either no dependence of flexibility on extent of hydration or the protein becomes stiffens with the addition of a water residue (Figure 3.7). This is a marked departure from the data at 1 bar which indicated the protein was more flexible when hydrated. The end regions of the hydrated protein are more flexible in the 1 bar studies then they were at 3 kbar, with the opposite trend for the empty protein. These changes may have a large influence on protein flexibility, but are not necessarily correlated to the extent of hydration. The secondary structure type by atom number and residue for BPTI

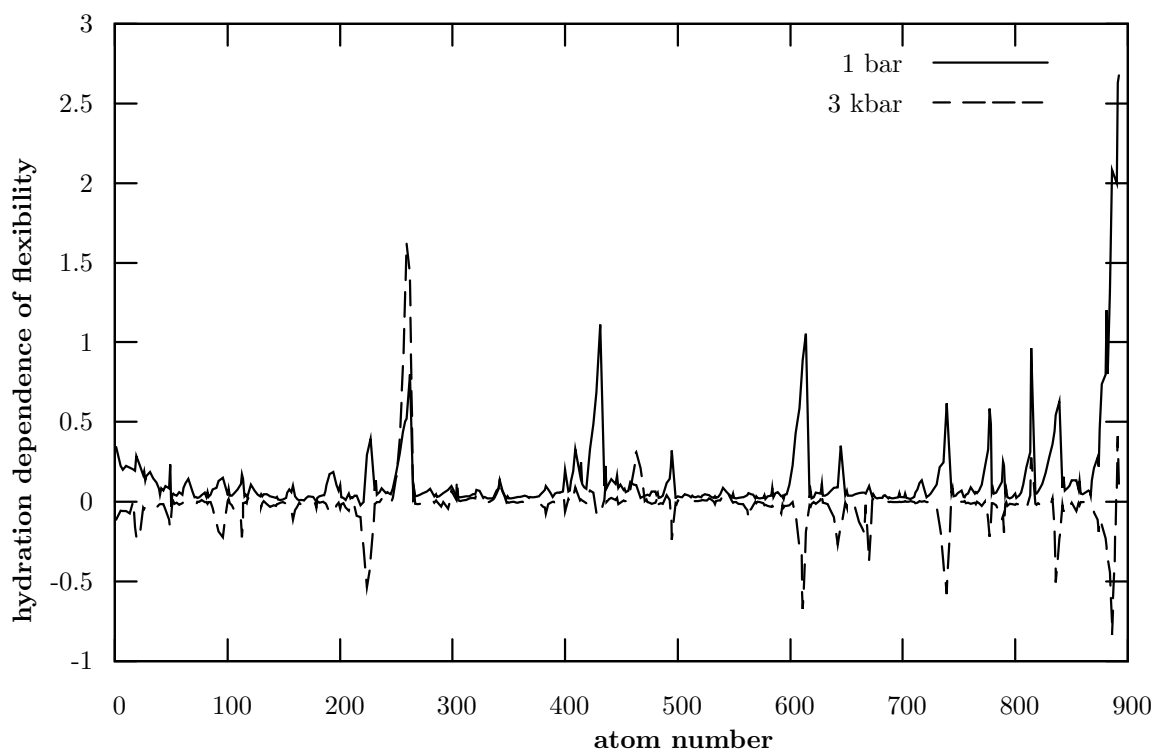


Figure 3.7: Slope of Mean square deviation for heavy atoms in BPTI 3 kbar versus 1 bar

Table 3.8: BPTI: Secondary Structure Details

Structure Type	Residue Numbers	Atom Numbers
helix 1 (G)	Asp 3–Leu 6	41–101
beta strand 1	Ile 18–Asn 24	267–404
beta turn 1	Asn 24–Ala 27	391–446
beta turn 2	Ala 25–Gly 28	405–452
beta strand 2	Leu 29–Tyr 35	454–570
beta turn 3	Lys 41–Asn 44	629–702
helix 2 (alpha)	Ala 48–Cys 55	765–867

can be found in Table ??.<sup>23</sup>

The percentage of the time a hydrogen bond falls within the cutoff limits of 4 Å and 50° allows for easy visualization of bonds that are gained or lost, however the average bond lengths and probability distributions allow the behavior of individual hydrogen bonds to be probed (Tables 3.9, 3.10 and 3.11)

While there is a lack of new interactions within 6 Å of the cavity, the amino acids which bind to those that make up the lining of the cavity do experience shifts. As seen in Table ??) the proton on Gly 36 is hydrogen bound to the oxygen of Thr 11. The bond is stretched with hydration from 1.9(1) Å when empty to 2.0(2) Å for the hydrated state (Figure 3.8). However the average hydrogen bond length for the hydrated state at 3 kbar is slightly shorter than the corresponding length at 1 bar limiting the change even this bond near the cavity experiences upon hydration.

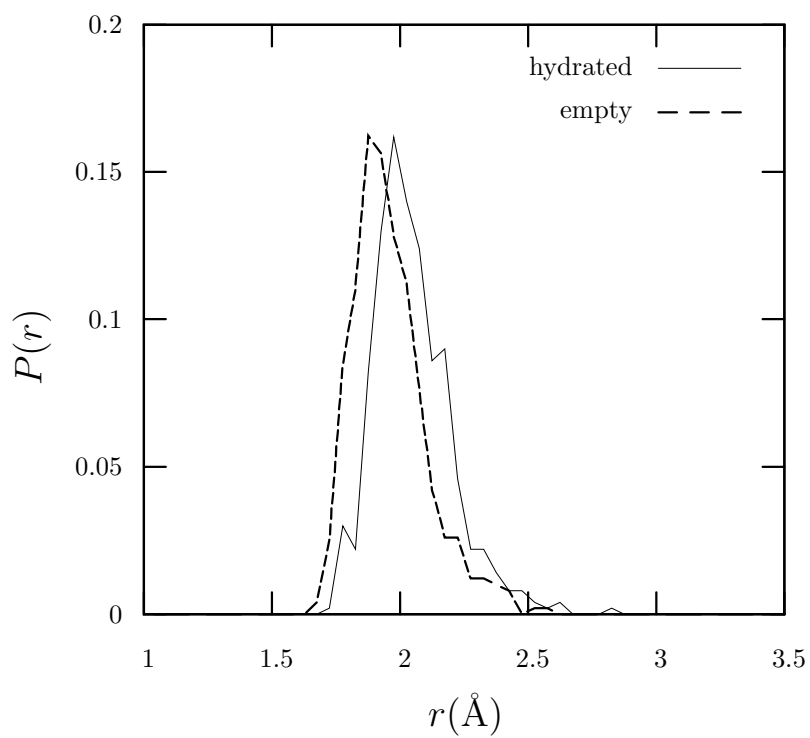


Figure 3.8: Distribution of hydrogen bond lengths between the H on Gly 36 and the O on Thr 11 in both the empty and hydrated simulations of BPTI.



Table 3.9: BPTI: Average hydrogen bond lengths. Bonds 1–25. Numbers in parenthesis represent 95% confidence limits.

Bond	Hydrated		Empty	
	3 kbar	1 bar	3 kbar	1 bar
HGlu7···OPhe4	2.25(0.24)	3.47(0.75)	2.24(0.26)	2.68(0.61)
HGlu7···NLeu6	2.35(0.13)	2.36(0.12)	2.36(0.12)	2.38(0.13)
HTyr10···OTyr10	2.87(0.18)	2.86(0.17)	2.84(0.20)	2.84(0.16)
HThr11···OG1Thr11	2.68(0.20)	2.72(0.18)	2.75(0.20)	2.74(0.17)
HCys14···NPro13	2.47(0.13)	2.47(0.12)	2.47(0.15)	2.44(0.12)
HAla16···NLys15	2.41(0.13)	2.41(0.16)	2.48(0.14)	2.51(0.21)
HAla16···OGly36	2.24(0.28)	2.28(0.35)	2.79(0.55)	2.65(0.51)
Hlle18···NArg17	2.67(0.30)	2.84(0.34)	2.75(0.37)	2.71(0.31)
HArg20···OArg20	2.58(0.20)	2.55(0.19)	2.47(0.15)	2.56(0.17)
HArg20···OPhe33	2.05(0.17)	2.07(0.17)	2.13(0.20)	2.05(0.17)
HTyr21···OTyr21	2.69(0.17)	2.74(0.17)	2.61(0.14)	2.72(0.15)
HTyr21···OPhe45	1.94(0.14)	1.95(0.14)	1.98(0.15)	1.93(0.13)
HPhe22···OPhe22	2.35(0.14)	2.38(0.13)	2.39(0.14)	2.40(0.13)
HPhe22···OGln31	1.95(0.11)	1.95(0.12)	1.96(0.12)	1.94(0.11)
HTyr23···OD1Asn43	1.96(0.14)	1.94(0.14)	1.96(0.15)	1.92(0.12)
HAsn24···OLEu29	1.99(0.18)	1.97(0.16)	1.94(0.13)	1.95(0.14)
HAla25···OTyr23	2.76(0.34)	2.83(0.35)	2.77(0.32)	2.77(0.29)
HLys26···OD1Asn24	2.66(0.40)	2.53(0.41)	2.74(0.40)	2.58(0.40)
HLys26···NAla25	2.50(0.14)	2.48(0.15)	2.50(0.15)	2.48(0.14)
HAla27···OD1Asn24	2.40(0.50)	2.84(0.64)	2.56(0.61)	2.86(0.63)
HAla27···OAsn24	2.75(0.39)	2.55(0.46)	2.72(0.45)	2.47(0.42)
HAla27···NLys26	2.58(0.16)	2.50(0.19)	2.57(0.18)	2.46(0.18)
HGly28···OAsn24	2.01(0.17)	2.03(0.20)	2.15(0.26)	2.06(0.21)
HGly28···NAla27	2.41(0.14)	2.40(0.13)	2.36(0.13)	2.38(0.12)
HLeu29···OAsn24	2.47(0.29)	2.63(0.34)	2.24(0.24)	2.51(0.30)

Table 3.10: BPTI: Average hydrogen bond lengths. Bonds 25–50. Numbers in parenthesis represent 95% confidence limits.

Bond	Hydrated		Empty	
	3 kbar	1 bar	3 kbar	1 bar
HLeu29···NGly28	2.47(0.15)	2.48(0.16)	2.55(0.17)	2.48(0.15)
HLeu29···OLeu29	2.54(0.18)	2.58(0.19)	2.67(0.16)	2.63(0.18)
HGln31···OPhe22	1.97(0.17)	2.02(0.18)	1.97(0.15)	2.01(0.17)
HGln31···OGln31	2.56(0.22)	2.62(0.22)	2.69(0.24)	2.61(0.21)
HThr32···OG1Thr32	2.51(0.18)	2.51(0.18)	2.63(0.23)	2.53(0.18)
HPhe33···OArg20	1.92(0.13)	1.94(0.12)	1.89(0.12)	1.93(0.11)
HPhe33···OPhe33	2.33(0.13)	2.35(0.13)	2.38(0.14)	2.34(0.12)
HPhe4···NAsp3	2.47(0.14)	2.50(0.14)	2.50(0.15)	2.51(0.14)
HTyr35···OIlle18	1.90(0.13)	1.92(0.13)	1.92(0.13)	1.89(0.11)
HTyr35···OTyr35	2.89(0.20)	2.85(0.19)	2.78(0.17)	2.83(0.17)
HGly36···OThr11	2.00(0.15)	2.04(0.17)	1.91(0.15)	1.92(0.13)
HGly37···NGly36	2.38(0.12)	2.38(0.13)	2.36(0.12)	2.37(0.11)
HCys38···NGly37	2.42(0.11)	2.43(0.12)	2.40(0.11)	2.42(0.11)
HCys38···OCys38	2.47(0.15)	2.48(0.15)	2.40(0.14)	2.39(0.14)
HAla40···NArg39	2.63(0.18)	2.66(0.18)	2.56(0.16)	2.57(0.15)
HLys41···OLys41	2.65(0.23)	2.71(0.25)	2.62(0.22)	2.71(0.22)
HAsn43···NArg42	2.61(0.14)	2.58(0.15)	2.58(0.15)	2.54(0.13)
HAsn44···OArg42	2.19(0.25)	2.24(0.25)	3.88(0.35)	2.22(0.25)
HAsn44···NAsn43	2.83(0.17)	2.86(0.16)	2.26(0.25)	2.84(0.14)
HAsn44···OAsn44	2.67(0.18)	2.73(0.18)	2.86(0.16)	2.72(0.18)
HPhe45···OTyr21	1.92(0.12)	1.93(0.13)	2.69(0.19)	1.92(0.12)
HPhe45···OPhe45	2.65(0.16)	2.67(0.16)	3.42(0.33)	2.69(0.16)
HSer47···NLys46	2.40(0.14)	2.45(0.15)	2.41(0.14)	2.43(0.14)
HSer47···OSer47	2.45(0.18)	2.46(0.17)	2.48(0.18)	2.48(0.16)
HCys5···OPro2	2.47(0.28)	2.19(0.26)	2.45(0.28)	2.40(0.27)

Table 3.11: BPTI: Average hydrogen bond lengths. Bonds 50–73. Numbers in parenthesis represent 95% confidence limits.

Bond	Hydrated		Empty	
	3 kbar	1 bar	3 kbar	1 bar
HCys5···NPhe4	2.41(0.12)	2.39(0.12)	2.41(0.12)	2.41(0.11)
HCys5···SGCys5	2.96(0.19)	3.24(0.31)	2.94(0.19)	3.01(0.21)
HGlu49···NAla48	2.49(0.15)	2.50(0.15)	2.48(0.13)	2.51(0.14)
HAsp50···OGSer47	2.18(0.22)	2.26(0.28)	2.49(0.54)	2.25(0.25)
HAsp50···NGlu49	2.65(0.16)	2.62(0.15)	2.61(0.16)	2.65(0.14)
HCys51···OSer47	2.26(0.23)	2.22(0.23)	2.19(0.22)	2.34(0.24)
HCys51···OAla48	2.98(0.29)	3.02(0.34)	2.98(0.32)	2.84(0.28)
HCys51···NAsp50	2.46(0.14)	2.48(0.14)	2.48(0.14)	2.45(0.13)
HMet52···OAla48	1.97(0.15)	2.02(0.19)	2.00(0.19)	1.95(0.14)
HMet52···NCys51	2.75(0.15)	2.71(0.15)	2.68(0.14)	2.67(0.13)
HArg53···OGlu49	2.05(0.20)	2.02(0.19)	2.00(0.17)	2.03(0.16)
HArg53···NMet52	2.49(0.13)	2.54(0.14)	2.53(0.14)	2.54(0.14)
HThr54···OAsp50	2.39(0.35)	2.26(0.36)	2.33(0.37)	2.09(0.23)
HThr54···OCys51	2.51(0.33)	2.62(0.35)	2.50(0.36)	2.85(0.33)
HThr54···NArg53	2.48(0.16)	2.52(0.17)	2.46(0.16)	2.62(0.19)
HThr54···OG1Thr54	2.74(0.23)	2.69(0.21)	2.77(0.22)	2.59(0.19)
HLeu6···OAsp3	2.21(0.24)	2.26(0.28)	2.40(0.41)	2.22(0.32)
HLeu6···NCys5	2.37(0.12)	2.36(0.12)	2.37(0.12)	2.37(0.11)
HCys55···OCys51	1.88(0.11)	1.93(0.14)	1.89(0.12)	1.94(0.13)
HCys55···NThr54	2.59(0.15)	2.59(0.16)	2.64(0.17)	2.61(0.16)
HGly56···OMet52	2.18(0.26)	2.45(0.42)	2.30(0.34)	2.48(0.44)
HGly56···NCys55	2.41(0.12)	2.39(0.12)	2.40(0.12)	2.41(0.13)
HGly57···NGly56	2.50(0.20)	2.75(0.54)	2.49(0.18)	2.52(0.26)

### 3.4 Lysozyme I106A Mutant

The lysozyme mutant exhibits an overall negative shift for  $\Delta G_{hyd}$  of 1.2 kcal/mol, however, the shift is not sufficient to make the process spontaneous (Table 3.12). The lack of

Table 3.12: Lysozyme  $\Delta G$  values at 1 bar and 3 kbar. Numbers in parenthesis represent 95% confidence limits.

