Summer 8-4-2011

# Water in Protein Cavities: Free Energy, Entropy, Enthalpy, and its Influences on Protein Structure and Flexibility

Hongtao Yu
*University of New Orleans*, htaoyu1@gmail.com

# Water in Protein Cavities: Free Energy, Entropy, Enthalpy, and Its Influences on Protein Structure and Flexibility

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirments for the degree of

Doctor of Philosophy
in
Chemistry

by

Hongtao Yu

B.S. University of Science and Technology of China, 2005

August, 2011

*In the loving memory of my father*
*Congquan Yu (1957–1998)*
*who inspired my interest in physics*

## Acknowledgments

First and foremost, I would like to express my respect for Professor Steven W. Rick who instilled in me the qualities of a good scientist. It was a great challenge for me to transfer from physics to chemistry. As my doctoral advisor, Professor Rick allowed me ample time to learn the fundamentals of chemistry and the programming techniques necessary for Molecular Dynamics simulation. I appreciate all his valuable advice, patient assistance and financial support throughout my graduate study.

I would also like to acknowledge my committee members, David Mobley, Edwin Stevens, John Wiley, and Yang Cai, for providing their time and great comments on this dissertation. I would like to especially thank David Mobley for teaching me computational chemistry and introducing me to powerful and sophisticated software packages and tools. Special thanks are given to the entire chemistry faculty of the University of New Orleans for their instruction and guidance, and to Alexis Lee for her help in the past years and for reviewing the manuscript of this dissertation.

It would have been impossible for me to finish my studies without the love and patience of my family. My parents, my grandma, and my sister have provided a constant and cherished source of love, concern, support and strength in my life. I am also grateful to my aunts and uncles for their supporting and loving encouragement.

This document was prepared using LaTeX with 12 pt Kerkis font. Protein structure figures were generated using PyMol. The chemical structure figure was made with MarvinSketch, and data plots were produced with matplotlib library of Python.

# Table of Contents

# List of Figures

# List of Tables

## Abstract

Complexes of the antibiotics novobiocin and clorobiocin with DNA gyrase are illustrative of the importance of bound water to binding thermodynamics. Mutants resistant to novobiocin as well as those with a decreased affinity for novobiocin over clorobiocin both involve a less favorable entropy of binding, which more than compensates for a more favorable enthalpy, and additional water molecules at the protein-ligand interface. Free energy, enthalpy, and entropy for these water molecules were calculated by thermodynamic integration computer simulations. The calculations show that addition of the water molecules is entropically unfavorable, with values that are comparable to the measured entropy differences. The free energies and entropies correlate with the change in the number of hydrogen bonds due to the addition of water molecules.

To examine the wide variety of cavities available to water molecules inside proteins, a model of the protein cavities is developed with the local environment treated at atomic detail and the nonlocal environment treated approximately. The cavities are then changed to vary in size and in the number of hydrogen bonds available to a water molecule inside the cavity. The free energy, entropy, and enthalpy change for the transfer of a water molecule to the cavity from the bulk liquid is calculated from thermodynamic integration. The results of the model are close to those of similar cavities calculated using the full protein and solvent environment. As the number of hydrogen bonds resulting from the addition of the water molecule increases, the free energy decreases, as the enthalpic gain of making a hydrogen bond outweighs the entropic cost. Changing the volume of the cavity has a smaller effect on the thermodynamics. Once the hydrogen bond contribution is taken into account, the volume dependence on free energy, entropy, and enthalpy is small and roughly the same for a hydrophobic cavity as a hydrophilic cavity.

The influences of bound water on protein structure and influences are also evaluated by performing molecular dynamics simulation for proteins with and without bound water. Four proteins are simulated, the wild-type bovine pancreatic trypsin inhibitor (BPTI), the wild-type hen egg white lysozyme (HEWL), and two variants of the wild-type Staphylococcal nuclease (SNase), PHS and PHS/V66E. The simulation reveals that all these four proteins suffer structural changes upon the removing of bound water molecules, as indicating by their increased RMSD values with respect to the crystal structures. Three out of the four proteins, BPTI, HEWL, and the PHS mutant of SNase have increased flexibility, while no apparent flexibility change is seen in the PHS/V66E variant of SNase.



*Keywords*: Molecular Dynamics, Water, Protein-Ligand Binding, Free Energy Calculation, Entropy, Enthalpy, Protein Structure, Protein Flexibility

# Chapter 1

# Introduction

## 1.1 Protein Folding and Cavities in Protein Interiors

Proteins are a class of biological macromolecules of major importance, whose functions involve catalyses of biological reactions, regulations of protein functions and gene expressions, transports of substances, storage of nutrients. All proteins are composed of one or more linear polypeptide chains built up from 20 amino acids. The amino acids in polypeptide chains are linked through covalent peptide bonds and their linear sequences are usually referred to as proteins' primary structures. Each protein has an unique amino acid sequence that is specified by the sequence of a gene [1, 2]. Segments of polypeptides often fold locally into ordered spatial arrangements such as *a helices* and *β pleated sheets*. The *a helices* are coiled structures stabilized by intra-chain hydrogen bonds, while *β sheets* are side-by-side arrays of polypeptide strands that are stabilized by hydrogen bonding between adjacent strands. These stable regular local structures are often called the secondary structures of proteins. Single polypeptide chains may bend and fold into irregular but nevertheless well-defined structures, these structures are usually stabilized by disulfide bonds, ionic bonds as well as weak interactions (such as hydrogen bonding and van der Waals interactions) and are defined as the tertiary structures of proteins. Most proteins consist of more than one polypeptide chains of characteristic tertiary structures, which are commonly referred to as subunits of proteins. The ways in which

these interacting subunits associate to form oligomeric proteins constitute the quaternary structures of proteins. Subunits that make up a oligomeric protein may be identical or different. The secondary, as well as tertiary and quaternary structures of proteins are intimately dependent upon their primary structures. Among them, the secondary structures are specified by the short-range sequences of proteins, while the tertiary structures are governed by their long-range sequences. The overall three dimensional architectures of proteins are generally termed as their conformations, by which the functions of proteins are completely determined.

Generally, proteins can be divided roughly into three global classes in terms of their overall morphologies and physicochemical properties: fibrous, globular, and membrane. Fibrous proteins are usually water insoluble proteins with relatively simple rod-like or wire-like structures, they play mainly structural roles in living systems for mechanical supports and protections. In contrast, globular proteins are mostly soluble proteins in aqueous solutions with roughly spherical conformations, and they perform primarily "chemical" roles such as binding, catalyses and switching. Membrane proteins are protein molecules that are noncovalently associated with various membrane systems of cells and organelles, they contain higher proportions of hydrophobic amino acids compared with globular proteins, and usually fold in such conformations that the apolar amino acids are oriented into their membrane-associated regions while the polar amino acids expose at the aqueous environments. Globular proteins are by far the most prevalent proteins among these three categories due to their numerous amounts and functions [3].

The globular proteins are usually characterized by their hydrophobic cores buried inside and hydrophilic surfaces exposed to and interacting with the water environments, the globular shapes adopted by them lower their surface-to-volume ratios and hence minimize their interactions with the surrounding medium. The well defined structure of a globular protein is usually referred to as its native state, which is formed by a so called folding process, and it is the most stable conformation for that protein in a given solvent

at a particular temperature. The folding process of a protein happens in a so highly cooperative manner that only the fully folded (native) state and the completely unfolded (denatured) state are present stably, while the intermediate partly folded structure that might interfere with processes within cells exists only transiently. It was believed that protein folding is a *cumulative selection* process that is largely driven by the hydrophobic effect. The folding pathway of a globular protein in aqueous solution starts with the formation and retention of the correct local conformations such as *α helices* and *β sheets*, the retained structures then undergo a hydrophobic collapse by aggregating the apolar groups inside the protein and meanwhile expelling the majority of water molecules from the interior, form a more compact global structure. Actually, the folded proteins are so well packed that their packing density is similar to that found in molecular crystals of amino acids [4, 5, 6, 7, 8, 9].