Pressure	$\Delta G_{protein}$ (kcal/mol)	$\Delta G_{wat}$ (kcal/mol)	$\Delta G_{hyd}$ (kcal/mol)
1bar	-3.5(1)	6.18(4)	2.7(1)
3kbar	-3.6(3)	5.08(5)	-1.5(3)

changes in  $\Delta G_{protein}$  is consistent with the hydrogen bonds and flexibility only experience small changes as well.

It is only within the hydrogen bond cutoff for 13.5% of the simulation (Table 3.13). Rough volume calculations were done with vertices defined as  $C_\beta$  Leu 31,  $N_{\epsilon 1}$  Trp 112,  $C_{\gamma 2}$

Table 3.13: Percent of simulation hydrogen bond exists between buried water and surrounding cavity.

Donor	Acceptor	% Simulation Time 1 bar	%Simulation Time 3 kbar
H WAT	O Asn 27	29.8	12.2
O WAT	HE1 Trp 112	9	1.3
H WAT	OD1 Asn 27	12.2	0
H WAT	O Ala 106	3.4	0
H WAT	O Trp 28	0.2	0

Ile 23, O Asn 27, and CB Ala 106. The overall volume of the cavity is slightly larger at 3 kbar than it is at 1 bar when the hydrated states are compared, however the differences are very small and it is difficult to make real comparisons (Table 3.14). The rmsd of the cavity water increases slightly, but this change is within the calculated error. The overall radius of gyration decreases slightly in both the hydrated and empty states relative to 1 atm.

Lysozyme at both 1 bar and 3 kbar has a mean-squared deviation that is largely

Table 3.14: Lysozyme: Volume of Cavity, Disorder of Water, and Protein Radius of Gyration. Numbers in parenthesis represent 95% confidence limits.

Pressure	Volume Å <sup>3</sup>		$\langle \delta r^2 \rangle^{0.5}$	Radius of Gyration Å	
	Empty	Hydrated		Empty	Hydrated
1 bar	36(8)	31(5)	0.9(5)	14.34(2)	14.34(5)
3 kbar	34(3)	31(4)	1.0(4)	14.32(3)	14.27(3)

independent of extent of hydration (Figure 3.9). There is also no change in hydration dependence for the protein within 6 Å of the interior water. The overall flexibility decreases at 3kbar versus 1 bar indicating that the atoms within the protein are less mobile under the applied pressure. There are a variety of changes within the tertiary structure at 3 kbar with changes occurring to coiled regions, and also to the secondary structure of helix 5, 6, and 7. While the water does not penetrate the interior of the protein, the residues are become more solvent exposed (Figure 3.10). The details of the secondary structure of the lysozyme mutant are given in Table 3.15.

## 3.5 Summary

The properties of the proteins studied at 3 kbar do not display a consistent dependence on hydration and the change in free energy of hydration is dependent on cavity type. The barnase and lysozyme mutants with their single buried waters in hydrophobic environments, demonstrate a clear shift towards a lower  $\Delta G_{Hyd}$ . While BPTI, with an isolated water in a hydrophilic environment, has a positive shift. While the change in free energy of hydration is small for both proteins, less than 1.0 kcal/mol, the values do appear statistically significant.

The protein at 1 bar exhibited a positive dependence on hydration, that is the

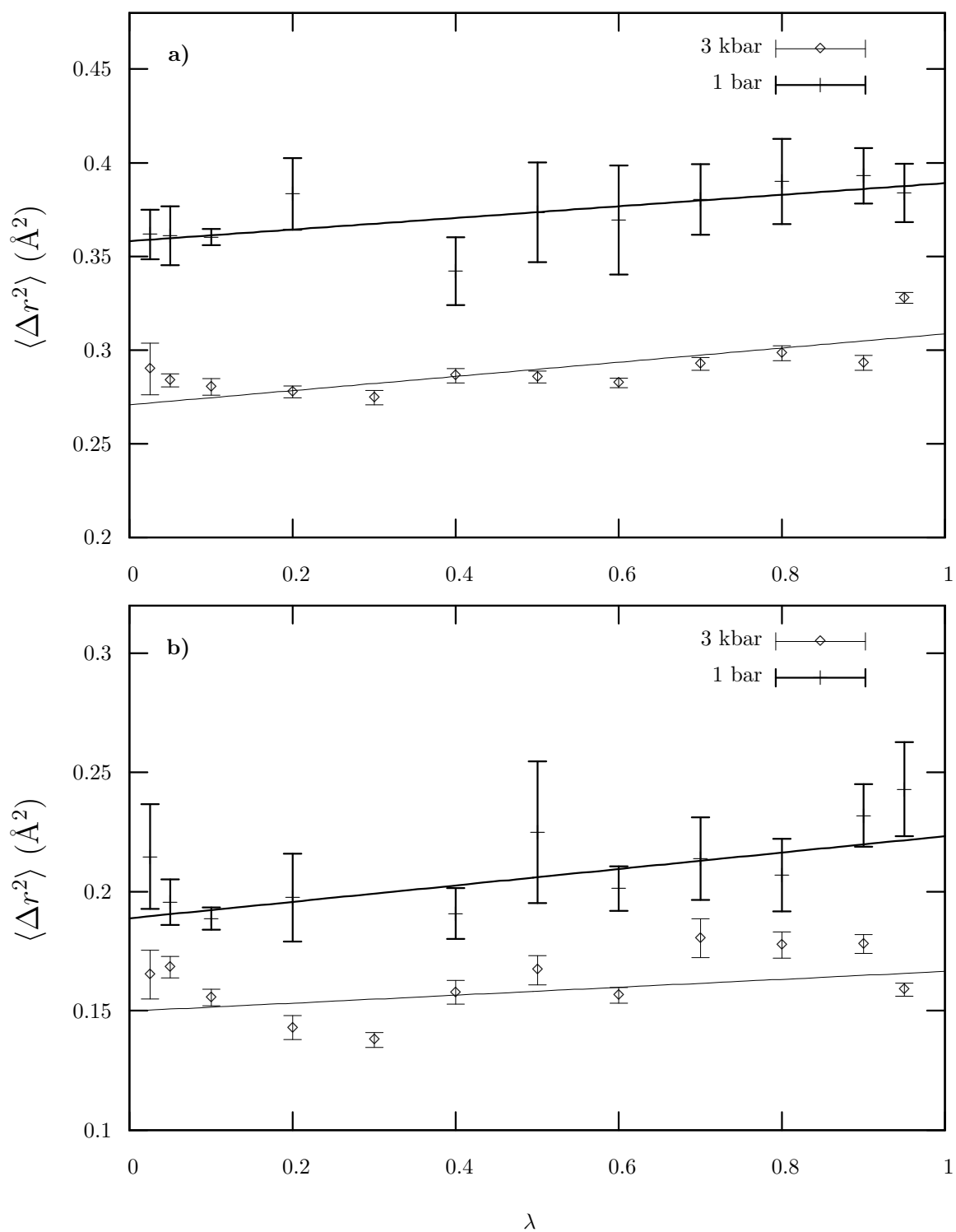


Figure 3.9: Mean square deviation of the full protein with side chains ( $\diamond$ ) and protein backbone atoms within 6  $\text{\AA}$  of the cavity water (+) from the average structure of Lysozyme. Errors reported are 95% confidence limits.

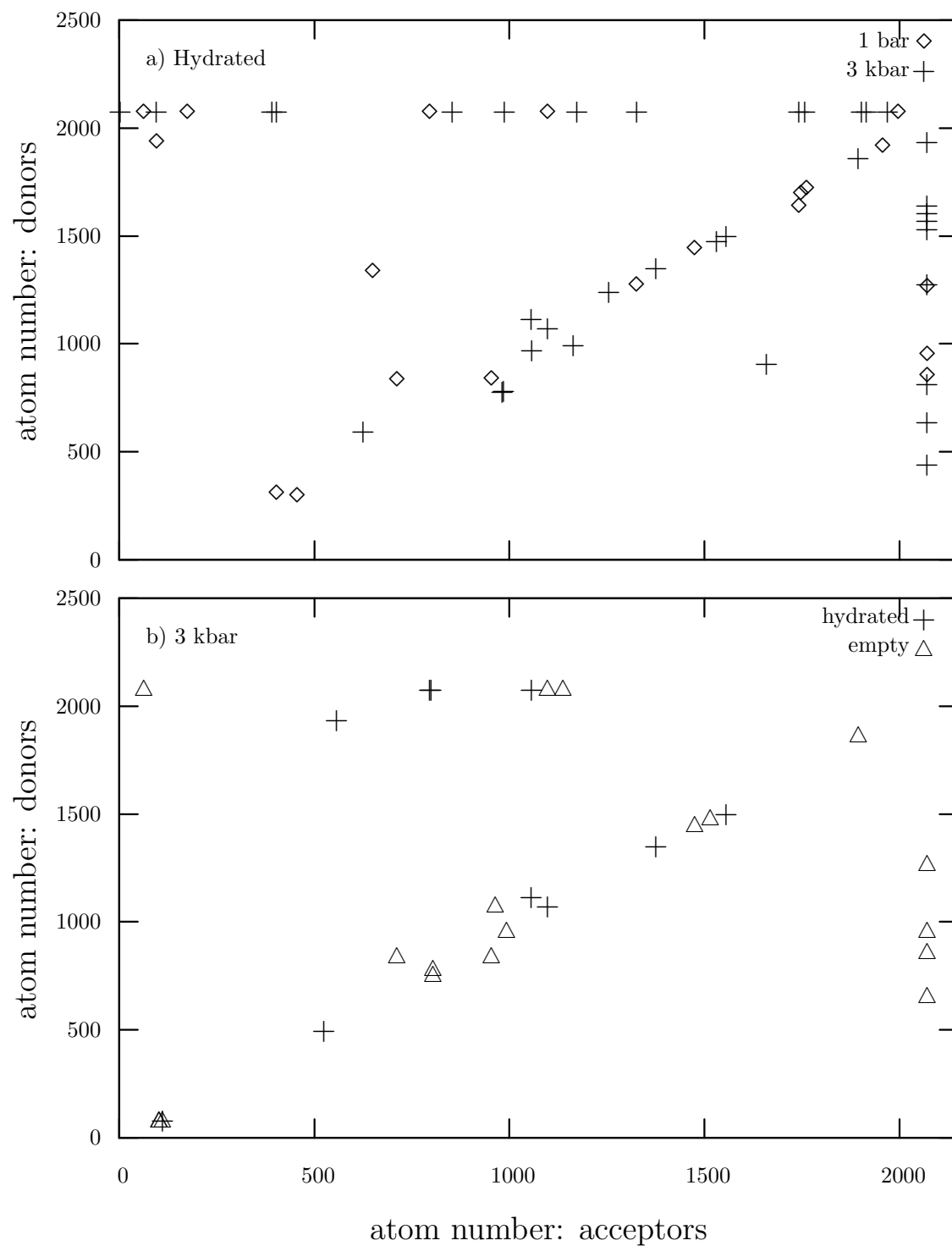


Figure 3.10: Hydrogen bonds for Lysozyme with at least 70% residency time at a) 1 bar( $\diamond$ ) and 3 kbar( $+$ ) and b) 3 kbar hydrated( $+$ ) and empty( $\triangle$ ). Bonds that exist at both pressures are removed for clarity. All solvent water is given an atom number of 2070.

Table 3.15: Lysozyme: Secondary Structure Details

Structure Type	Residue Numbers	Atom Numbers
helix 1 (alpha)	Arg 5–Arg 14	76–256
beta turn 1	Met 17–Tyr 20	283–339
helix 2 (alpha)	Leu 25–Glu 35	401–584
beta turn 2	Ser 36–Asn 39	585–637
beta turn 3	Asn 39–Ala 42	624–685
beta strand 1	Thr 43–Tyr 45	686–734
beta turn 4	Asn 46–Asp 49	735–777
beta turn 5	Ala 47–Arg 50	749–801
beta strand 2	Thr 52–Tyr 54	813–859
beta turn 6	Gly 55–Gln 58	860–922
beta strand 3	Ile 59–Asn 60	923–955
beta turn 7	Asn 60–Tyr 63	942–1011
beta turn 8	Ser 61–Trp 64	956–1035
beta turn 9	Asp 67–Thr 70	1060–1114
beta turn 10	Thr 70–Ala 73	1101–1145
beta turn 11	Asn 75–His 78	1162–1212
helix 3 (G)	Cys 81–Leu 85	1243–1311
beta turn 12	Gln 86–Ile 89	1312–1373
helix 4 (alpha)	Ala 90–Val 100	1374–1529
beta turn 13	Asp 102–Gly 105	1554–1603
helix 5 (G)	Gly 105–Ala 108	1597–1647
helix 6 (alpha)	Val 110–Arg 115	1672–1783
beta turn 14	Cys 116–Arg 119	1673–1848
helix 7 (G)	Arg 122–Tyr 124	1877–1938
beta turn 15	Val 125–Cys 128	1939–1988

hydrated protein was more flexible than the empty protein. However, at 3 kbar the flexibility of BPTI shows little dependence on hydration, and the sections within 6 Å of the cavity water exhibit a slight negative dependence. The slight increase in cavity size upon hydration and decrease in near by hydrogen bond lengths is consistent with results at 1 bar and can not be used to explain this behavior. Most residues in the protein exhibit less motion during the simulation indicating the hydrated protein is stiffens as pressure is increased.

Buried water in the hydrophobic barnase cavity has a same effect on flexibility at both pressures. However, the water at higher pressure exhibits both longer periods between



flips and a smaller  $\langle \delta r^2 \rangle$  of the water oxygen. This apparent lack of mobility should effect the entropic contribution to the  $\Delta G_{hyd}$ .

Perhaps most interesting was the effect of pressure on lysozyme, where despite the lack of any stable hydrogen bonds during the course of the simulation, the  $\Delta G_{hyd}$  was over 1 kcal/mol more favorable. While the resulting free energy still did not indicate a spontaneous process it does give credence to the idea that hydrogen bonding alone can not be used to predict extent of hydration.

Examination of the  $\langle \delta r^2 \rangle$  data for both proteins on a residue-by-residue basis and by structure type failed to generate any patterns to the shift other than the structure changes that occur at turns, bends, and residues next to, but not part of, major structure types. This agrees with the concept that the areas that are altered first under increased pressure are those that do not have a rigid hydrogen bond network. The role of buried water in the denaturation process is still difficult to discern and our flexibility results indicate that it may depend greatly on the environment in which the water is located.

# Bibliography

- [1] Buckle, A. M.; Cramer, P.; Fersht, A. R. *Biochemistry* **1996**, *35*, 4298–4305.
- [2] Kraulis, P. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.
- [3] Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301–329.
- [4] Takano, K.; Funahashi, J.; Yamagata, Y.; Fujii, S.; Yutani, K. *J. Mol. Biol.* **1997**, *274*, 132–142.
- [5] Refaee, M.; Tezuka, T.; Akasaka, K.; Williamson, M. P. *J. Mol. Biol.* **2003**, *327*, 857–865.
- [6] Wroblowski, B.; Diaz, J. F.; Heremans, K.; Engelborghs, Y. *PROTEINS: Struct. Funct. Gen.* **1996**, *25*, 446–455.
- [7] Paci, E.; Marchi, M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11609–11614.
- [8] Kitchen, D. B.; Reed, L. H.; Levy, R. M. *Biochemistry* **1992**, *31*, 10083–10093.
- [9] Peng, X.; Jonas, J.; Silva, J. L. *Biochemistry* **1994**, *33*, 8323–8329.
- [10] Weber, G.; Drickamer, H. G. *Q. Rev. Biophys.* **1983**, *16*, 89–112.
- [11] Samarasinghe, S. D.; Campbell, D. M.; Jonas, A.; Jonas, J. *Biochemistry* **1992**, *31*, 7773–7778.
- [12] Jonas, J.; Ballard, L.; Nash, D. *Biophys. J.* **1998**, *75*, 445–452.

- [13] Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. *Proc. Natl. Acad. Sci.* **1998**, *95*, 1552–1555.
- [14] Harpaz, Y.; Gerstein, M.; Chothia, C. *Structure* **1994**, *2*, 641–649.
- [15] Royer, C. A. *Methods in Enzymology* **1995**, *259*, 357–377.
- [16] Payne, V. A.; Matubayasi, N.; Murphy, L. R.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101*, 2054–2060.
- [17] Hillson, N.; Onuchic, J. N.; Garcia, A. E. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14848–14853.
- [18] Zhang, J.; Peng, X.; Jonas, A.; Jonas, J. *Biochemistry* **1995**, *34*, 8631–8641.
- [19] Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem* **1990**, *59*, 631–660.
- [20] Buck, M.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1994**, *237*, 247–254.
- [21] Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry*. John Wiley and Sons, Inc., New York 2002.
- [22] Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. I.; Vincent, J. J.; Crowlet, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 6*, University of California, San Francisco, 1999.
- [23] Laskowski, R. A.; Chistyakov, V. V.; Thornton, J. M. *Nucleic Acids Res.* **2005**, *33*, D266–D268.

# Chapter 4

## Fluctuating Charge Algorithms

### 4.1 Introduction

Partial charges on atoms in traditional non-polarizable molecular dynamics simulations are constants defined by the force field. This eliminates a charges ability to migrate in response to changes in their environments electric field. In this way interactions between molecules A and B will have no effect on the interactions between molecules A and C, regardless of their polarity. This will cause difficulties in systems where molecular motions changes in the electric field.<sup>1-6,8,15</sup>

The implementation of polarizability in molecular dynamics simulations requires a self-consistent method to find the variables that define the ground-state energy of the system at every time step.<sup>1,9</sup> The two main methods for determining these variables are the iterative and the extended Lagrangian. Each method has its advantages and disadvantages. The iterative method usually allows longer time steps then the extended Lagrangian does, however, it is more computationally expensive as it requires additional energy evaluations at every step.<sup>1,10</sup>

The fluctuating charge method treats point charges as variables that can respond to changes in the environment subject to a conservation of charge constraint.<sup>1,11</sup> The

extended Lagrangian method is implemented by assigning each charge a fictitious mass and propagating it using low temperature Newtonian dynamics. While the original method, developed for water, was capable of 1 fs time steps, the transition to protein systems has resulted in a  $\delta t$  of only 0.5 fs.<sup>1,2</sup> This length of time step seriously impedes the wide spread implementation of fluctuating charge into traditional molecular dynamics simulations.

In an attempt to increase the time steps available using the extended Lagrangian method the following three changes were made to the original fluctuating charge methodology.

Method 1. Assign each charge the same mass and propagate them independently.<sup>11</sup> This is parameterizing the original method for a longer time step.

Method 2. Assign each atom type a different charge mass, and again propagate them independently.<sup>2, 12, 13</sup>

Method 3. Replace the charge degrees of freedom with normal modes, which are assigned masses and propagated. In the case of water there are two normal modes to propagate.

The data presented in the following chapter was previously published.<sup>14</sup>

## 4.2 Methods

The basic potential for the electronegativity equalization method of adding polarizability to molecular dynamics simulations can be written

$$U(q) = \sum_{\alpha} (E_{0,\alpha} + \chi_{\alpha}^0 q_{\alpha} + 1/2 J_{\alpha\alpha} q_{\alpha}^2) + \sum_{\alpha} \sum_{\beta} J_{\alpha\beta}(r_{\alpha\beta}) q_{\alpha} q_{\beta}. \quad (4.1)$$

where  $J_{\alpha\beta}$  is dependent on atomic distances and approaches  $1/r_{\alpha\beta}$  at large distances. For a collection of molecules

$$U(\{q\}, \{\mathbf{r}\}) = \sum_i \left( \sum_{\alpha} \tilde{\chi}_{\alpha}^0 q_{i\alpha} + \frac{1}{2} \sum_{\alpha} \sum_{\beta} q_{i\alpha} q_{i\beta} J_{\alpha\beta}(r_{i\alpha, i\beta}) - E_i^{gp} \right) \\ + \sum_{j < i} \sum_{\alpha} \sum_{\beta} \left( 4\epsilon_{\alpha\beta} \left[ \left( \frac{\sigma_{\alpha\beta}}{r_{i\alpha, j\beta}} \right)^{12} - \left( \frac{\sigma_{\alpha\beta}}{r_{i\alpha, j\beta}} \right)^6 \right] + q_{i\alpha} q_{j\beta} / r_{i\alpha, j\beta} \right) \quad (4.2)$$

where the sums over  $i$  and  $j$  are over molecules,  $\alpha$  and  $\beta$  are over the atoms in the molecule,  $E_i^{gp}$  is the energy of an isolated or gas-phase molecule and  $\tilde{\chi}_{\alpha}^0$  are parameters of the potential that depend on atom types. The charges are determined by minimizing the energy with a conservation of charge constraint for each molecule. For proteins the constraint is usually set to each amino acid.<sup>4,11,12,16,17</sup> The charge constraint can be used to eliminate one of the charges,  $q_{i1} = -\sum_{\alpha \neq 1} q_{i\alpha}$ , The TIP4P water is a 4-site water model where the charge on the oxygen is displaced to a M-site to better represent the charge placement in water, whenever M-site is mentioned it refers to the charge associated with the oxygen. For TIP4P-FQ water  $\alpha=1$  is assigned as the charge on the M-site,  $\alpha=2$  and  $\alpha=3$  are the charges on each of the protons. Thus, the charge on the oxygen can be represented as a linear combination of the two proton charges. The intra-molecular portion of Eq 4.2 can be written

$$U_i = \sum_{\alpha=2}^{N_{atoms}} q_{i\alpha} \Delta \tilde{\chi}_{\alpha} + \frac{1}{2} \sum_{\alpha=2}^{N_{atoms}} \sum_{\beta=2}^{N_{atoms}} q_{i\alpha} q_{i\beta} A_{\alpha\beta} \quad (4.3)$$

where

$$\Delta \tilde{\chi}_{\alpha} = \tilde{\chi}_{\alpha} - \tilde{\chi}_1 \quad (4.4)$$

and

$$A_{\alpha\beta} = J_{\alpha\beta} r_{i\alpha, i\beta} - J_{\alpha 1} r_{i\alpha, i1} - J_{1\beta} r_{i1, i\beta} + J_{11} r_{i1, i1}. \quad (4.5)$$

The equation of motion for each charge  $\alpha$  can be expressed as

$$m_{q\alpha}\ddot{q}_{i\alpha} = -\Delta\tilde{\chi}\alpha - \sum_{\beta=2}^{N_{atoms}} q_{i\alpha}A_{\alpha\beta}. \quad (4.6)$$

The actual equations for each of the methods are derived from the above equations.

For Method 1 enforcing the charge neutrality with Lagrange multiplier<sup>1</sup>

$$\lambda_i = -\frac{1}{N_{atoms}} \sum_{\alpha=1}^{N_{atoms}} \frac{\partial U}{\partial q_{i\alpha}} \quad (4.7)$$

the equation of motion for the charge is

$$m_q\ddot{q}_{i\alpha} = -\frac{\partial U}{\partial q_{i\alpha}} - \lambda_i. \quad (4.8)$$

The Hamiltonian for the system is

$$H = U + \frac{1}{2} \sum_i \sum_{\alpha=1}^{N_{atoms}} m_{\alpha} \dot{r}_{i\alpha}^2 + \frac{1}{2} \sum_i \sum_{\alpha=1}^{N_{atoms}} m_q \dot{q}_{i\alpha}^2. \quad (4.9)$$

In Method 2 charge masses are assigned based on atom type,<sup>2,12,13</sup> with a different value for hydrogen and for the M-site charges. The equations of motion are

$$m_{q\alpha}\ddot{q}_{i\alpha} = -\frac{\partial U}{\partial q_{i\alpha}} - \lambda_i \quad (4.10)$$

with

$$\lambda_i = -\frac{\sum_{\alpha=1}^{N_{atoms}} \frac{\partial U}{\partial q_{i\alpha}} / m_{q\alpha}}{\sum_{\alpha=1}^{N_{atoms}} 1/m_{q\alpha}} \quad (4.11)$$

with the Hamiltonian expressed as

$$H = U + \frac{1}{2} \sum_i \sum_{\alpha=1}^{N_{atoms}} m_{\alpha} \dot{r}_{i\alpha}^2 + \frac{1}{2} \sum_i \sum_{\alpha=1}^{N_{atoms}} m_{q\alpha} \dot{q}_{i\alpha}^2. \quad (4.12)$$

Transforming to normal modes is done to decrease the coupling between the atomic degrees of freedom and the charge degrees of freedom. As seen in Eq 4.6 the charges are linearly coupled, the transform to normal coordinates,  $\mathbf{Q}$ , from the point charge coordinates,  $\mathbf{q}$ , can be performed by  $\mathbf{q} = \mathbf{C}\mathbf{Q}$ , where  $\mathbf{q}$  and  $\mathbf{Q}$  are length vectors, and  $\mathbf{C}$  is the matrix that diagonalizes  $A_{\alpha\beta}$ . This is done such that

$$M_Q \ddot{Q}_{i\alpha} = M_Q \omega_\alpha^2 Q_{i\alpha} \quad (4.13)$$

$M_Q$  is the mass of the normal mode coordinate, and  $\omega_\alpha$  is the frequency of the oscillation.

For the case of TIP4P-FQ water, as applied in Method 3,  $m_{q2} = m_{q3} = m_q$ ,  $A_{22} = A_{33}$ , and  $A_{23} = A_{32}$  because of symmetry. The matrix  $\mathbf{C}$  is

$$\mathbf{C} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \quad (4.14)$$

such that the two normal modes are

$$Q_{i1} = q_{i2} + q_{i3} \quad (4.15)$$

$$Q_{i2} = q_{i2} - q_{i3} \quad (4.16)$$

and the oscillation frequencies are calculated to be

$$\begin{aligned} \omega_1^2 &= \frac{(A_{22} + A_{23})}{m_q} \\ &= \frac{(J_{HH}(0) + J_{HH}(r_{HH}) - 4J_{MH}(r_{MH}) + 2J_{MM}(0))}{m_q} \end{aligned} \quad (4.17)$$

and

$$\omega_2^2 = \frac{(A_{22} - A_{23})}{m_q}$$



$$= \frac{(J_{HH}(0) - J_{HH}r_{HH})}{m_q}. \quad (4.18)$$

The hydrogen–hydrogen distance and hydrogen–M–site distance are represented by  $r_{HH}$  and  $r_{MH}$ . A normal mode charge mass of  $M_q$  and the original charge  $m_q/2$  have equivalent kinetic energies. The motion for the two charge normal modes can be expressed as

$$M_Q \ddot{Q}_{iA} = -\frac{\partial U}{\partial Q_{iA}} = -\sum_{\alpha=1}^{N_{atoms}} \frac{\partial U}{\partial q_{i\alpha}} \frac{\partial q_{i\alpha}}{\partial Q_{iA}} = -\sum_{\alpha=1}^{N_{atoms}} C_{A\alpha} \frac{\partial U}{\partial q_{i\alpha}} \quad (4.19)$$

where the force on the normal modes can be expressed in terms of the charge forces using the coefficients from the matrix  $\mathbf{C}$  and also, from Equation 4.16 and  $q_{i1} = -q_{i2} - q_{i3}$ , the relation  $Q_{i1} = -q_{i1}$ . These values are  $C_{1\alpha}=(-1, 1/2, 1/2)$  and  $C_{2\alpha}=(0, 1/2, -1/2)$ . The Hamiltonian is

$$H = U + \frac{1}{2} \sum_i \sum_{\alpha=1}^{N_{atoms}} m_{\alpha} \dot{r}_{i\alpha}^2 + \frac{1}{2} \sum_i \sum_{\alpha=1}^{N_{atoms}-1} m_Q \dot{Q}_{i\alpha}^2 \quad (4.20)$$

The changes to the program involved in the three methods do not extend simulation times, the process of finding the normal modes and the optimal masses are in the set-up/parameterization stage. All of the simulations were done on 256 water molecules with SHAKE to enforce the bond constraints, Ewald sums for long-range electrostatics, in the canonical ensemble with a Nosé-Hoover temperature bath.<sup>9,17,18,20</sup> The data was gathered at 298K with using densities such that a pressure of 1 atm is achieved. Data was gathered in 10 runs of either 20ps or 100ps each.

### 4.3 Fluctuating Charge

The results of the various extended Lagrangian methods were compared using potential energy (E/N), average dipole moment ( $\langle\mu\rangle$ ), the root-mean-square of this dipole moment ( $\langle\mu^2\rangle^{1/2}$ ) and the translational diffusion constant (see Eq 4.21.

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle |r_i^{cm}(t) - r_i^{cm}(0)|^2 \rangle \quad (4.21)$$

As a source of comparison these values were calculated for a simulation performed with TIP4P-FQ using the iterative method without any corrections for the Lennard-Jones cutoffs (Table 4.1). As the time step,  $\delta t$ , increases the average energy increases slightly and

Table 4.1: Results for the TIP4P-FQ model for the potential energy per molecule, the average dipole moment, the root-mean-square of the dipole moment, and the translational diffusion constant as a function of the time step. Also shown is the average energy for the TIP4P-EW model. Using the Iterative method.

$\delta t$ (fs)	E/N (kcal/mol)	$\langle\mu\rangle$ (Debye)	$\langle\delta\mu^2\rangle^{1/2}$ (Debye)	D ( $10^{-9} \text{ m}^2/\text{s}$ )	E/N (TIP4P-EW) (kcal/mol)
0.5	-9.93(2)	2.643(3)	0.201(2)	2.0(5)	-11.53(1)
1.0	-9.93(4)	2.643(6)	0.202(1)	1.8(1)	-11.53(2)
1.25	-9.91(2)	2.640(3)	0.200(1)	2.0(1)	-11.51(2)
1.5	-9.91(2)	2.639(2)	0.201(1)	1.86(1)	-11.51(1)
1.75	-9.89(2)	2.637(4)	0.200(1)	1.9(1)	-11.50(2)
2.0	-9.88(2)	2.635(3)	0.201(1)	1.97(9)	-11.49(1)

average dipole moment decreases, this second value is consistent with the small increase in energy. If the calculations had included a correction for the Lennard-Jones cutoffs the drift in energy with time would likely have been less. The non-polarizable model also shows a shift in average energy, however this shift is on the order is approximately 100 cal/mol less than the shift for the TIP4P-FQ model.