The tight packing of atoms in natively folded proteins, especially in their core regions, are critical for structural specificity and stability [4, 7, 10]. Considering the spontaneous aggregation of oil in water, this process is largely driven by the hydrophobic effect. However, the lack of specific architecture in the droplet interior implies that the hydrophobic interaction can not be regarded as the only factor governing protein folding; specific packing interaction within protein interiors is believed to play a major role in protein conformation and stability [4]. Numerous studies have shown that the protein interiors are closely packed, with an average packing density of about 0.75 [4, 7, 11, 12]. In a folded protein, the complementary side chains of amino acids fit tightly with each other, giving an unique folding structure with striking packing efficiency.

Although the protein interiors are tightly packed, the packing is not uniform; packing defects exist ubiquitously in proteins in the form of interior cavities. Indeed, internal cavities of atomic size have been found by X-ray crystallography [13, 14, 15, 16, 17, 18] and by theoretical methods [19, 20, 21]. Hubbard et al. performed a comprehensive analysis on internal cavities in 121 globular protein crystal structures and pointed out that,

internal cavities are not artificial packing defects created in structure determination, they present commonly in proteins of all size, particularly in proteins containing more than 100 residues, and are most happened in core regions. The sizes of cavities in protein interiors spread over a large range, sometimes can be as large as 200 $Å^3$. The total volume of cavities accounts for approximately two percent of the total protein volume [20, 22].

Extensive site-directed mutagenesis experiments have identified that the internal cavities may have influences on both the biological functions [23, 24] and the structural stabilities [25, 26, 27, 28] of proteins. For example, Lambright et al. have suggested that the rebinding kinetics of CO in myoglobin may be affected by the internal cavities [23], and Lee et al. proposed that the inhibitory activity of $a_1$-antitrypsin correlates positively with the cavity volume [24]. The appearance of cavities in protein interiors may destabilize their folded structures due to the loss of van der Waals contacts [28, 29, 30], notwithstanding in some cases the cavities do increase proteins' structural stabilities by eliminating unfavorable strains [25]. Eriksson et al. created a number of "cavity-creating" mutations in T4 lysozyme by substituting large hydrophobic residues with small ones (Leu → Ala) and noted that the loss of protein stability is approximately linearly related to the volume of the cavity [28]. A similar linear relationship between the protein stability changes and the cavity size was also obtained by Buckle et al. when studying the barnase Ile/Leu → Ala mutants [31].

## 1.2 Cavities at Protein Binding Interfaces

Many of the most important biological processes in living organisms involve the association of proteins to other proteins, DNAs/RNAs, or small ligands. For examples, in living cells, the replication of DNAs are carried out by large protein complexes that consist of a large number of protein components. Furthermore, the regulation of gene expression and translation requires the binding of proteins to either double-strand or single-strand DNAs/RNAs. It has long been recognized that protein binding, which represents inter-

molecular recognition, shares many similar features with folding, which represents intramolecular recognition [32, 33, 34, 35, 36]. For example, both the protein binding and folding processes involve the repelling of water molecules and the assembling of complementary residue groups. Although the surfaces of binding subunits have a high degree of geometry match [37] and hydrophobic complementarity [38, 39], they are far from perfect packing. Lawrence and Colman analyzed the shape complementarity at surfaces between oligomeric proteins and found that the protein interfaces are not optimally matched [40]. This conclusion was further confirmed by Hubbard and Argos on an analysis of a larger set of data [41]. Interfacial cavities may be present due to the imperfection in surface complementarity between binding subunits. Several statistical analyses have demonstrated that cavities occur frequently at the protein-protein interfaces [41, 42], and protein-DNA/RNA interfaces [43].

The cavities present at protein-protein interfaces share some common properties with those in protein interiors, such as wide distributions in cavity volume and hydrophobicity and rough relationship between total cavity volume and protein size. However, the interfacial cavities have their unique characteristics. They happen more frequently than interior cavities, their volumes spread a broader range and on average are larger than internal cavity volumes [41, 42]. The cavities at the protein-DNA/RNA interfaces were characterized by Sonavane and Chakrabarti and no significant differences are found between them and protein-protein interfacial cavities [43]. The presence of interfacial cavities in protein complexes may play some functional roles, though they sometimes have harmful influences on stability and binding affinity. Hubbard and Argos, for example, have suggested that the cavities at protein domain interfaces may involve shear or hinge domain-domain motions [44], which are critical for some proteins to perform their functions.

## 1.3 Water in Protein Cavities and Its Biological Functions

Nature abhors a vacuum. As long as the empty sites not occupied by protein atoms are large enough to accommodate water molecule(s), the solvent waters will try to fill them. Crystal structure analysis has shown that cavities that large enough to hold at least one water molecules are commonly found in protein interiors and contributed 1% to the total protein volume [45]. Buried water molecules have been observed in a number of experiments [31, 46, 47, 48, 49, 50, 51, 52, 53] and occupy 18% of these cavity sites, giving on average 1 buried water for every 27 amino acid residues. The hydrated water molecules prefer to reside in larger and polar cavities, almost all cavities with polarity (the ratio of polar to apolar solvent accessible surface) above 0.55 and more than 92% cavities that exceed 50 $Å^3$ in solvent accessible volume are hydrated. Each hydrated water molecule, on average, makes 3 polar contacts with the protein and other buried waters [22, 45]. The hydration waters at the protein binding interfaces are analyzed by Lu et al. in 2007. Their study involves 392 protein-ligand crystal structures with resolutions $\leq 2$ Å. Over 85% of these protein-ligand structures are found have at least 1 bound water molecules and totally 1839 bound waters are found in these 392 complexes, 72% of which are interfacial waters that take part in the mediation of protein-ligand interaction. Each interfacial water molecule averagely has 3 polar contacts

One of the reasons for experimental observations of water molecules reside frequently in hydrophilic cavities is because the detection of water molecules is highly environment dependent. Detection by conventional X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy methods requires the hydration water molecules either have low temperature factors or long residence times (see Section 1.4). In polar cavities, the hydrogen bonds formed with the cavity residues immobilize the translational and rotational degree of freedom of the hydration water molecules, giving well defined water structures. These buried water molecules can be easily determined in the X-ray

and NMR experiments. The locating of hydration water molecules in hydrophobic cavities is hampered by their disorder. Interleukin-1$\beta$ (IL1B), for example, is a cytokine protein that involved in wide-ranging cellular activities such as inflammatory response and hematopoiesis. Structure analysis reveals five cavities in this protein interior, among which four are hydrophilic cavities located around the trefoil structure (labels 1, 2, 3, 4 in Figure 1.1). Hydration water molecules in these polar cavities have been well identified by both crystallography [55, 56, 57, 58, 59] and NMR [60, 61] methods. These water molecules make multiple hydrogen bonds with the protein and involve in bridging the backbone interactions. The remaining one cavity is a large hydrophobic cavity located in the center of the protein (label 5 in Figure 1.1). This cavity can hold one to four water molecules and is completely lined with nonpolar residues. The hydration state of this cavity has caused much controversy. Several X-ray studies reveals no water molecules in this cavity [55, 56, 57, 58, 59]. In contrary, a NMR examination conducted by Ernst et al. demonstrates that there are positionally disordered water molecules. These water molecules have residence times ranging from 1 ns to 200 $\mu$s, although they are too disordered to be determined by standard crystallography technique, they are detectable in some NMR experiments [48]. A theoretically based crystallography analysis by Yu et al. also suggests a hydrated water dimer within this cavity [52]. It should be noted, however, that the Yu et al. method is based on an iterative modification procedure of the solvent density, the validity of this modification is not well justified. The hydration of the central hydrophobic cavity is also supported by a 2007 molecular dynamics study, in which water molecules with residence times exceeding 500 ps are observed. The number of hydration water molecules vary dynamically between one to four [62]. Although there appear some evidences about the hydration of this large hydrophobic cavity, a recent free energy calculation shows unfavorable (positive) free energy changes for transferring water molecules from the bulk into the cavity, the free energy cost increases monotonically with the number of water molecules in the cavity [63]. The significant disagreements between

these experiments and the simulation as well as the free energy calculation suggest that the hydration pattern of hydrophobic cavities is still unclear and further work is needed.