When choosing a "mass" for the charge a balance must be found between to

opposing considerations. The mass must be small enough to respond quickly to changes in the environment thus staying close to the potential energy minima and limiting coupling between the charge degrees of freedom and the molecular degrees of freedom. Coupling between the degrees of freedom with result in the charge coordinate warming and losing contact with the energy minima. A mass that is too small requires a small time step. The relationship between mass and time step can be understood by a sine function. Ideally for a sine function there should be at least two evaluations per period. Given the sampling rate for the Hamiltonian of  $2\pi/\delta t$ , the Nyquist critical frequency for a 1 fs time step is  $\omega_N=16,600 \text{ cm}^{-1}$  for  $\delta t=1.25 \text{ fs}$ .<sup>22</sup> The results of insufficient sampling can be seen when the charge trajectories of the extended Lagrangian using the original charge mass and time step are plotted versus the trajectories for the iterative method starting from the same coordinates. The extended Lagrangian solution oscillates around the exact solution in a smooth sinusoidal function when plotting the trajectory of the m-site charge (Figure 4.1). However, when the trajectories of the charges on the protons are analyzed the high frequency oscillations are not smooth in appearance (Figure ??).<sup>11,14</sup>

The time scales for the charge oscillation are shown clearly by the frequency spectrum of the charge velocity autocorrelation function,

$$\tilde{C}_{\dot{q}_\alpha \dot{q}_\alpha}(\omega) = \int_0^\infty \left( \frac{1}{N} \sum_{i=1}^N \dot{q}_{i\alpha}(0) \dot{q}_{i\alpha}(t) \right) \cos(\omega t) dt \quad (4.22)$$

where  $\dot{q}_{i\alpha}(t)$  is the velocity of the charge of atom  $\alpha$  on molecule  $i$ . Time correlation functions have been used to characterize Car-Parrinello dynamics in the past.<sup>23</sup> The spectrum for extended Lagrangian charges using the original method with a time step of 1 fs and a charge mass of  $6.0 \times 10^{-5} \text{ (ps/e)}^2$  is shown in Figure 4.3. The low frequency region, up to  $2500 \text{ cm}^{-1}$ , corresponds to a charge response due to the molecular motion. While the lower frequency section of the protein is maintained throughout the different simulation methods, the higher frequency parts are from the fictitious nature of the extended Lagrangian charge dynamics. This region is broad, from  $4000$  to over  $12000 \text{ cm}^{-1}$  and this

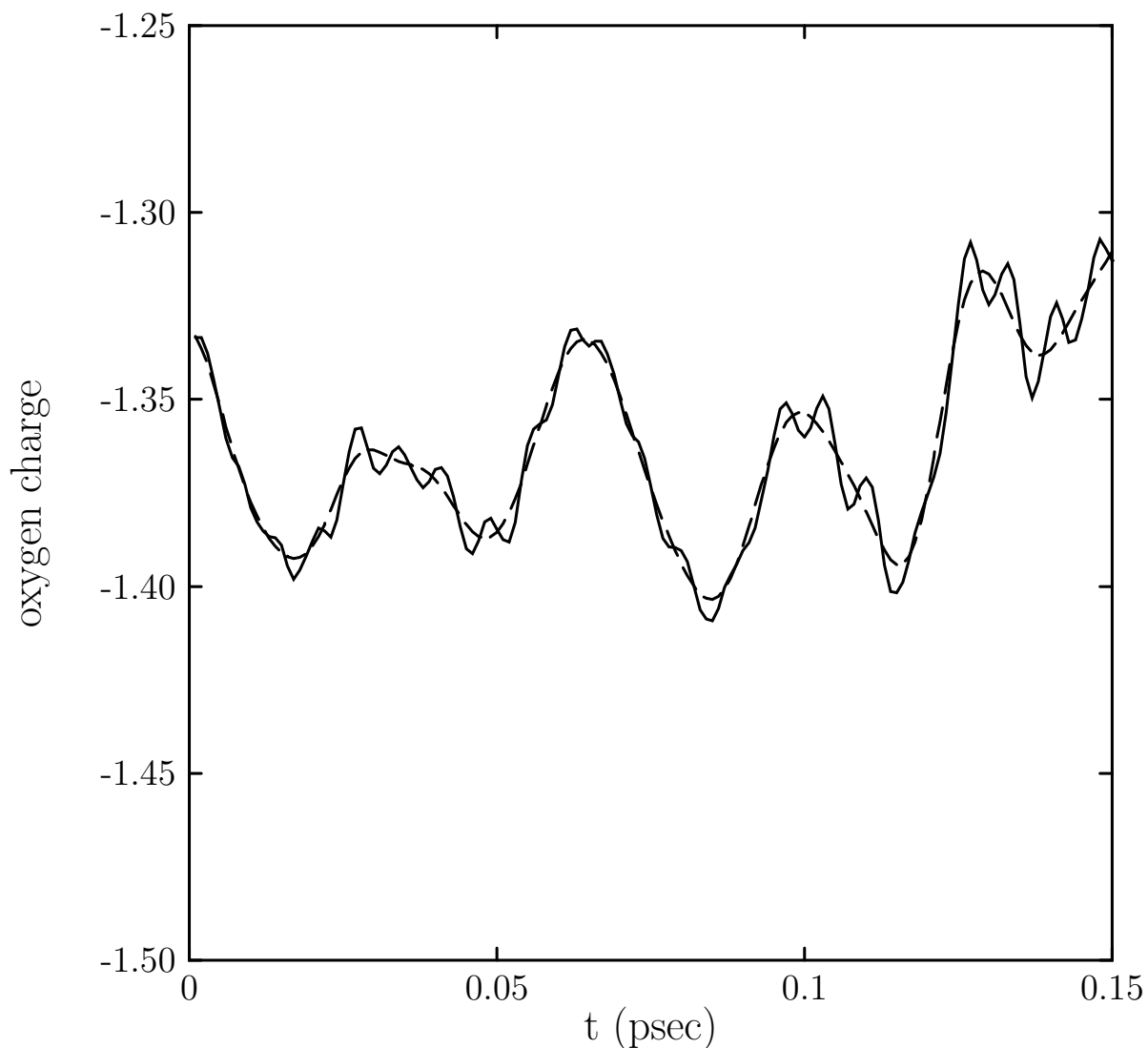


Figure 4.1: Oxygen atom charge (m-site) versus time comparing the exact (dashed line) and the extended Lagrangian values (solid line).

broadness does not allow for much flexibility in choosing the charge mass. A smaller charge mass will push the frequencies higher and the charge dynamics may become too fast to be used with a standard time step. A large charge mass will allow for a larger time step but will push the low frequency end of the spectrum towards the molecular frequencies. This will cause strong thermal coupling between the charge and molecular degrees-of-freedom, as well as a slowed response to changes in the molecular structure of the liquid. Both the

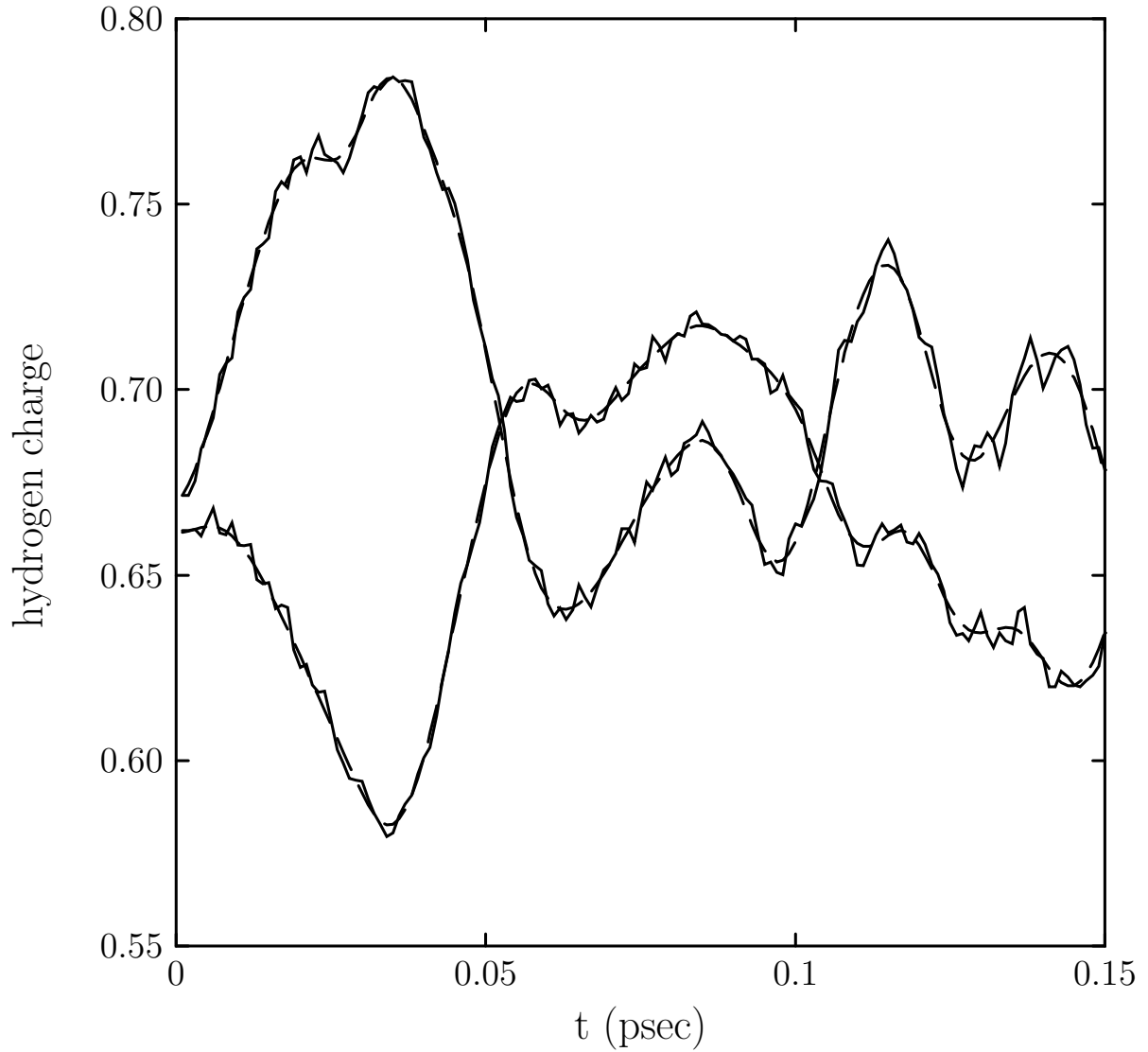


Figure 4.2: Hydrogen atom charge versus time comparing the exact (dashed line) and the extended Lagrangian values (solid line).

hydrogen and M-site charges show a feature around  $4000$  to  $6000\text{ cm}^{-1}$  and the hydrogen charges only show a peak starting at  $8000\text{ cm}^{-1}$ . This peak is in the region of the charge normal mode frequencies in the gas-phase, from Equations 4.17 and 4.18. By decreasing the M-site charge mass from  $6.0 \times 10^{-5}\text{ (ps/e)}^2$  to  $2.0 \times 10^{-5}\text{ (ps/e)}^2$  and increasing the hydrogen charge mass to  $1.4 \times 10^{-4}\text{ (ps/e)}^2$ , the frequencies can be brought together (Figure 4.4). The modification of the charge masses to bring the frequencies in the same

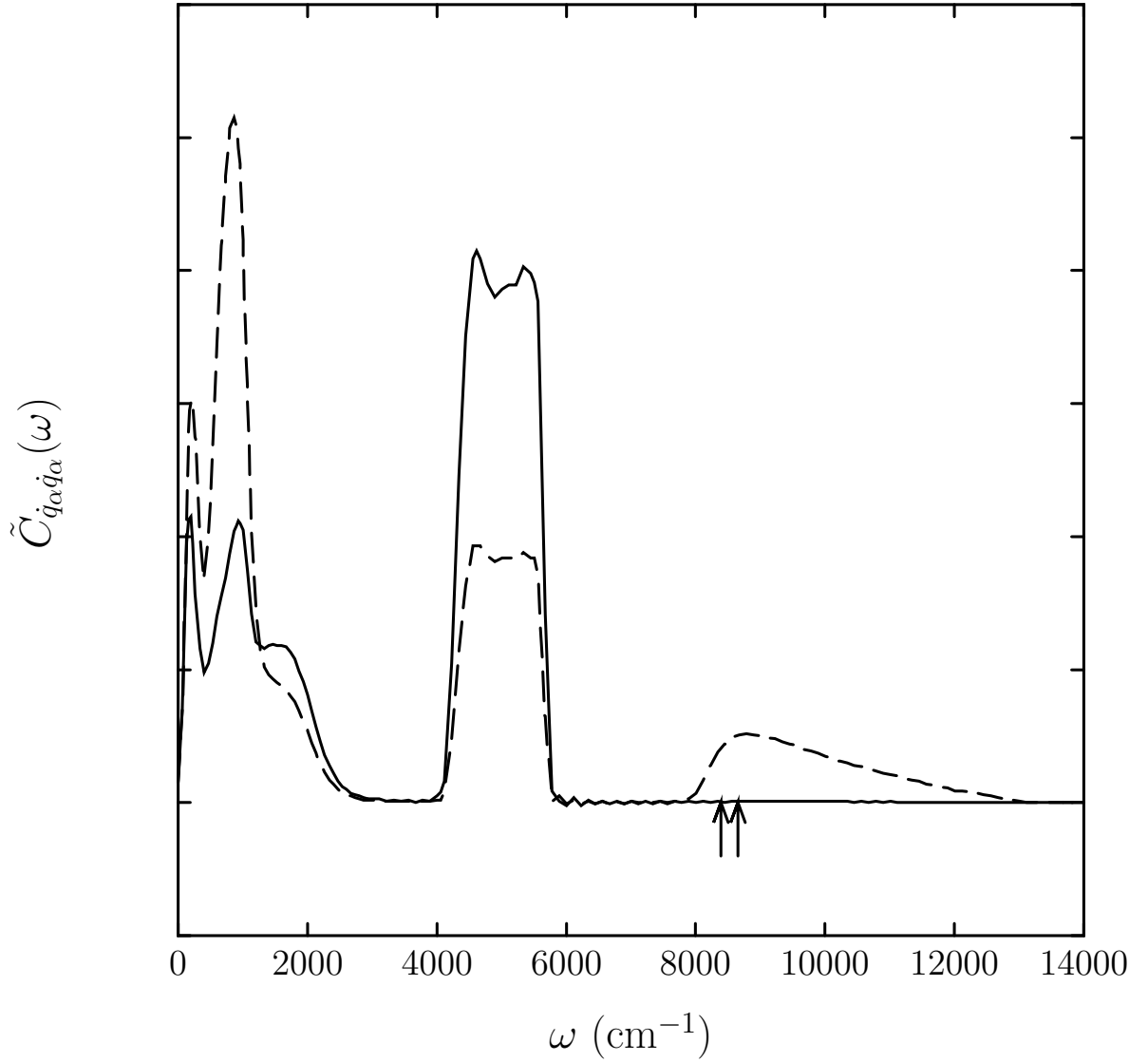


Figure 4.3: Frequency spectrum of the charge velocity autocorrelation function using the original extended Lagrangian method (Method 1) for the hydrogen atom (dashed line) and M-site charges (solid line). The arrows show the frequencies of the gas-phase charge normal modes for the system.

range was done by Stern, *et al.*<sup>13</sup> The normal mode method eliminates the peak around 5000  $\text{cm}^{-1}$ . Figure 4.5 shows the results using a normal mode charge mass of  $3.0 \times 10^{-5} \text{ (ps/e)}^2$ . This mass corresponds to the same charge mass as  $6.0 \times 10^{-5} \text{ (ps/e)}^2$  for Method 1. The absence of this peak for the normal mode method means that the charge mass can be decreased without thermal coupling between the charge and nuclear degree-of-freedom.

Table 4.2: Results for the TIP4P-FQ model using the extended Lagrangian method with a single charge mass,  $m_q$ .

$\delta t$ (fs)	$m_q$ ( $10^{-4}$ [ps/e] $^2$ )	E/N (kcal/mol)	$\langle\mu\rangle$ (Debye)	$\langle\delta\mu^2\rangle^{1/2}$ (Debye)	D ( $10^{-9}$ m $^2$ /s)	$T_q(t=20 \text{ ps})$ (K)
0.5	0.60	-9.92(2)	2.641(3)	0.202(3)	1.82(9)	3.0(1)
0.5	1.40	-9.80(4)	2.646(6)	0.236(2)	1.8(2)	120(5)
0.5	2.40	-9.57(3)	2.651(5)	0.264(2)	1.89(9)	241(7)
1.0	0.60	-9.90(2)	2.637(3)	0.199(1)	1.90(8)	2.5(2)
1.0	1.40	-9.81(3)	2.647(5)	0.236(3)	1.9(1)	104(4)
1.0	2.40	-9.59(3)	2.646(3)	0.265(2)	2.0(1)	219(10)
1.25	1.40	-9.78(1)	2.642(2)	0.236(1)	1.83(7)	100(5)
1.25	2.40	-9.60(2)	2.650(4)	0.265(2)	1.9(1)	205(7)
1.5	1.40	-9.75(3)	2.638(5)	0.235(2)	1.9(1)	87(5)
1.5	2.40	-9.56(3)	2.651(6)	0.270(2)	2.1(2)	192(7)
1.75	2.40	-9.54(4)	2.647(5)	0.272(1)	2.1(2)	184(8)
2.0	2.40	-9.49(1)	2.643(2)	0.266(7)	2.0(1)	157(5)

In order to use Method 1 with a larger time step, larger charge masses have to be used. If we want to keep the critical frequency the same relative to the fastest frequency of the system, then increasing  $\delta t$  by a factor of two would require increasing the charge mass by a factor of 4 to  $2.4 \times 10^{-4}$  (ps/e) $^2$ . With this charge mass, trajectories are stable with time steps up to 2 fs, but it does not result in accurate values for the average energy or other properties of the liquid (see Table 4.2). For a time up to 1.5 fs, stable trajectories can be achieved with a mass equal to  $1.4 \times 10^{-4}$  (ps/e) $^2$ , intermediate between  $6.0 \times 10^{-5}$  (ps/e) $^2$  and  $2.4 \times 10^{-5}$  (ps/e) $^2$ . These results are also not in good agreement with the results from Table 4.1. So Method 1 is not successful with any value of the charge mass with time steps of 1.25 fs or larger. The problems with the method can be seen in the charge temperature, given by

$$T_q = \sum_{i\alpha} \langle \dot{q}_{i\alpha} \rangle / kf \quad (4.23)$$

where  $f$  is the number of charge degrees of freedom, 2 times the number of molecules. The last column of Table 4.2 shows the  $T_q$  after a 20 ps trajectory. In all cases except for time steps of 0.5 or 1.0 fs and a charge mass of  $6.0 \times 10^{-5}$  (ps/e) $^2$ , the temperature gets large and

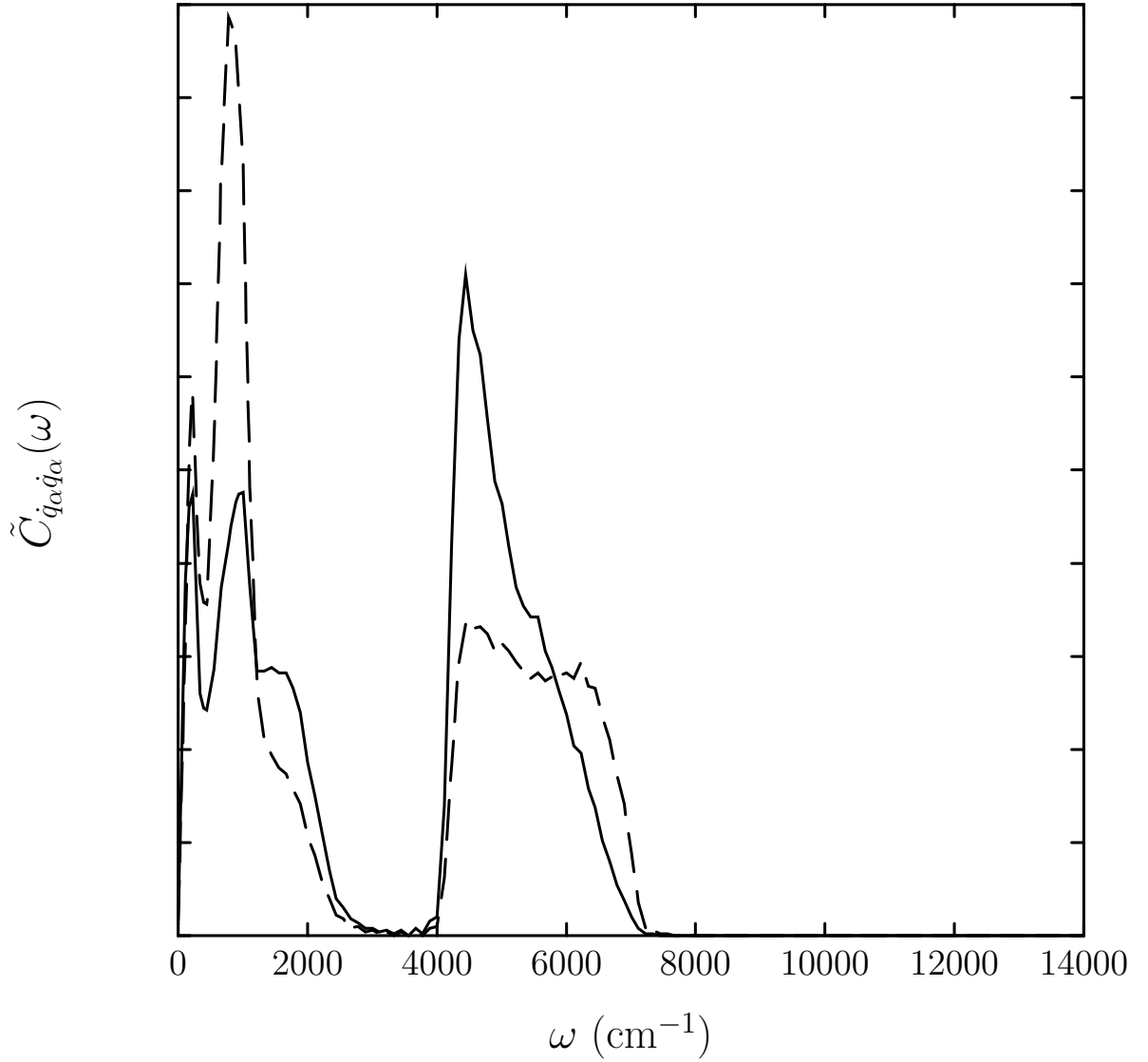


Figure 4.4: Frequency spectrum of the charge velocity autocorrelation function with different charge masses for the hydrogen (dashed line) and M-site charges (solid line).

approaches the temperature of the nuclear coordinates. After an initial jump from the starting point with charge velocities equal to zero,  $T_q$  only shows a slow increase as originally implemented with  $\delta t = 1$  fs and  $m_q = 6.0 \times 10^{-5}$  (ps/e)<sup>211</sup> (the solid line in Figure 4.6).

The other two methods work much better at larger time steps (Tables 4.3 and 4.4). Just like Method 1, in order to use larger time steps, the charge mass for the hydrogen



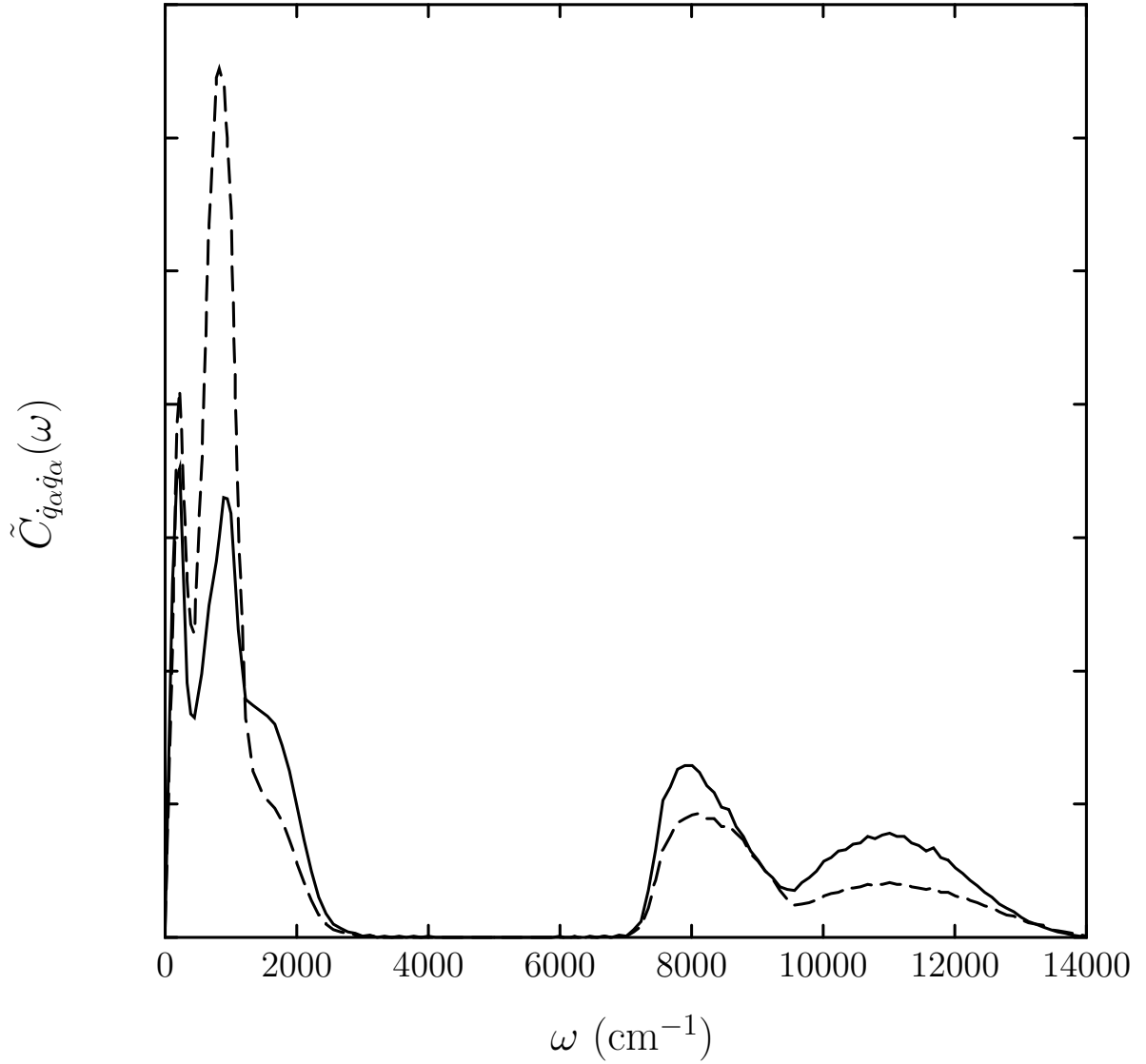


Figure 4.5: Frequency spectrum of the charge velocity autocorrelation function using charge normal modes for the hydrogen (dashed line) and M-site charges (solid line).

atom has to be increased. For a 2 fs time step, the mass of the hydrogen charge has to be at least  $2.4 \times 10^{-4} \text{ (ps/e)}^2$ , which corresponds to a normal mode charge mass,  $m_Q$ , equal to  $1.2 \times 10^{-4} \text{ (ps/e)}^2$ . Different values for the M-site charge mass were tried as well, but a value of  $2.0 \times 10^{-5} \text{ (ps/e)}^2$  was found to be optimal. At this time step both these methods give energies of  $-9.83 \pm 0.03 \text{ kcal/mol}$  slightly higher than the value the iterative method gives at the same time step ( $-9.90 \pm 0.04 \text{ kcal/mol}$ ). Other properties, including the average

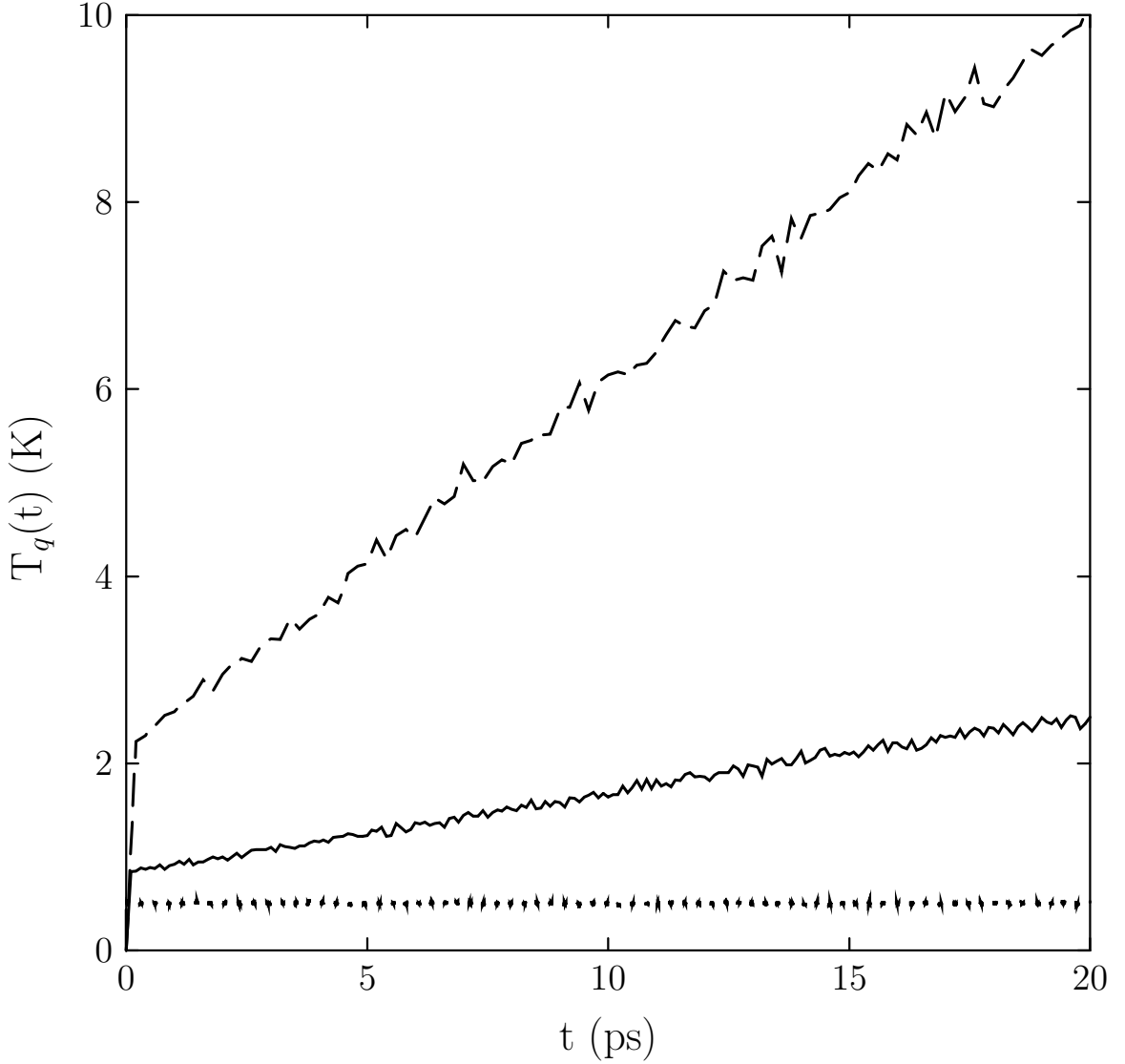


Figure 4.6: The average time dependence of the charge temperature with a single charge mass of  $6.0 \times 10^{-5} \text{ (ps/e)}^2$  and  $\delta t = 1.0 \text{ fs}$  (solid line), charge normal modes with a charge mass of  $3.0 \times 10^{-5} \text{ (ps/e)}^2$  and  $\delta t = 1.0 \text{ fs}$  (dotted line), and charge normal modes with a charge mass of  $1.2 \times 10^{-4} \text{ (ps/e)}^2$  and  $\delta t = 2.0 \text{ fs}$  (dashed line).

dipole moment,  $\langle \mu \rangle$ , the root-mean-square of the dipole moment, and the diffusion constant are also in close agreement with both the iterative method with  $\delta t = 2 \text{ fs}$  and the original implementation with  $\delta t = 1 \text{ fs}$  and a single charge mass equal to  $6.0 \times 10^{-5} \text{ (ps/e)}^2$ . The average pressure was also calculated and did not show any differences among the methods. The structure of the liquid is also in good agreement (Figure 4.7). The  $g_{OO}(r)$  for the

Table 4.3: Results for the TIP4P-FQ model using the extended Lagrangian method with different charge masses for the two charge sites, M and H.