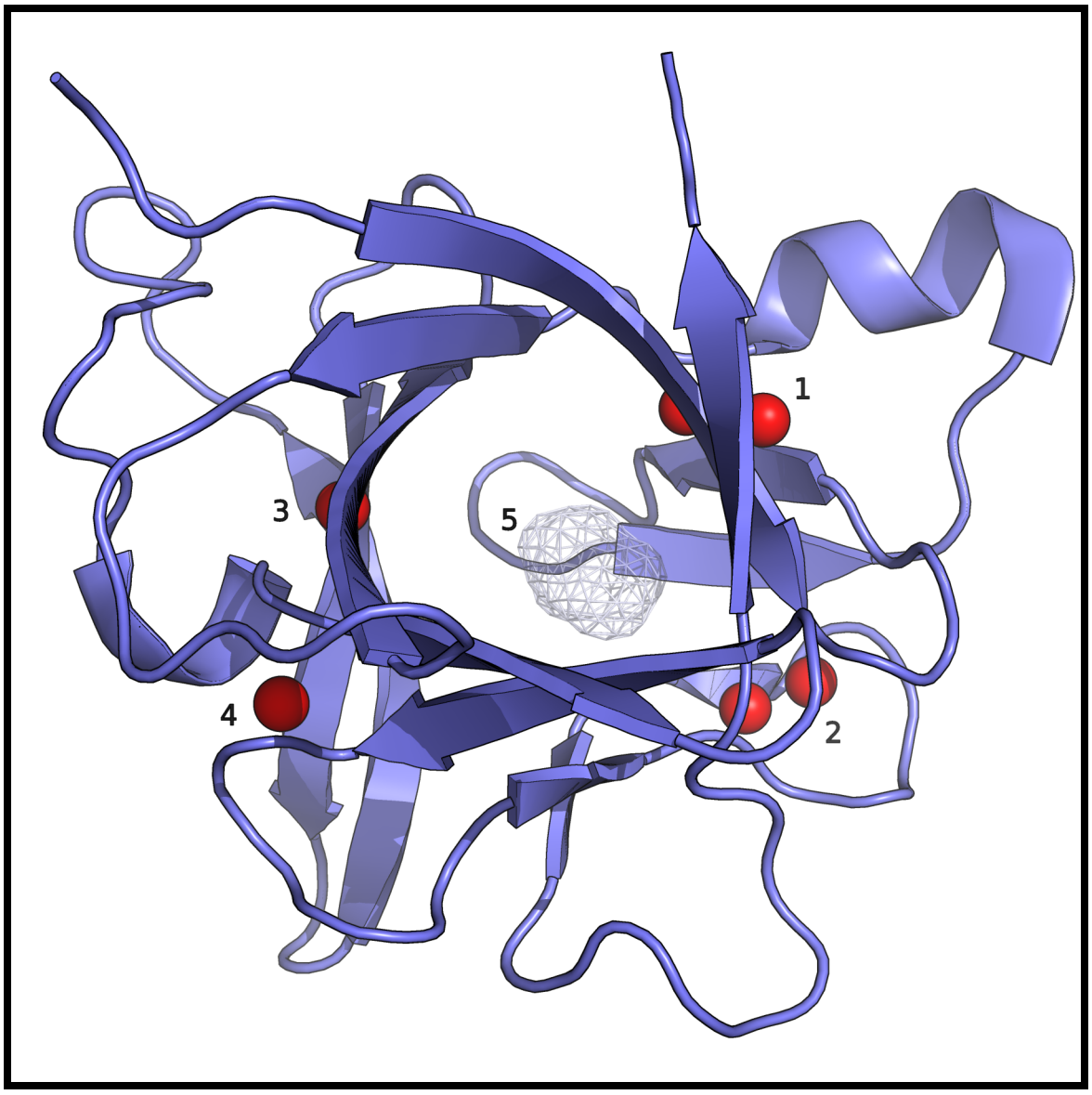


Figure 1.1: 2NVH crystal structure of Interleukin-1$\beta$ [59]. Five interior cavities are found in this protein, of which four polar cavities are hydrated by water molecules (red spheres) and one apolar cavity located in the center of the protein is observed to be empty (wireframe at label 5)

The occupation of water molecules in protein cavity can promote the conformational stability by making favorable interactions with the protein. First, the buried water molecules can increase the van der Waals contacts without introducing any new internal

strains [22, 31, 64]. Second, the unique structure allows water can act as both hydrogen bond donor and accepter. Although the protein core consists of predominantly hydrophobic residues, polar atoms are unavoidable due to the structural property of peptide bond; solvated water(s) can stabilize protein by linking otherwise separated polar groups in backbones and side chains via hydrogen bonding or hydrogen bonding networks [65, 66]. The influences of buried water on protein stability can be studied experimentally by introducing them into or repelling them out from the protein interior through structural mutations.

A well studied example about water in protein cavity is bovine pancreatic trypsin inhibitor (BPTI). BPTI is a small serine protease inhibitor consisting of 58 amino acids in a single polypeptide chain. It exists in blood and many tissues. The function of BPTI involves the inhibition of several proteolytic enzymes such as trypsin, chymotrypsin, kallikrein and plasmin, etc. Four water molecules are found to be buried completely in the protein interior, and conserve in several wild type BPTI structures [67, 68, 69, 70, 71], even in an extensively mutated form, which contains 20 alanines [72]. These four conserved water molecules, which are usually considered as an integral part of the protein, located in two cavities and are completely inaccessible to the bulk solvent. One isolated water (WAT122 in 5PTI structure, see Figure 1.2) is harbored in a small cavity which is composed of two loops containing residues 11–14 and 36–38; the remaining three waters (WAT111, WAT112, and WAT113 in 5PTI) form a cluster and reside in a channel-like cavity created by residues 8–10 and 40–44 (). The most deeply buried water molecule in the cluster, WAT113, is about 7 Å away from the isolated WAT122 water. The four structural water molecules form totally 9 water-protein and 2 water-water hydrogen bonds, with 4 water-protein hydrogen bonds are contributed by WAT122 (Figure 1.2) [67, 68, 69]. A mutagenesis experiment has shown that, the mutation of a glycine residue, Gly-36, to a serine residue to replace the WAT122 water molecule with a hydroxyl group will break the hydrogen bond between the water hydrogen and the carbonyl oxygen in cysteine 38, since the hydroxyl moiety

can donate only one proton to potential hydrogen bond accepters. This interrupting of hydrogen bond slightly destabilizes the BPTI mutant by approximately 0.7 kcal/mol [65].

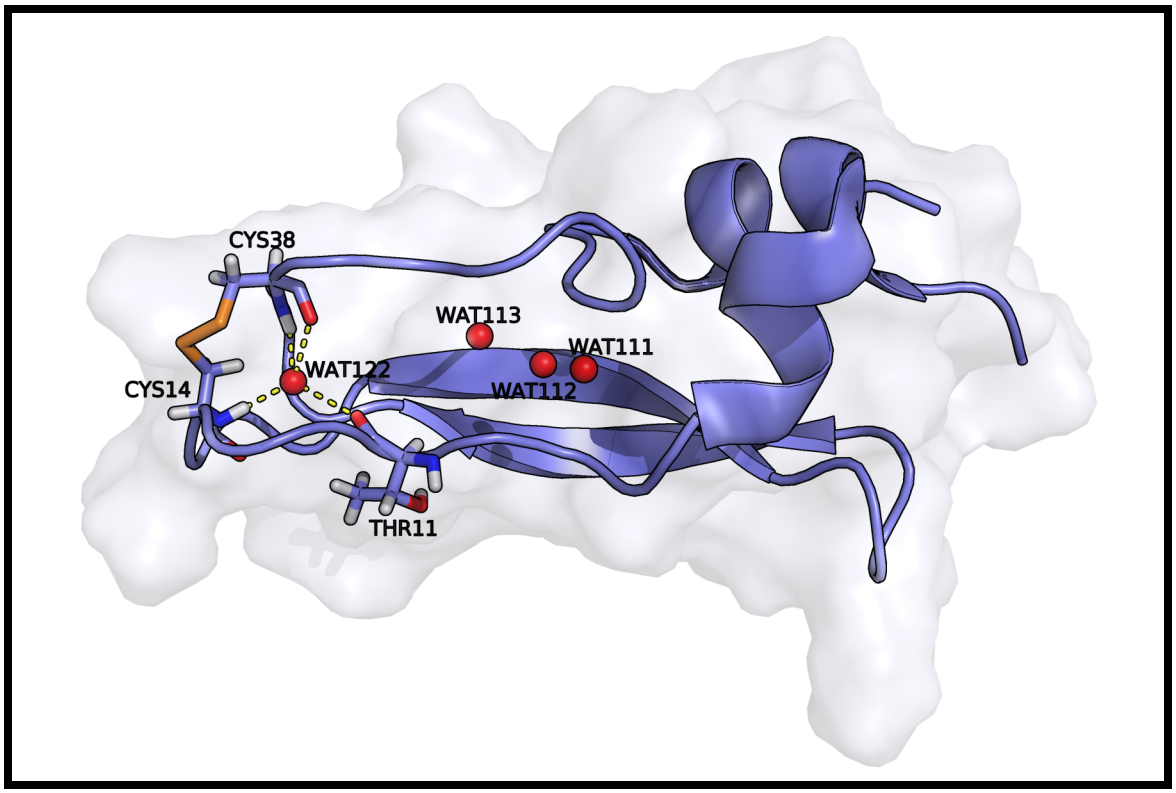


Figure 1.2: 5PTI crystal structure of BPTI [68]. 4 buried water molecules are observed in 5PTI structure (red spheres). 1 water molecule locates in an isolated cavity and the remaining 3 water molecules reside in a larger cavity forming a water cluster.

Beside their stabilizing function, buried water molecules may also have some influence on the protein flexibility. The structural flexibility has been proved to have strong influences on many proteins' functions. Several studies have shown that the protein function can be altered by mutations which do not affect the structure but do change the flexibility [73, 74, 75]. One dramatic example is demonstrated in the work of Adams et al.. Single point mutations in the protein Cdc42Hs, a member of the Ras superfamily of proteins, can be oncogenic, not due to structural differences but solely through changes in flexibility [74]. In another recent study, mutations in the C-terminal Src kinase, Csk, disrupt the function of that enzyme by decreasing its flexibility [75]. The hydration of protein interior changes the interactions among residues and thus affects its flexibility.

Intuitively, the water occupancy of empty sites in protein will result in a tighter atomic packing and hence a more rigid structure. Mao et al. has shown that the burial of a water molecule apparently decreases the protein's flexibility [76]. In contrast, the work of Fischer, Verma and co-workers, also on the same protein (BPTI) and the same water molecule (WAT122 in 5PTI, see Figure 1.2), found that the protein flexibility is increased upon the addition of that water [77, 78, 79]. The promotion of protein flexibility by buried waters are also observed by Olano and Rick in the study of BPTI and I76A mutant of barnase. They found that both protein's flexibilities are increased by introducing a water molecule into the protein interiors [80]. How the water molecules influence the flexibility of the protein is still not fully established.

In addition to their structural role, conserved water molecules have been found in proteins belonging to the same homologous family [81, 82], implying that they may serve more of functional roles [83, 84, 85]. For example, a X-ray study indicated that the releasing of internal bound water molecules from the active sites of serine proteases plays a primary role in the substrate binding [86]. A more recent study conducted by Tashiro and Stuchebrukhov suggested that the presence of the structural water molecules in the internal catalytic center of Cytochrome *c* Oxidase plays a dual role in the oxygen reduction process: (1) proton transfer and, (2) protein conduction [85].

Water molecules appear in the cavities of protein interfaces are considered to mediate the protein-protein and protein-ligand interactions. These water molecules usually have significant influences on protein binding thermodynamics. Large changes in binding affinities between the proteins and ligands can be attributed to the presence of one or more tightly bound water molecules at the binding sites [87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97]. These water molecules have attracted considerable attention due to their potential applications in protein structure prediction and pharmaceutical design. For example, in structure prediction, an improved prediction has been archieved by "wetting" the Hamiltonian with a knowledge-based potential to include the water mediated long-

range interactions between polar or charged groups [98, 99]. The introduction of ordered water molecules at the binding interfaces is generally considered to decrease the binding affinities due to the significant entropic costs. This decrease sometimes can exceed over an order of magnitude [92, 100]. This rule has been used as a strategy to design protein inhibitors with greater binding affinities by replacing water molecules from the binding interfaces [87]. However, there are exceptions that in some cases the binding affinities are improved by introducing water molecules into the binding sites [101]. Although the water molecules seem play import roles in protein binding, their thermodynamic properties are still poorly understood.

## 1.4 Methods to Study Water in Protein Cavities

Experimentally, various techniques could be used to investigate the presence of buried water molecules in protein interiors, *e.g.*, X-ray crystallography [31, 49, 52], neutron diffraction [47, 53], NMR spectroscopy [46, 48, 50], and mass spectrometry [51]. X-ray and neutron diffraction determine the water molecules by detecting the electron densities of water atoms in protein crystals. Since the electron densities observed in the experiment represent a linear superposition of all possible positions during the data collection process, which typically lasts several hours, only well ordered water molecules with thermal fluctuations smaller than 1 $Å^2$ (the upper limit for resolving electron densities ) could be determined [102]. If the positional fluctuations larger than this limitation, the resulting electron density will be smeared and, as a result, the water molecules are undetectable. Such disordered water molecules can be detected in solution NMR spectroscopy. NMR does not require the uniform ordering of water molecules but relies primarily on the intermolecular nuclear Overhauser effect (NOE) between water protons and hydrogen atoms of the proteins [46]. Since the NOE intensity depends on the inverse sixth power of the proton-proton distance, a spatial proximity of protons between water molecules and

the proteins are required. In addition, a residence time of 500 ps or longer is necessary for the water molecules in order to generate a detectable signal [46, 60, 103].