$\delta t$ (fs)	$m_{qM}:m_{qH}$ ( $10^{-4}$ [ps/e] <sup>2</sup> )	E/N (kcal/mol)	$\langle\mu\rangle$ (Debye)	$\langle\delta\mu^2\rangle^{1/2}$ (Debye)	D ( $10^{-9}$ m <sup>2</sup> /s)	$T_q(t=20 \text{ ps})$ (K)
0.5	0.20 : 1.40	-9.92(3)	2.642(4)	0.203(1)	1.88(8)	5.2(4)
0.5	0.20 : 2.40	-9.92(3)	2.646(5)	0.206(4)	1.7(2)	23(1)
1.0	0.20 : 1.40	-9.90(4)	2.639(5)	0.201(1)	1.9(2)	4.2(3)
1.0	0.20 : 2.40	-9.88(4)	2.645(6)	0.204(2)	1.9(2)	37(1)
1.25	0.20 : 1.40	-9.91(3)	2.640(4)	0.202(2)	2.0(1)	3.5(2)
1.25	0.20 : 2.40	-9.88(2)	2.647(4)	0.205(5)	1.9(1)	33(6)
1.5	0.20 : 1.40	-9.88(2)	2.637(3)	0.202(1)	1.9(1)	2.8(1)
1.5	0.20 : 2.40	-9.88(2)	2.644(3)	0.208(1)	1.8(1)	29(1)
1.75	0.20 : 2.40	-9.84(2)	2.637(4)	0.207(1)	1.8(2)	22(1)
2.0	0.20 : 2.40	-9.83(3)	2.642(5)	0.203(2)	2.0(1)	17.4(8)

method using charge normal modes (with  $\delta t=2$  fs and  $M_Q=1.2 \times 10^{-4}$  (ps/e)<sup>2</sup>) is almost indistinguishable from the  $g_{OO}(r)$  using the original method, with slightly smaller peak heights. The other correlation functions  $g_{HH}(r)$  and  $g_{OH}(r)$  are in good agreement as well (data not shown). For all the methods listed in Tables 4.2 and 4.3, the pair correlation functions are very close to each other.

The charge temperature rises fairly linearly with time (Figure 4.6) and the effects of the amount of thermal coupling to the charges are more apparent in simulations longer than 20 ps. In Method 1 with  $\delta t=1$  fs and  $M_Q=6.0 \times 10^{-5}$  (ps/e)<sup>2</sup> the charge temperature after 100 ps is only 9.2 K (from an average of ten 100 ps runs). The averages from the 100 ps simulation are close to those from the 20 ps simulation ( $E/N=-9.86 \pm 0.01$  kcal/mol). For Method 3 with  $\delta t=1$  fs and  $M_Q=3.0 \times 10^{-5}$  (ps/e)<sup>2</sup> the charge temperature increases very slowly and after 100 ps is only 0.5 K. The results using this method,  $\delta t$ , and  $M_Q$  from ten 100 ps simulations are identical to the results from ten 20 ps simulations. Similarly, using Method 3 with a time step of 1.5 fs and  $M_Q=7.0 \times 10^{-5}$  (ps/e)<sup>2</sup> gives a value for  $T_q$  of 3.0 K after 100 ps and the simulation gives the same results as the shorter simulations. In cases where the  $T_q$  increases more quickly, the results will depend on the simulation length. The dashed line in Figure 4.6 is for Method 3 with  $\delta t=2$  fs and  $M_Q=2.4 \times 10^{-4}$  (ps/e)<sup>2</sup>. From

Table 4.4: Results for the TIP4P-FQ model using the extended Lagrangian method with charge normal modes.

$\delta t$ (fs)	$m_Q$ ( $10^{-4}$ [ps/e] <sup>2</sup> )	E/N (kcal/mol)	$\langle\mu\rangle$ (Debye)	$\langle\delta\mu^2\rangle^{1/2}$ (Debye)	D ( $10^{-9}$ m <sup>2</sup> /s)	$T_Q(t=20 \text{ ps})$ (K)
0.5	0.30	-9.95(3)	2.646(5)	0.203(1)	1.8(1)	0.50(2)
0.5	0.70	-9.93(3)	2.643(5)	0.202(1)	1.9(1)	2.5(1)
0.5	1.20	-9.90(3)	2.643(6)	0.205(1)	1.9(1)	28(2)
1.0	0.30	-9.91(2)	2.642(4)	0.202(4)	1.9(1)	0.52(1)
1.0	0.70	-9.93(3)	2.638(2)	0.200(1)	1.8(1)	2.2(2)
1.0	1.20	-9.90(2)	2.643(3)	0.205(1)	1.86(8)	22.2(8)
1.25	0.70	-9.89(3)	2.637(5)	0.199(1)	2.0(1)	1.7(1)
1.25	1.20	-9.87(3)	2.641(5)	0.204(2)	1.9(1)	18(1)
1.5	0.70	-9.91(2)	2.640(3)	0.202(1)	1.9(2)	1.5(1)
1.5	1.20	-9.89(3)	2.642(4)	0.205(1)	1.9(1)	17.2(8)
1.75	1.20	-9.88(2)	2.640(3)	0.205(1)	1.8(2)	13.7(8)
2.0	1.20	-9.83(3)	2.643(5)	0.203(6)	2.0(1)	10.1(4)

simulations of 20 ps,  $T_q$  stays under 10 K and the results are close to the iterative method results at the same time step. After 100 ps,  $T_q$  will reach about 40 K and the resulting properties are much different ( $E/N=-9.755\pm0.009$  kcal/mol). It appears that  $T_q$  should be kept below some threshold, perhaps around 10 K. The temperature can be kept low by quenching the charges to their minimum energy values when  $T_q$  exceeds some limit<sup>24</sup> or at periodic intervals.<sup>25</sup> Using Method 3 ( $\delta t=2$  fs,  $M_Q=2.4\times10^{-4}$  (ps/e)<sup>2</sup>) and quenching the charges every 10 ps over 100 ps simulations (keeping  $T_q < 10$  K) gives results that are again close to the iterative method results ( $E/N=-9.86\pm0.01$  kcal/mol). The charge temperature can also be kept low using thermostatting,<sup>12,26</sup> although this was not done in any of the present simulations.

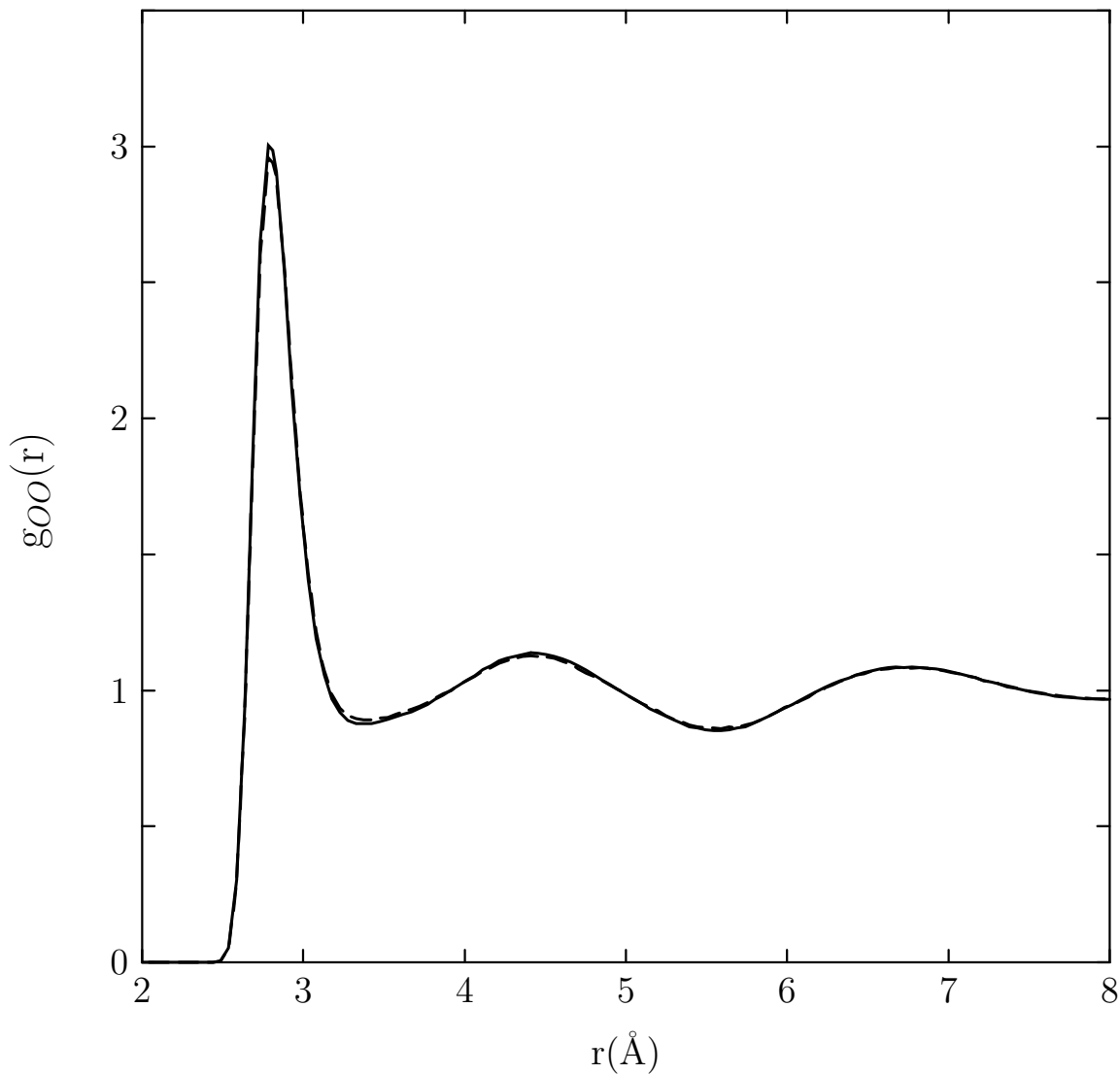


Figure 4.7: The oxygen-oxygen correlation function, comparing the results using a 1 fs time step with a single charge mass for both charge types equal to  $6.0 \times 10^{-5} \text{ (ps/e)}^2$  (solid line) and using a 2 fs time step with the normal mode method for charge dynamics and a charge mass equal to  $1.2 \times 10^{-4} \text{ (ps/e)}^2$  (dashed line).

## 4.4 Summary

At a 1 ps time step all methods 2 and 3 give results with good agreement to the original implementation of fluctuating charge (method 1). While all three methods are similar in terms of energetics, the charge normal modes maintain a lower temperature throughout the

simulation, however both methods 1 and 2 have final temperatures that do not exceed the 10 K upper limit to stay close to the potential energy minimum. Accurate results can be obtained up to  $\delta t=1.5$  fs, past this point the coupling between the charge and atomic degrees of freedom becomes an issue resulting in an increase in potential energy.

While the temperature of the charge may cause problems as the time step increases it is not an insurmountable problem. The temperature is still within the low temperature range at 10 ps, this allows for a combination of the extended Lagrangian and the iterative method to be employed at larger time steps. The extended Lagrangian can be used in 10 ps simulations, or another appropriate value, at which time the system can be annealed to the exact charges using the iterative method. This would only slightly increase the simulation times associated with fluctuating charge method. However, as the time step for these simulations this would still be a net decrease in computational cost. The other possibility is thermostatting the charges.

While the charge normal modes give good results at larger values of  $\delta t$ , this is not necessarily the best option for protein simulations. The use of normal modes that are not coupled might decrease method 3's charge temperature at  $\delta t=2$  ps, however the process of adding normal modes to protein simulations is not trivial. The geometry of each amino acid would help determine the appropriate diagonalizing matrix, and any changes in overall geometry during the course of the simulation would result in changes to the normal modes.

# Bibliography

- [1] Rick, S. W.; Stuart, S. J. Potentials and Algorithms for Incorporating Polarizability in Computer simulations. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Ed.; volume 18 John Wiley and Sons, Inc., New York 2002; pp 89–146.
- [2] Patel, S.; III, C. L. B. *J. Computational Chemistry* **2004**, *25*, 1–16.
- [3] Jorgensen, W. L.; Gao, J. *J. AM. Chem. Soc.* **1988**, *110*, 4212–4216.
- [4] Williams, D. E. *Biopolymers* **1990**, *29*, 1367–1386.
- [5] Dinur, U.; Hagler, A. T. *J. Comp. Chem.* **1995**, *16*, 154–170.
- [6] Urban, J. J.; Famini, G. R. *J. Comp. Chem.* **1993**, *14*, 353–362.
- [7] Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1994**, *116*, 3949.
- [8] Cieplak, P.; Kollman, P. *J. Comp. Chem.* **1991**, *12*, 1232–1236.
- [9] Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- [10] Vassilev, P.; Hartnig, C. B.; Koper, M. T. .; Frechard, F.; van Santen, R. A. *J. Chem. Phys* **2001**, *115*, 9815–9820.
- [11] Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.
- [12] Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, *118*, 672–679.
- [13] Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys* **2001**, *115*, 2237–2251.

- [14] Olano, L. R.; Rick, S. W. *J. Computational Chemistry* **2005**, *26*, 699–707.
- [15] Rick, S. W.; Berne, B. J. *J. Phys. Chem. B* **1997**, *101*, 10488.
- [16] Ribeiro, M. C. C.; Almeida, L. C. J. *J. Chem. Phys.* **1999**, *110*, 11445–11448.
- [17] Patel, S.; Mackerell, A. D.; III, C. L. B. *J Computational Chemistry* **2004**, *25*, 1504.
- [18] Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
- [19] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- [20] Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- [21] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*. Oxford University Press, Oxford 1987.
- [22] Olano, L. R.; Rick, S. W. *J. Am. Chem. Soc.* **2004**, *126*, 7991–8000.
- [23] Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *120*, 300–311.
- [24] Stuart, S. J.; Berne, B. J. *J. Phys. Chem.* **1996**, *100*, 11934–11943.
- [25] Rick, S. W. *J. Chem. Phys.* **2001**, *114*, 2276–2283.
- [26] Stöckelmann, E.; Hentschke, R. *J. Chem. Phys.* **1999**, *110*, 12097–12107.