The thermodynamics of buried water molecules in protein interior cavities can be measured indirectly by investigating their influences on protein thermal stability and protein-ligand interaction. In a typical experiment, the protein (or ligand) is altered to introduce or eliminate buried water molecules, The thermodynamics of corresponding buried water molecules are then interpreted from the difference in thermodynamics of the mutated and native structures. Several methods could be used to study the protein, protein-ligand binding thermodynamics, such as Circular Dichroism(CD) [28, 65], Differential Scanning Calorimetry (DSC) [88], Isothermal Titration Calorimetry (ITC) [64, 92, 95, 96], *etc.*. CD spectroscopy monitors the absorption of circularly polarized light. Protein secondary structures, such as *α helices* and *β sheets* are chiral structures, and thus have different absorption properties for the left- and right-handed circularly polarized light. The protein unfolding thermodynamics can be measured by monitoring the change in the absorption for the two types of circularly polarized as a function of denaturant concentration or temperature. DSC is an alternative technique widely used to study the thermal denaturation of protein. DSC is designed to detect the amount of heat required to raise the temperature of a protein system. By comparing with the reference system, the heat capacity changes associated with protein thermal denaturation as a function of temperature are obtained, and thus, free energy, entropy, and enthalpy. ITC is particularly suitable for studying the protein-ligand binding interaction. In a basic ITC experiment, the ligand is injected gradually into the solution of protein, the heat absorbed or released by the protein-ligand binding reaction is measured at a constant temperature. Both equilibrium association constant, $K_a$, and enthalpy change, $\Delta H$, can be obtained from a single ITC measurement.

Although spectroscopy and calorimetry techniques provide practical approaches to examine the influences of buried water molecules on protein and protein-ligand thermody-

namics, these methods are invariably rely upon the mutation of specific chemical groups. The consequential changes of the structural conformations and interaction patterns of the protein complexes induced by the mutated groups hamper the decomposition of the water contribution from other factors. Computer simulation proved to be an efficient tool to study the hydration of protein cavities without introducing too much perturbation to the systems. With computational experiments, the hydration processes could be analyzed at an exceedingly spatial and temporal detail; Both accurate thermodynamic and kinetic properties, such as hydration free energies, entropies, and residence times of hydration water molecules, can be extracted from the simulations. In addition, the simulations are not limited by the physical availability of the protein systems, artificial systems could be made and the hydration can be investigated in a systematic way. Moreover, hydration properties under conditions that are difficult or impossible in the laboratory (for example, extreme temperature or pressure) can also be explored by computer simulations.

## 1.5 Molecular Dynamics and Free Energy Calculation

One of the principal methods in the computational study of the water in protein cavities is the Molecular Dynamics (MD) simulation. Since first introduced by Alder and Wainwright in the late 1950's [104, 105]. Molecular Dynamics has become a standard computer simulation method [106, 107, 108, 109]. In classical Molecular Dynamics simulations of biomolecular systems, each atom in the system is treated as a point with mass and charge, interactions between atoms are described by simple force rules (Force Field). The trajectory of the system (*i.e.*, the coordinate and velocity of each atom as a function of time) are obtained by integrating Newton's second law of motion

$$-\nabla_i U(\mathbf{r}^N) = m_i \frac{d^2\mathbf{r}_i}{dt^2} \qquad i = 1, \cdots, N \tag{1.1}$$

where $m_i$ and $\mathbf{r}_i$ represent the mass and coordinate of atom $i$ and $U(\mathbf{r}^N)$ is the potential energy of the system that depends on the configuration of the $N$ atoms. The analytical

form of $U(\mathbf{r}^N)$ and the parameters therein are usually referred to as the force filed, which are usually obtained by fitting to either *ab initio* calculations or experimental data. A commonly used empirical force field formulation for the biomolecules has the form:

$$U(\mathbf{r}^N) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} k_\vartheta (\vartheta - \vartheta_0)^2 + \sum_{torsions} \frac{1}{2} V_n (1 + \cos(n\omega - \gamma)) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} + \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.2}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between atom $i$ and $j$. The first two terms in Equation 1.2 describe the bond stretching and angle bending energies. The $k_b$, $k_\vartheta$ and $l_0$, $\vartheta_0$, are their corresponding force and equilibrium constants. The third term is the torsion energy. $\omega$, $V_n$, $n$ and $\gamma$ are the dihedral angle, barrier height, multiplicity and phase shift, respectively. The last two terms represent nonbonded van der Waals and electrostatic interactions. $\varepsilon_{ij}$, $\sigma_{ij}$, $q_i$, and $\epsilon_0$ are the dispersion well depth, Lennard-Jones diameter, atomic charge, and the vacuum permittivity, respectively. The reliability of the simulation depends intimately on the force field. With a well defined force field, the kinetic informations can be accessed directly from the simulation trajectory at a high degree of accuracy; the thermodynamic properties could be estimated with statistical mechanics methods.

One of the most important quantities in equilibrium thermodynamics study is the free energy.[110, 111, 112] In biomolecular systems, The free energies describe the association and reaction abilities of biomolecules, and are closely related to their equilibrium constants through $\Delta G = -RT \ln K$. This quantity can be estimated from the simulation via equation [113]

$$G = -k_B T \ln Q = k_B T \ln \left\langle \exp \frac{H(\mathbf{p}^N, \mathbf{r}^N)}{k_B T} \right\rangle \tag{1.3}$$

where $Q = \iint d\mathbf{p}^N \, d\mathbf{r}^N e^{H(\mathbf{p}^N, \mathbf{r}^N)/k_B T}$ is the partition function and $\langle\rangle$ denotes an ensemble average. It should be noted that an accurate estimation of free energy requires an adequate sampling in the phase space, which is hampered by the currently available computational power and rugged energy surfaces of bimolecules. For example, the term

$\exp(H(\mathbf{p}^N, \mathbf{r}^N)/k_BT)$ in Equation 1.3 implies that high energy regions contributed exponentially to the free energy, while the conventional Molecular Dynamics methods perfer to explore the low energy regions. For a systems with large energy barriers, the simulation can be trapped in a local energy minimum that close to the initial conformation during the whole simulation process, incorrect estimation may be obtained due to the pathological sampling. Although considerable number of approaches have been proposed to accelerate the sampling[114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125], an efficient algorithm for absolute free energy calculation is still in high demand.

An alternative approach is to calculate a free energy difference. Consider two well-defined states, 0 and 1. For the hydration of a protein cavity, 0 could be the state that cavity is empty, and 1 could be the state that the cavity is hydrated by a water molecule. The free energy difference, $\Delta G$, between the two states is given by

$$\Delta G = G_1 - G_0 = -k_BT \ln \frac{Q_1}{Q_0} = -k_BT \ln \langle \exp [(H_1 - H_0)/k_BT] \rangle_0 \tag{1.4}$$

The subscript 0 indicates the average is taken over the equilibrium ensemble of state 0. This method was first proposed by Zwanzig in 1954 and is sometimes called Free Energy Perturbation (FEP) [126]. Although FEP proved to be a effective method to calculate the free energy difference, it suffers the drawback that the phase space of the two states must be well overlapped to give an accurate estimation of the free energy difference. A clever strategy is to introduce some "intermediate" state between 0 and 1, the free energy difference is calculated consecutively between adjacent states and $\Delta G$ is given by integrating all these free energy differences [127]

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{1.5}$$

This method is usually referred to as the Thermodynamic Integration method (TI) in the literature.

## 1.6 Overview of the Present Study

The main focus of this dissertation is to understand the thermodynamics of water molecules in protein cavities, as well as their influences on protein-ligand binding and protein structure and flexibility. In Chapter 2, the free energy, entropy, and enthalpy of water molecules at the DNA gyrase/novobiocin interface are calculated, their contributions to the gyrase-novibiocin binding affinity are discussed. In Chapter 3, the study goes beyond specific protein systems and is extended to general case. A model of protein cavities that can accommodate only single water molecules is developed. The hydration thermodynamics of water molecules in various protein cavities are presented in terms of the cavity sizes and hydrophobicities. In Chapter 4, Molecular Dynamics simulations are performed for four proteins with and without bound water molecules in the protein interiors. The influences of the bound water on protein structure and stability are investigated.