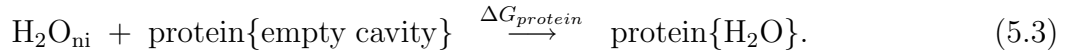
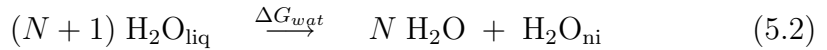
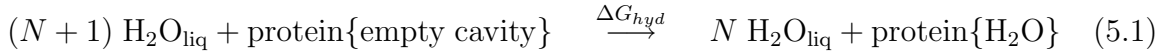


# Chapter 5

## Experimental

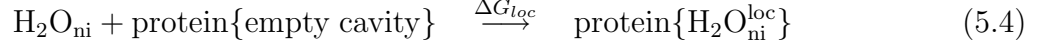
### 5.1 Protein Simulations

The free energy of hydration is a state function and thus it can be expressed as the sum of two terms,  $\Delta G_{protein}$ . The first,  $\Delta G_{wat}$ , involves the removal of a water molecule from a box of pure liquid water (1). While the second,  $\Delta G_{hyd}$  is the introduction of a water molecule into a protein cavity(2).<sup>1</sup>



where  $\text{H}_2\text{O}_{\text{liq}}$  is liquid water and  $\text{H}_2\text{O}_{\text{ni}}$  is a non-interacting water molecule. To correctly and reversibly calculate the free energy for Equation 5.3, the simulation would have to allow for the water molecule to search the entire simulation box, which would require prohibitively long simulation times. Also, if  $\Delta G_{protein}$  is positive, the water molecule will not occupy the cavity of interest. These problems can be avoided by first localizing the molecule in the binding site, then allowing the molecule to interact with the protein.<sup>1,3-6</sup>

The process represented by Equation 5.3 becomes a two step process,



where  $\text{H}_2\text{O}_{\text{ni}}^{\text{loc}}$  indicates a non-interacting molecule which is localized in the protein cavity.

If the molecule is localized using a harmonic potential,

$$U_{harm}(r) = k_{harm} (r_O - r_X)^2 \quad (5.6)$$

where  $k_{harm}$  is the force constant,  $r_O$  is the position of the oxygen atom of the water molecule, and  $r_X$  is the center of the binding cite, then the free energy to localize the molecule is given by<sup>4-6</sup>

$$\Delta G_{loc} = -kT \ln[\rho(\pi kT/k_{harm})^{3/2}] \quad (5.7)$$

where  $k$  is Boltzmann's constant,  $T$  is temperature, and  $\rho$  is the bulk density of water.

This method correctly satisfies the conditions regarding a standard state which is the pure liquid of density  $\rho$ .<sup>1,3-6</sup> Note that some previous studies of cavity hydration<sup>1,7,8</sup> have not satisfied conditions of the standard state and reversibility.<sup>4,6</sup> Simulations of BPTI were run without restraint to find the mean square fluctuation in the oxygen position,  $\langle \delta r^2 \rangle$ . The optimal force constant, was found from  $k_{harm} = 6kT/\langle \delta r^2 \rangle$  to give  $k_{harm} = 3 \text{ kcal/mol/\AA}^2$ .<sup>5</sup>

Because the protein may rotate and translate during the course of the simulation, the center of the restraint potential,  $r_o$ , must rotate along with the protein. This was implemented by rotating the crystal structure geometry onto the simulation structure every time step by minimizing the root-mean-square deviation between the  $\text{C}_\alpha$  atoms. The value of  $r_X$  is taken to be the location of the oxygen atom of the crystal water, after the structure is rotated onto the current coordinates. The free energy for the hydration process

is then given by

$$\Delta G_{hyd} = \Delta G_{wat} + \Delta G_{protein} = \Delta G_{wat} + \Delta G_{loc} + \Delta G_{inter} \quad (5.8)$$

The initial state of Eq. 1 is the empty cavity. To ensure that other water molecules do not enter the cavity as  $\lambda$  goes to zero, an additional short-ranged interaction between the position of the reference cavity water oxygen and the oxygen atoms on the solvent water is added with this form,  $U_r = \sum_k \epsilon_r (r_{Ok}/\sigma_r)^{-12}$ .

To calculate the free energy of the process given by Equation 5.5, free energy perturbation can be used, with a potential energy of the water molecule in the binding pocket is given by

$$U_\lambda = \lambda \left[ \sum_j 4\epsilon_{Oj} \left[ \left( \frac{r_{Oj}}{\sigma_{Oj}} \right)^{-12} - \left( \frac{r_{Oj}}{\sigma_{Oj}} \right)^{-6} \right] + \sum_{i=1}^3 \sum_j q_i q_j / r_{ij} \right] \\ + (1 - \lambda) \sum_k \epsilon_r \left( \frac{r_{Ok}}{\sigma_r} \right)^{-12} + (1 - \lambda) k_{harm} (r_O - r_X)^2 \quad (5.9)$$

where O denotes the position of the oxygen atom of the cavity water, the sum over  $i$  is over the atoms of the cavity water, the sum over  $j$  is over all the other atoms (all protein, ions, and all other water molecules) and the sum over  $k$  is over the oxygen positions of all other water molecules. The parameter  $\lambda$  therefore scales in the Lennard-Jones and electrostatic interactions of the cavity water molecule with the surrounding molecules while simultaneously scaling out both the repulsive term keeping out other water molecules from the site ( $\epsilon_r=0.152$  kcal and  $\sigma_r=2.0$  Å) and the harmonic restraint term. The protein is fully hydrated at  $\lambda = 1$ .

The calculation of  $\Delta G$  from free energy perturbation requires the calculation of averages of  $\langle e^{-(U_{\lambda_i} - U_{\lambda_{i+1}})/kT} \rangle_{U_\lambda}$  (see, for example, Allen and Tildesley<sup>9</sup>). The bias from the

repulsive potential  $U_r$  keeping the cavity empty as  $\lambda$  goes to zero can be corrected using

$$\langle e^{-(U_{\lambda_i}-U_{\lambda_{i+1}})/kT} \rangle_{U_\lambda}^0 = \frac{\langle e^{-(U_{\lambda_i}-U_{\lambda_{i+1}})/kT} e^{(1-\lambda_i)U_r/kT} \rangle_{U_\lambda}}{\langle e^{(1-\lambda_i)U_r/kT} \rangle_{U_\lambda}} \quad (5.10)$$

where  $\langle \dots \rangle^0$  denotes the average without  $U_r$ . If the value of the perturbation term  $U_{\lambda_i} - U_{\lambda_{i+1}}$  is uncorrelated with  $U_r$  then

$$\langle e^{-(U_{\lambda_i}-U_{\lambda_{i+1}})/kT} \rangle_{U_\lambda}^0 = \frac{\langle e^{-(U_{\lambda_i}-U_{\lambda_{i+1}})/kT} \rangle_{U_\lambda} \langle e^{(1-\lambda_i)U_r/kT} \rangle_{U_\lambda}}{\langle e^{(1-\lambda_i)U_r/kT} \rangle_{U_\lambda}} = \langle e^{-(U_{\lambda_i}-U_{\lambda_{i+1}})/kT} \rangle_{U_\lambda} \quad (5.11)$$

This is the assumption we use. For the barnase cavity, water molecules never attempt to enter so  $U_r$  is never much greater than zero. For BPTI, this term is required, but is effectively zero except at small  $\lambda$  values. At larger values of  $\lambda$ , the interactions with the cavity water are sufficient to keep other molecules out of the cavity.

All protein molecular dynamics were performed using the Amber6 suite of programs with the Cornell *et al.* 1994 force field<sup>3</sup> and TIP3P water.<sup>11</sup> The protein contributions to the free energies,  $\Delta G_{proteins}$ , were determined with free energy perturbation in the module Gibbs from the Amber 6.0 suite of programs.<sup>12</sup> A minimum of 500 ps of simulation was performed at each of 12  $\lambda$  values ranging from  $\lambda = 0.95$  to  $\lambda = 0.025$  with a prior 20 ps of equilibration performed at each  $\lambda$  value. In the wild-type BPTI (PDB entry 5PTI) and the lysozyme mutant (PDB entry 2HEA) all  $\lambda$  values were equilibrated for 220 ps. The barnase mutant (PDB entry 1BRI) 298K simulations included an additional 500 ps of molecular dynamics on all values of  $\lambda$  from a different equilibration point, as well as an additional 2000 ps at the end points in an attempt to minimize error in the free energy and other structural data. Only forward (insertion) values of  $\Delta G$  were used in the protein simulations with the exception of  $0.025 \rightarrow 0.0$  where the backward (deletion) value was used due to increased noise at the endpoints. Insertion and deletion free energies were not averaged as the magnitude of errors in the two measurements are not identical.<sup>13</sup> The insertion free energy values were less noisy than their deletion counterparts, but values for

$\Delta G_{hyd}$  using both forward and reverse as well as half steps were very similar. To ensure a complete set of data the amber code was altered to output data required to calculate  $\Delta G_{\lambda+\Delta\lambda}$ ,  $\Delta G_{\lambda-\Delta\lambda}$ , as well as  $\Delta G_{\lambda+(\Delta\lambda)/2}$ , and  $\Delta G_{\lambda-(\Delta\lambda)/2}$ .

The free energy for the removal of a water molecule from the pure liquid,  $\Delta G_{wat}$ , was calculated using our group’s own program, using the separated-shifted scaling method to avoid singularities.<sup>14,15</sup> The free energy calculations used 12  $\lambda$  values, ranging from 0.05 to 0.95, and ran for 5 nanoseconds at each  $\lambda$  value. These simulations used 256 molecules, a 1 fs time step, and SHAKE for bond constraints.<sup>9</sup> The simulations were done in the isothermal-isobaric (constant T,P,N) ensemble, by coupling to a pressure bath and a Nosé-Hoover temperature bath.<sup>16–18</sup>

The entropy can be found from a finite difference approximation of the temperature derivative which requires calculating the free energy at two different temperatures ( $T \pm \Delta T$ ),<sup>19</sup> given by

$$\Delta S = -(\Delta G(T + \Delta T) - \Delta G(T - \Delta T))/(2\Delta T). \quad (5.12)$$

The entropies are about an order of magnitude more uncertain than the  $\Delta G$  and so require longer simulations.<sup>19</sup> Using a finite difference approximation to the entropy is equivalent to assuming that the free energy is linear over this temperature range. Therefore, rather than calculating  $\Delta S$  through Equation 5.12, the free energy at the three temperatures ( $T - \Delta T, T, T + \Delta T$ ) can be fit to a line and the slope of the line can be used to get  $\Delta S$ . A temperature difference of 15 Kelvin is used, which in previous studies of aqueous solvation has been shown to be effective.<sup>19–21</sup> A similar method has been used to estimate the entropy of binding between compounds and nucleic acids.<sup>22</sup>

Structure files of the I76A mutant of barnase, the I106A mutant of lysozyme and wild-type BPTI were obtained from the protein data bank and the original counter ions were removed. Three proteins are in the unit cell of the 1BRI structure, only the third (labeled C) was used. In barnase the histidine residues were assumed to be neutral and

changed to reflect protonation of the  $N_\delta$  (HID). Lysozyme also had histidines but residues were protonated at the epsilon position of the  $N_\delta$  (HIE). In BPTI and lysozyme all the original crystal waters were kept while for barnase only those waters surrounding the protein chain C were retained. The protein data bank files were first loaded into the xLeap program within the Amber6 package to ensure no residues were absent. The resulting coordinate system was saved in protein data bank (pdb) format and protonated using the appropriately named PROTONATE program. The subsequent protonated pdb file was again loading into xLeap. The proteins were loaded into xLeap, and cross-linking of the appropriate cysteine residues was performed for BPTI and lysozyme using the connect command. Chlorine ions were added to create a neutral simulation box, 2 ions were added to barnase, 8 for lysozyme and 6 were added to BPTI. Additional solvent waters were added to create an 8 Å box around the three proteins, 3132, 6318 and 4769 residues were added to BPTI, lysozyme and barnase, respectively. At this point the required coordinate and topology files were created, as was the topology file in which the buried water is made non-interacting. Simulations were performed with particle-mesh Ewald. A 7 Å cutoff for non-bonded pairs during simulations for Lennard-Jones and real space Ewald interactions.

Steepest descent minimization was performed for 10 ps to eliminate bad contacts and the proteins were warmed to 298 K over the course of 26 ps for BPTI and 140 ps for the larger proteins barnase and lysozyme using the Sander program in the canonical ensemble (constant T,V,n). After equilibrating at 298 for 20 ps, an additional 20 ps of isothermal-isobaric equilibration was performed at 283, 298, and 313 K.

A sample input for free energy calculations using the Gibbs program follows. This example is for the  $\lambda = 0.6$ .

Gibb's input for 2HEA

```
&cntrl
  irest = 0, ibelly=0, iewald=1,
  ntx = 7, ntxo=1,
  tempi=298., heat = 0.000000E+00,
  ntb=2, iftres=1,ibxrd = 0,
```

```

nrun = 1, NTT=5, , TEMP0=298.,
dtemp =10, tautp=0.5, tauts=0.5,
isolv=2023, nsel = 0, dtuse=0.5,
ntp=1, npscal=0, pres0=1.0,
comp=44.6000, taup = 0.6,
ndfmin = 0, ntc=1, nscm=-1,
isvat=1, nstlim=50000, INIT = 4,
T=0.0, DT = 0.001,
ivemax = 0,
ntc=3,
tol=0.00001, tol2 = 0.001, ncorc=0,
ishkfl = 1, itimth=0, jfastw = 0,
ntf=3, ntid=0, ntnb=1, nsnb = 50,
ielper=0, imgslt=0,
idsx0=0, itrslu=1, ioleps=0,
intprt=1, itip=0,
cut=7.0, scnb = 2.0, scee= 1.2,
dielc = 1.0, cutprt = 0.0,
ntpr=1000, ntwx=1000, ntwv=-1, ntwe=-1,
isande=1, iperat = 0, iatcmp= 0,
ntatdp = 0, ntwprt=0,
ntr=1, ntrx = 1,
taur = 1.0, intr = 0, ibigm=1,
isftrp=0, rwell = 5.0,
ctimt = 0.0, isldyn=-3,
almda=0.60, almdel = 0.10, idifrg=0,
nstmeq=0, nstmul = 50000, ndmpmc=10,
islp=0, corrs1 = 0.8, almdl0=0.001, amxdel=0.1,
idiel=1,idwide = 0,
&end
63.530530 59.651815 58.422944 90.0000000 90.0000000 90.0000000
64 60 60 3 0 0
0.000001
Title, restraining calphas
0.0
FIND
CA * * *
SEARCH
RES 1 130
END
Title, wtp
3.0
ATOM 2021
END
END

```

Temperature coupling during the constant T, P, n simulations was performed using Berendsen coupling<sup>23</sup> with the temperature relaxation time  $\tau_T$  optimized for each structure in an effort to maintain good temperature control for the cavity water. The resulting couplings were  $\tau_T = 0.5$  ps for barnase and lysozyme and  $\tau_T = 0.2$  ps for 5PTI. The cavity water was considered part of the solute for all calculations. The constant pressure calculations were also performed with isotropic position scaling and SHAKE was used to constrain all bonds. Trajectories were performed in 50 ps segments, each containing 50,000 steps of 1 fs each and coordinates and energy data was sent to output every 1 ps (1000 steps). This allowed for limited loss of data should a crash occur and a larger set of energy data. Thus, a 500 ps simulation, the smallest performed, involved 10 runs per  $\lambda$  value resulting in 120 trajectories over the course of the calculation.