## Bibliography


[1] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[2] H. Robert Horton, Laurence A. Moran, K. Gray Scrimgeour, Marc D. Perry, and J. David Rawn. *Principles of Biochemistry*. Prentice Hall, New Jersey, fourth edition, 2006.

[3] Gregory A. Petsko and Dagmar Ringe. *Protein Structure and Function*. New Science Press, London, 2003.

[4] Frederic M. Richards. The interpretation of protein structures: Total volume, group volume distributions and packing density. *J. Mol. Biol.*, 82:1–14, 1974.

[5] Cyrus Chothia. Structural invariants in protein folding. *Nature*, 254:304–308, 1975.

[6] J. L. Finney. Volume occupation, environment and accessibility in proteins. the problem of the protein surface. *J. Mol. Biol.*, 96:721–732, 1975.

[7] Frederic M. Richards. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.*, 6:151–176, 1977.

[8] Georg E. Schulz and R. Heiner Schirmer. *Principles of Protein Structure*. Springer Advanced Texts in Chemistry. Springer-Verlag, New York, 1979.

[9] Yehouda Harpaz, Mark Gerstein, and Cyrus Chothia. Volume changes on protein folding. *Structure*, 2:641–649, 1994.

[10] George D. Rose and Richard. Hydrogen bonding, hydrophobicity, pakcing, and protein folding. *Ann. Rev. Biophys. Biomol. Struct.*, 22:381–415, 1993.

[11] Patrick J. Fleming and Frederic M. Richards. Protein packing: Dependence on protein size, secondary structure and amino acid composition. *J. Mol. Biol.*, 299: 487–498, 2000.

[12] Jie Liang and Ken A. Dill. Are proteins well-packed? *Biophys. J.*, 81:751–766, 2001.

[13] B. P. Schoenborn, H. C. Watson, and J. C. Kendrew. Binding of xenon to sperm whale myoglobin. *Nature*, 207:28–30, 1965.

[14] B. P. Schoenborn. Binding of xenon to horse haemoglobin. *Nature*, 208:760–762, 1965.

[15] B. P. Schoenborn and C. L. Nobbs. Binding of xenon to sperm whale deoxymyoglobin. *Mol. Pharmacol.*, 2:495–498, 1966.

[16] Benno P. Schoenborn. Structure of alkaline metmyoglobin-xenon complex. *J. Mol. Biol.*, 45:297–298, 1969.

[17] Robert F. Tilton, Jr, Irwin D. Kuntz, Jr, and Gregory A. Petsko. Cavities in proteins: Structure of a metmyoglobin-xenon complex solved to 1.9 Å. *Biochemistry*, 23: 2849–2857, 1984.

[18] R. K. Wierenga, M. E. M. Noble, and R. C. Davenport. Comparison of the refined crystal structures of liganded and unliganded chicken, yeast, and trypanosomal triosephosphate isomerase. *J. Mol. Biol.*, 224:1115–1126, 1992.

[19] B. Lee and F. M. Richards. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.*, 55:379–400, 1971.

[20] Alexander A. Rashin, Michael Iofin, and Barry Honig. Internal cavities and buried waters in globular proteins. *Biochemistry*, 25:3619–3625, 1986.

[21] Robert F. Tilton, Jr., U. C. Singh, S. J. Weiner, M. L. Connolly, Irwin D. Kuntz, Jr, P. A. Kollman, N. Max, and D. A. Case. Computational studies of the interaction of myoglobin and xenon. *J. Mol. Biol.*, 192:443–456, 1986.

[22] Simon J. Hubbard, Karl-Heinz Gross, and Patrick Argos. Intramolecular cavities in globular proteins. *Protein Eng.*, 7:613–626, 1994.

[23] David G. Lambright, Sriram Balasubramanian, Sean M. Decatur, and Steven G. Boxer. Anatomy and dynamics of a ligand-binding pathway in myoglobin: The roles of residue 45, 60, 64, and 68. *Biochemistry*, 33:5518–5525, 1994.

[24] Cheolju Lee, Soon-Ho Park, Min-Youn Lee, and Myeong-Hee Yu. Regulation of protein function by native metastability. *J. Mol. Biol.*, 97:7727–7731, 2000.

[25] Mihail Karpusas, Walter A. Baase, Masazumi Matsumura, and Brian W. Matthews. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc. Natl. Acad. Sci. USA*, 86:8237–8241, 1989.

[26] James T. Kellis, Jr., Kerstin Nyberg, and Alan R. Fersht. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry*, 28:4914–4922, 1989.

[27] A.E. Eriksson, W. A. Baase, J. A. Wozniak, and B. W. Matthews. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature*, 355:371–373, 1992.

[28] A. E. Eriksson, W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthews. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, 255:178–183, 1992.

[29] A. A. Pakula and R. T. Sauer. Genetic analysis of protein stability and function. *Annu. Rev. Genet.*, 23:289–310, 1989.

[30] Brian W. Matthews. Studies on protein stability with T4 lysozyme. *Adv. Prot. Chem.*, 46:249–278, 1995.

[31] Ashley M. Buckle, Patrick Cramer, and Alan R. Fersht. Structural and energeetic responses to cavity-creating mutations in hydrophobic cores: Observation of a buried water molecule and the hydrophobic nature of such hydrophobic cavities. *Biochemistry*, 35:4298–4305, 1996.

[32] Patrick Argos. An investigation of protein subunit and domain interfaces. *Protein Eng.*, 2:101–113, 1988.

[33] Joēl Janin, Susan Miller, and Cyrus Chothia. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.*, 204:155–164, 1988.

[34] Joēl Janin and Cyrus Chothia. The structure of protein-protein recognition sites. *J. Biol. Chem.*, 265:16027–16030, 1990.

[35] Lawren C. Wu, Rita Grandori, and Jannette Carey. Autonomous subdomains in protein folding. *Protein Sci.*, 3:369–371, 1994.

[36] Susan Jones and Janet M. Thornton. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA*, 93:13–20, 1996.

[37] Cyrus Chothia and Joēl Janin. Principles of protein-protein recognition. *Nature*, 256:705–708, 1975.

[38] Alex P. Korn and Roger M. Burnett. Distribution and complementary of hydropathy in multisubunit proteins. *Proteins: Struct. Funct. Bioinf.*, 9:37–55, 1991.

[39] L. Young, R. L. Jernigan, and D. G. Covell. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.*, 3:717–729, 1994.

[40] Michael C. Lawrence and Peter M. Colman. Shape complementary at protein/protein interfaces. *J. Mol. Biol.*, 234:946–950, 1993.

[41] Simon J. Hubbard and Patrick Argos. Cavities and packing at protein interfaces. *Protein Sci.*, 3:2194–2206, 1994.

[42] Shrihari Sonavane and Pinak Chakrabarti. Cavities and atomic pakcing in protein structures and interfaces. *PLoS. Comput. Biol.*, 4:e1000188, 2008.

[43] Shrihari Sonavane and Pinak Chakrabarti. Cavities in protein-DNA and protein-RNA interfaces. *Nucleic Acids Res.*, 37:4613–4620, 2009.

[44] Simon J. Hubbard and Patrick Argos. A functional role for protein cavities in domain:domain motions. *J. Mol. Biol.*, 261:289–300, 1996.

[45] Mark A. Williams, Julia M. Goodfellow, and Janet M. Thornton. Buried waters and internal cavities in monomeric proteins. *Protein Sci.*, 3:1224–1235, 1994.

[46] Gottfried Otting, Edvards Liepinsh, and Kurt Wüthrich. Protein hydration in aqueous solution. *Science*, 254:974–980, 1991.

[47] A. A. Kossiakoff, M. D. Sintchak, J. Shpungin, and L. G. Presta. Analysis of solvent structure in proteins using neutron $D_2O$-$H_2O$ solvent maps: Pattern of primary and secondary hydration of trypsin. *Proteins*, 12:223–236, 1992.

[48] James A. Ernst, Robert T. Clubb, Huan-Xiang Zhou, Angela M. Gronenborn, and G. Marius Clore. Demonstration of positionally disordered water within a protein hydrophobic cavity by nmr. *Science*, 267:1813–1817, 1995.

[49] Kazufumi Takano, Jun Funahashi, Yuriko Yamagata, Satoshi Fujii, and Katsuhide Yutani. Contribution of water molecules in the interior of a protein to the conformational stability. *J. Mol. Biol.*, 274:132–142, 1997.

[50] Gottfried Otting, Edvards Liepinsh, Bertil Halle, and Urban Frey. NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. *Nat. Struct. Biol.*, 4:396–404, 1997.

[51] Jürgen Woenckhaus, Robert R. Hudgins, and Martin F. Jarrold. Hydration of gas-phase proteins: A special hydration site on gas-phase BPTI. *J. Am. Chem. Soc.*, 119: 9586–9587, 1997.

[52] B. Yu, M. Blaber, A. M. Gronenborn, G. M. Clore, and D. L. D. Caspar. Disordered water within a hydrophobic protein cavity visualized by x-ray crystallography. *Proc. Natl. Acad. Sci. USA*, 96:103–108, 1999.

[53] Cécile Bon, Mogens S. Lehmann, and Clive Wilkinson. Quasi-laue neutron-diffraction study of the water arrangement in crystals of triclinic hen egg-white lysozyme. *Acta Cryst. D*, 55:978–987, 1999.

[54] Yipin Lu, Renxiao Wang, Chao-Yie Yang, and Shaomeng Wang. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.*, 47:668–675, 2007.

[55] B. C. Finzel, L. L. Clancy, D. R. Holland, S. W. Muchmore, K. D. Watenpaugh, and H. M. Einspahr. Crystal structure of recombinant human interleukin-1$\beta$ at 2.0 å resolution. *J. Mol. Biol.*, 209:779–791, 1989.

[56] J. P. Priestle, H. P. Schär, and M. G. Grütter. Crystallographic refinement of interleukin 1$\beta$ at 2.0 å resolution. *Proc. Natl. Acad. Sci. USA*, 86:9667–9671, 1989.

[57] A. C. Treharne, D. H. Ohlendorf, P. C. Weber, J. J. Wendoloski, and F. R. Salemme. X-ray structural studies of the cytokine interleukin 1-$\beta$. *Prog. Clin. Biol. Res.*, 349: 309–319, 1990.

[58] B. Veerapandian, Gary L. Gilliland, Reetta Raag, Anders L. Svensson, Yoshihiro Masui, Yoshikatsu Hirai, and Thomas L. Poulos. Functional implications of interleukin-1$\beta$ based on the three-dimensional structure. *Proteins: Struct. Funct. Bioinf.*, 12: 10–23, 1992.

[59] Michael L. Quillin, Paul T. Wingfield, and Brian W. Matthews. Determination of solvent content in cavities in IL-1$\beta$ using experimentally phased electron density. *Proc. Natl. Acad. Sci. USA*, 103:19749–19753, 2006.

[60] G. Marius Clore, Ad Bax, Paul T. Wingfield, and Angela M. Gronenborn. Identification and localization of bound internal water in the solution structure of interleukin 1$\beta$ by heteronuclear three-dimensional $^{1}$H rotating-frame overhause $^{15}$N-$^{1}$H multiple quantum coherence NMR spectroscopy. *Biochemistry*, 29:5671–5676, 1990.

[61] G. Marius Clore, Paul T. Wingfield, and Angela M. Gronenborn. High-resolution three-dimensional structural of interleukin 1$\beta$ in solution by three- and four- dimensional nuclear magnetic resonance spectroscopy. *Biochemistry*, 30:2315–2323, 1991.

[62] Sandeep Somani, Choon-Peng Chng, and Chandra S. Verma. Hydration of a hydrophobic cavity and its functional role: A simulation study of human interleukin. *Proteins: Struct. Funct. Bioinf.*, 67:868–885, 2007.

[63] Hao Yin, Guogang Feng, G. Marius Clore, Gerhard Hummer, and Jayendran C. Rasaiah. Water in the polar and nonpolar cavities of the protein interleukin-1$\beta$. *J. Phys. Chem. B*, 114:16290–16297, 2010.

[64] Kazufumi Takano, Yuriko Yamagata, and Katsuhide Yutani. Buried water molecules contribute to the conformational stability of a protein. *Protein Eng.*, 16:5–9, 2003.

[65] Kurt D. Berndt, Jürgen Beunink, Werner Schröder, and Kurt Wüthrich. Designed replacement of an internal hydration water molecule in BPTI: Structural and functional implications of a glycine-to-serine mutations. *Biochemistry*, 32:4564–4570, 1993.

[66] James C. Covalt Jr, Melinda Roy, and Patricia A. Jennings. Core and surface mutations affect folding kinetics, stability and cooperativity in IL-1$\beta$: Does alteration in buried water play a role? *J. Mol. Biol.*, 307:657–669, 2001.

[67] Johann Deisenhofer and Wolfgang Steigemann. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 å resolution. *Acta Cryst. B*, 31:238–250, 1975.

[68] Alexander Wlodawer, Jochen Walter, Robert Huber, and Lennart Sjölin. Structure of bovine pancreatic trypsin inhibitor: Results of joint neutron and x-ray refinement of crystal form ii. *J. Mol. Biol.*, 180:301–329, 1984.

[69] Alexander Wlodawer, Joseph Nachman, Gary L. Gilliland, Warren Gallagher, and Clare Woodward. Strcuture of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.*, 198:469–480, 1987.

[70] Alexander Wlodawer, Johann Deisenhofer, and Robert Huber. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.*, 193: 145–156, 1987.

[71] Sean Parkin, Bernhard Rupp, and Hôkon Hope. Structure of bovine pancreatic trypsin inhibitor at 125 K: Definition of carboxyl-terminal residues Gly57 and Ala58. *Acta Cryst. D*, 52:18–29, 1996.

[72] Mohammad Monirul Islam, Shihori Sohya, Keiichi Noguchi, Masafumi Yohda, and Yutaka Kuroda. Crystal structure of an extensively simplified variant of bovine pancreatic trypsin inhibitor in which over one-third of the residues are alanines. *Proc. Natl. Acad. Sci. USA*, 105:15334–15339, 2008.

[73] Lewis E. Kay. Protein dynamics from NMR. *Nat. Struct. Mol. Biol.*, 5:513–517, 1998.

[74] Paul D. Adams, Adrienne P. Loh, and Robert E. Oswald. Backbone dynamics of an oncogenic mutant of Cdc42Hs shows increased flexiblity at the nucleotide-bingding site. *Biochemistry*, 43:9968–9977, 2004.

[75] Lilly Wong, Scot Lieser, Barbara Chie-Leon, Osamu Miyashita, Brandon Aubol, Jennifer Shaffer, Josè N. Onuchic, Patricia A. Jennings, Virgil L. Woods Jr, and Joseph A. Adams. Dynamic coupling between the SH2 domain and active site of the COOH terminal Src kinese, Csk. *J. Mol. Biol.*, 341:93–106, 2004.

[76] Yi Mao, Mark A Ratner, and Martin F. Jarrold. One water molecule stiffens a protein. *J. Am. Chem. Soc.*, 122:2950–2951, 2000.