Immediately following the main control statements is the data needed to run particle-mesh Ewald. The length in angstroms of each side of the simulation box is given follows by the angle, as these simulations were performed within a rectangle the angles involved were 90 degrees. The subsequent line contains the values for NFFT 1, 2 and 3 which are used to define the charge grid. These values were as close to the box lengths as possible to give an approximately 1Å grid for the reciprocal space calculations. Those calculations involve the use of the fast Fourier transform and as such the NFFT value is chosen to be a product of 2, 3 and 5 or the powers of those numbers. Also required for the calculation is a value for the order of the calculation, in this case a fairly small value 3 was chosen, the cost of the PME calculation is determined in large part by this value (Order<sup>3</sup>). The tolerance was set to  $10^{-6}$  and any charges caused by roundoff errors were neutralized prior to each run.<sup>12</sup> The final lines are involved in calculating the restraints defined earlier.

Calculations at 3 kbar were performed using the same methodology. However the spring constant for the harmonic constraint holding the water in the cavity was increased to ensure it did not leave the pocket during the equilibration to the higher pressure. This was scaled back to the same spring constant in the 1 bar studies after the pressure was



finished equilibrating. The protein calculations at high pressure were performed for 500-1000 ps of simulation time. Longer simulations were used primarily at the end points of the simulation and lower values of  $\lambda$  where noise becomes a problem and flexibility measurements require more data points. Special consideration was given to ensure that the water did not exit the cavity while the water was scaled out, as at increased pressure this was a possibility despite the harmonic constraint. When this occurred the simulation was equilibrated from a different point and the calculations were restarted.

# Bibliography

- [1] Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. . *Amer.. Am. Chem. Soc.* **1990**, *112*, 7057–7059.
- [2] Zhang, L.; Hermans, J. *Proteins: Struct. Funct. Gen.* **1996**, *24*, 433–438.
- [3] Mann, G.; Hermans, J. *J. Mol. Bio.* **2000**, *302*, 979–989.
- [4] Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.
- [5] Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. *Biophys J* **1996**, *71*, 670–681.
- [6] Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- [7] Helms, V.; Wade, R. C. *Biophys. J.* **1995**, *69*, 810–824.
- [8] Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. *Biopolymers* **1991**, *31*, 919–931.
- [9] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*. Oxford University Press, Oxford 1987.
- [10] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [11] Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.

- [12] Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. I.; Vincent, J. J.; Crowlet, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. Amber 6, University of California, San Francisco, 1999.
- [13] Lu, N.; Kofke, D. A. *J. Chem. Phys.* **2001**, *114*, 7303–7311.
- [14] Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
- [15] Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1994**, *116*, 3949.
- [16] Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.
- [17] Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
- [18] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- [19] Smith, D. E.; Haymet, A. D. J. *J. Chem. Phys.* **1993**, *98*, 6445–6454.
- [20] Rick, S. W. *J. Phys. Chem. B* **2000**, *104*, 6884–6888.
- [21] Lüdemann, S.; Schreiber, H.; Abseher, R.; Steinhauser, O. *J. Chem. Phys.* **1996**, *104*, 286.
- [22] Norberg, J.; Nilsson, L. *J. Phys. Chem.* **1995**, *99*, 13056–13058.
- [23] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

# Chapter 6

## Conclusions

### 6.1 Protein Simulations

The calculated free energy changes for the hydration of the cavities in the two different proteins vary considerably depending on the type of cavity. the entropy change also varies considerably. For the BPTI cavity, the hydration process is entropically unfavorable and for barnase it is entropically favorable. The polar cavity in BPTI, in which the water can form four hydrogen bonds, would be predicted to be hydrated, based on the calculate free energy change of -4.7 kcal/mol. The less polar cavity of the barnase mutant is not predicted to be hydrated. This result, that polar cavities are hydrated and non-polar are not, is consistent with calculations of cavities on other proteins,<sup>1</sup> but our  $\Delta G_{hyd}$  for the barnase mutant is in apparent disagreement with the X-ray results of Buckle, *et al.*<sup>2</sup>

Before discussing the differences between the simulations and experiments, it is worth stating that there is some ambiguity in the X-ray data as well. The unit cell of the protein contains three protein molecules and only one (molecule C) contains electron density well-defined enough to indicate a water molecule in that position. The electron density in the other two structures is too weak to assign a water molecule to this site. The three protein structures in the unit cell are very similar (with root-mean-square deviations of 0.169 for protein A and 0.128 for protein B from C) so it is not obvious why only one of

the structures contains the buried water. The simulation and X-ray (for protein structure C) results may be different for several reasons. The experiments are done under different conditions, including not only the crystal environment but also at a temperature of 4°C and at a pH of 7.5.<sup>2</sup> The temperature difference does not seem significant enough to change results, and since the calculations find that the process is entropically favorable,  $\Delta G_{hyd}$  should be even less at 25°C than at 4°C. Another explanation may be that the interior water in the X-ray structure is thermodynamically unstable, in agreement with the calculated results, but is kinetically trapped. This does not seem too likely, since the cavity is not too far from the surface and our simulations reveal that the water molecule can leave the cavity (when unconstrained) on a short time scale. The differences may also be due to problems with the potential models. For the Cornell, *et al.* force field, the parameters which describe the interactions between non-polar groups and water are chosen to reproduce the solvation free energies for a set of molecules, including methane, ethane, and butane.<sup>3</sup> These solvation energies depend not only on the strength of the water-solute interaction but also the water-water potential. For example, several models with much different methane-water interactions have similar solvation free energies.<sup>4</sup> The water-methane interaction for Cornell, *et al.*-TIP3P models has a potential energy minimum equal to -0.28 kcal/mol, much less than the *ab initio* value of -0.71 kcal/mol.<sup>5</sup> Other non-polarizable models have similarly weak methane-water interactions, while a polarizable model has a deeper minimum (-0.57 kcal/mol), while still giving the same solvation free energy.<sup>4</sup> This may indicate a possible area of improvement for potential models. Stronger interactions between water and non-polar groups would tend to decrease the value of  $\Delta G_{hyd}$ , bringing it in closer agreement with the apparent X-ray result. It appears that a closer look at both the X-ray data and the potential models is necessary before the hydration of this cavity and perhaps other hydrophobic cavities is fully understood.

The dynamics of the water in the cavity depends on the type of cavity. Mobility in

the non-polar barnase cavity is high with the water molecule undergoing two types of motion not seen in the polar BPTI cavity. The molecule rotates to interchange which of its hydrogen atoms is near the Phe 7 O atom, on a 5 picosecond time scale, and also translates, about once every 14 ps, to a secondary binding site 7.3 Å away to form a single hydrogen bond, with the Try 97 O atom (Figure 2.2). For the water molecule in BPTI, no such rotations or translations to other regions of the protein are seen.

For all proteins, the addition of the water molecule to the interior cavity increases the flexibility of the protein, as seen in an increase in the mean square fluctuations in atomic positions,  $\langle \Delta r^2 \rangle$  (Figures 2.3, 2.15 and 2.21). The interior water molecules may increase the protein flexibility by increasing and weakening the length of nearby protein-protein hydrogen bonds, this trend can be seen in barnase and BPTI (Figures 2.6 and 2.16). Other studies have indicated that buried water molecules shield charge-charge interactions of the protein leading to a higher dielectric constant, this would lead to an increased protein flexibility.<sup>6</sup> Two studies of BPTI in the gas-phase have examined how protein flexibility changes, based on calculations of the change in vibrational entropy upon binding the water molecule. These calculations reached opposite conclusions, as stated in the Introduction, with one concluding that the flexibility increases<sup>7</sup> and the other that it decreases.<sup>8</sup> In the simulations of Mao, *et al.*, the water molecule did not stay in the cavity, so this may explain the difference.<sup>8</sup> Our results for aqueous BPTI agree with the results of Fischer and Verma<sup>7</sup> and also a later study by Fischer, Smith and Verma.<sup>9</sup> The binding of molecules larger than water may increase flexibility as well. A study by Tidor and Karplus of the dimerization of the protein insulin, in the gas-phase, demonstrated that the change in vibrational entropy suggests that the monomer protein gets more flexible upon binding to form the dimer. Another study showed that the binding of an inhibitor to the the rhinovirus capsid protein again increases the flexibility of the protein, as seen by an increase in  $\langle \Delta r^2 \rangle$ .<sup>10</sup> This capsid protein study is different in that the binding process involved the displacement of water, whereas the other studies involved the displacement of

empty space, so this study is not looking at quite the same thing.

In general our calculations at 1 atm found that the thermodynamics for the hydration of a protein cavity depend significantly on the properties of the cavity. The entropy of hydration,  $\Delta S_{hyd}$ , varies considerably for the two proteins, which, given the large differences in the mobility of the buried water molecules is perhaps not surprising. The  $\Delta S_{hyd}$  for the barnase cavity is positive, indicating that the process is entropically favorable. For the BPTI cavity  $\Delta S_{hyd}$  is negative. Different entropies for different hydration sites on proteins are also indicated by the mass spectroscopy data on gas-phase BPTI<sup>11</sup> and by the librational amplitudes of water in BPTI,<sup>12</sup> although all the reported  $\Delta S_{hyd}$  are negative. Those hydration sites are likely to be more hydrophilic than the barnase cavity. The entropy estimates based on the librational amplitudes find that the entropy changes are closer to zero for the water molecules that form fewer than four hydrogen bonds.<sup>12</sup> Our results show that  $\Delta S_{hyd}$  can be negative for water molecules which form only one hydrogen bond. For the BPTI cavity,  $\Delta S_{hyd}$  is  $-12 \pm 9$  cal/mol/K. The gas phase mass spectroscopy data on BPTI gives an entropy change of  $-33 \pm 5$  cal/mol/K for the transfer from the liquid phase to gas phase BPTI although the binding site of the water is uncertain.<sup>11</sup> Our calculated value is significantly larger than this but is smaller than the Dunitz lower bound estimate of  $-7$  cal/mol/K.<sup>13</sup> That value is certainly within the error bars of our calculations, but our result, together with the gas phase measurements for BPTI and the theoretical estimate for HIV-1 protease,<sup>14</sup> indicates that entropy changes may in fact be less than the “lower bound” value of Dunitz. The lower entropy implies that the addition of the buried water has a significant influence on the protein. The magnitude of the influence may be larger in the gas phase than in the liquid phase. Our results, particularly the  $\Delta S_{hyd}$  values agree with the conclusions of Takano, et al that “all water molecules do not contribute equally to stability, owing to differences in the environment of water molecules in proteins, such as the number of hydrogen bonds.”<sup>15</sup>

While superficially our calculations at 1 bar had free energies that related to cavity

environment, the properties of the proteins studied at 3 kbar do not display a consistent dependence on hydration and the change in free energy of hydration is dependent on cavity type. The barnase and lysozyme mutants with their single buried waters in hydrophobic environments, demonstrate a clear shift towards a lower  $\Delta G_{Hyd}$ . While BPTI, with an isolated water in a hydrophilic environment, has a positive shift. While the change in free energy of hydration is small for both proteins,  $<1.0$  kcal/mol, the values do appear statistically significant. The lack of significant changes in protein volume and radius of gyration is not unexpected as prior to denaturation there will be areas of localized increased volumes that offset negative volume changes.

The protein at 1 bar exhibited a positive dependence on hydration, that is the hydrated protein was more flexible than the empty protein. However, at 3 kbar the flexibility of BPTI shows little dependence on hydration, and the sections within 6 Å of the cavity water exhibit a slight negative dependence. The slight increase in cavity size upon hydration and decrease in near by hydrogen bond lengths is consistent with results at 1 bar and can not be used to explain this behavior. Most residues in the protein exhibit less motion during the simulation indicating the hydrated protein is stiffens as pressure is increased. The increase in solvent bound to the protein when the cavity was empty and the minimal changes in hydrogen bonding when the to hydrated states were observed indicate that wild-type protein is more stable to pressure changes.

Buried water in the hydrophobic barnase cavity results in no shift in the hydration dependence of flexibility, however this is not unexpected as the water is not strongly interacting. However, the water appears more stable within the cavity at high pressure, exhibiting both longer periods between flips and a smaller  $\langle r^2 \rangle$  of the water oxygen. This apparent stability should decrease the entropic contribution to the  $\Delta G_{hyd}$ .

Perhaps most interesting was the effect of pressure on lysozyme, where despite the lack of any stable hydrogen bonds during the course of the simulation the  $\Delta G_{hyd}$  was over 1 kcal/mol more favorable. While the resulting free energy still did not indicate a



spontaneous process it does give credence to the idea that hydrogen bonding alone can not be used to predict extent of hydration.

Examination of the  $\langle \delta r^2 \rangle$  data for both proteins on a residue-by-residue basis and by structure type failed to generate any patterns to the shift other than the structure changes that occur at turns, bends, and residues next to, but not part of, major structure types. This agrees with the concept that the areas that are altered first under increased pressure are those that do not have a rigid hydrogen bond network. The role of buried water in the denaturation process is still difficult to discern and our flexibility results indicate that it may depend greatly on the environment in which the water is located. However, the concept that the cavity water serves to nucleate the process is unlikely, at least the three proteins we examined. There was no real hydration dependence to flexibility and no changes in hydrogen bonding occurring near the water when the pressure is increased to 3kbar.

# Bibliography

- [1] Zhang, L.; Hermans, J. *Proteins: Struct. Funct. Gen.* **1996**, *24*, 433–438.
- [2] Buckle, A. M.; Cramer, P.; Fersht, A. R. *Biochemistry* **1996**, *35*, 4298–4305.
- [3] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [4] Rick, S. W.; Berne, B. J. *J. Phys. Chem. B* **1997**, *101*, 10488.
- [5] Szczęśniak, M. M.; Chałasiński, G.; Cybulski, S. M.; Cieplak, P. *J. Chem. Phys.* **1993**, *98*, 3078.
- [6] Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S. *Biophys. J.* **2000**, *79*, 1610–1620.
- [7] Fischer, S.; Verma, C. S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9613–9615.
- [8] Mao, Y.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 2950–2951.
- [9] Smith, S.; Smith, J. C.; Verma, C. S. *J. Phys. Chem. B* **2001**, *105*, 8050–8055.
- [10] Phelps, D. K.; Rossky, P. J.; Post, C. B. *J. Mol. Biol.* **1998**, *276*, 331–337.
- [11] Woenckhaus, J.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 9586–9587.

- [12] Denisov, V. P.; Venu, K.; Peters, J.; Hörlein, H. D.; Halle, B. *J. Phys. Chem. B* **1997**, *101*, 9380–9389.
- [13] Dunitz, J. D. *Science* **1994**, *264*, 670.
- [14] Li, Z.; Lazardis, T. *J. Am. Chem. Soc.* **2003**, *125*, 6636–6637.
- [15] Takano, K.; Funahashi, J.; Yamagata, Y.; Fujii, S.; Yutani, K. *J. Mol. Biol.* **1997**, *274*, 132–142.

## Vita

Lisa Renee Olano was born in Silver Spring, Maryland. She received her B.S. degree from Emory University where she did research for David Goldsmith in natural product synthesis. She received a M.Ph. in Chemistry from George Washington University where her research involved organophosphonate chromium complexes under the direction of Dr. D. Andrew Knight. She switched to the University of New Orleans in 2001, where she joined Steven Rick's research group.