[77] Stefan Fischer and Chandra S. Verma. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. USA*, 96:9613–9615, 1999.

[78] Stefan Fischer, Jeremy C. Smith, and Chandra S. Verma. Dissecting the vibrational entropy change on protein/ligand binding: Burial of a water molecule in bovine pancreatic trypsin inhibitor. *J. Phys. Chem. B*, 105:8050–8055, 2001.

[79] Jeremy C. Smith, Franci Merzel, Chandra S. Verma, and Stefan Fischer. Protein hydration water: Structure thermodynamics. *J. Mol. Liq.*, 101:37–33, 2002.

[80] L. Renee Olano and Steven W. Rick. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.*, 126:7991–8000, 2004.

[81] J. T. Edsall and H. A. McKenzie. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. *Adv. Biophys.*, 16:53–183, 1983.

[82] Uma Sreenivasan and Paul H. Axelsen. Buried water in homologous serine proteases. *Biochemistry*, 31:12785–12791, 1992.

[83] Mounir Tarek and Douglas J. Tobias. The dynamics of protein hydration water: A quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophys. J.*, 79:3244–3257, 2000.

[84] Jian Xu, Walter A. Baase, Michael L. Quillin, Enoch P. Baldwin, and Brian W. Matthews. Structural and thermodynamic analysis of the binding of solvent at internal sites in T4 lysozyme. *Protein Sci.*, 10:1067–1078, 2001.

[85] Motomichi Tashiro and Alexei A. Stuchebrukhov. Thermodynamic properties of internal water molecules in the hydrophobic cavity around the catalytic center of cytochrome *c* oxidase. *J. Phys. Chem. B*, 109:1015–1022, 2005.

[86] Edgar Meyer, Greg Cole, R. Radhakrishnan, and Otto Epp. Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Cryst. B*, 44:26–38, 1988.

[87] Patrick Y. S. Lam, Prabhakar K. Jadhav, Charles J. Eyermann, C. Nicholas Hodge, Yu Ru, Lee T. Bacheler, James L. Meek, Michael J. Otto, Marlene M. Rayner, Y. Nancy Wong, Chong-Hwan Chang, Patricia C. Weber, David A. Jackson, Thomas R. Sharpe, and Susan Erickson-Viitanen. Rational design of potent,

bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, 263:380–384, 1994.

[88] P. C. Weber, M. W. Pantoliano, D. M. Simons, and F. R. Salemme. Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.*, 116: 2717–2724, 1994.

[89] Vincent Mikol, Christos Papageorgiou, and Xaver Borer. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: Synthesis, biological activity, and crystallographic analysis with cyclophilin a. *J. Med. Chem.*, 38:3361–3367, 1995.

[90] J. V. N. Vara Prasad, Elizabeth A. Lunney, Donna Ferguson, Peter J. Tummino, J. Ronald Rubin, Eric. L. Reyner, Barbra H. Stewart, Robert J. Guttendorf, John M. Domagala, Leonid I. Suvorov, Sergei V. Gulnik, Igor A. Topol, T. N. Bhat, and John W. Erickson. HIV protease inhibitors possessing a novel, high-affinity, and achiral $p_1$'/$p_2$' ligand with a unique pattern of *i*n vitro resistance. importance of a conformationally-restricted template in the design of enzyme inhibitors. *J. Am. Chem. Soc.*, 117:11070–11074, 1995.

[91] William E. Royer, Jr., Animesh Pardanani, Quentin H. Gibson, Eric S. Peterson, and Joel M. Friedman. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. USA*, 93:14526–14531, 1996.

[92] Geoffrey A. Holdgate, Alan Tunnicliffe, Walter H. J. Ward, Simon A. Weston, Gina Rosenbrock, Peter T. Barth, Ian W. F. Taylor, Richard A. Pauptit, and David Timms. The entropic penalty of ordered water acccounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: A thermodynamic and crystallographic study. *Biochemistry*, 36:9663–9673, 1997.

[93] Christian Heiss, Maris Laivenieks, J. Gregory Zeikus, and Robert S. Phillips. The stereospecificity of secondary alcohol dehydrogenase from *t*hermoanaerobacter ethanolicus is partially determined by active site water. *J. Am. Chem. Soc.*, 123: 345–346, 2001.

[94] Christopher Clarke, Robert J. Woods, John Gluska, Alan Cooper, Margaret A. Nutley, and Geert-Jan Boons. Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.*, 123:12238–12247, 2001.

[95] Mark J. Snider and Richards Wolfenden. Site-bound water and the shortcomings of a less than perfect transition state analogue. *Biochemistry*, 40:11364–11371, 2001.

[96] David L. Zechel, Alisdair B. Boraston, Tracey Gloster, Catherine M. Boraston, James M. Macdonald, D. Matthew G. Tilbrook, Robert V. Stick, and Gideon J. Davies. Iminosugar glycosidase inhibitors:â‰ structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to Î²-glucosidases. *J. Am. Chem. Soc.*, 125:14313–14323, 2003.

[97] Tracey M. Gloster, Spencer J. Williams, Shirley Roberts, Chris A. Tarling, Jacqueline Wicki, Stephen G. Withers, and Gideon J. Davies. Atomic resolution analyses of the

binding of xylobiose-derived deoxynojirimycin and isofagomine to xylanase Xyn10A. *Chem. Commun.*, pages 1794–1795, 2004.

[98] Garegin A. Papoian, Johan Ulander, and Peter G. Wolynes. Role of water mediated interaction in protein-protein recognition landscapes. *J. Am. Chem. Soc.*, 125:9170–9178, 2003.

[99] Garegin A. Papoian, Johan Ulander, Michael P. Eastwood, Zaida Luthey-Schulten, and Peter G. Wolynes. Water in protein structure prediction. *Proc. Natl. Acad. Sci. USA*, 101:3352–3357, 2004.

[100] Daniel Lafitte, Valérie Lamour, Philippe O. Tsvetkov, Alexander A. Makarov, Michel Klich, Pierre Deprez, Dino Moras, Claudette Briand, and Robert Gilli. DNA gyrase interaction with coumarin-based inhibitors: The role of the hydroxybenzoate isopentenyl moiety and the 5'-methyl group of the noviose. *Biochemistry*, 41:7217–7223, 2002.

[101] Antonio Hernandez Daranas, Hiroki Shimizu, and Steve W. Homans. Thermodynamics of binding of d-galactose and deoxy derivatives thereof to the l-arabinose-binding protein. *J. Am. Chem. Soc.*, 126:11870–11876, 2004.

[102] Michael Levitt and Britt H. Park. Water: Now you see it, now you donâ™t. *Structure*, 1:223–226, 1993.

[103] Gottfried Otting and Kurt Wüthrich. Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Am. Chem. Soc.*, 111:1871–1875, 1989.

[104] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.

[105] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. general method. *J. Chem. Phys.*, 31:459–466, 1959.

[106] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, Oxford, 1987.

[107] J. M. Haile. *Molecular Dynamics Simulation: Elementary Methods*. John Wiley & Sons, New York, 1992.

[108] Daan Frenkel and Smit Berend. *Understanding Molecular Simulations: From Algorithm to Applications*. Academic Press, San Diego, first edition, 1996.

[109] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, Cambridge, second edition, 2004.

[110] Walter Greiner, Neise Ludwig, and Horst Stöcker. *Thermodynamics and Statistical Mechanics*. Springer-Verlag, New York, 1995.

[111] R. K. Pathria and Paul D. Beale. *Statistical Mechanics*. Academic Press, Boston, 2001.

[112] Benjamin Widom. *Statistical Mechanics: A Concise Introduction for Chemists*. Cambridge University Press, Cambridge, 2002.

[113] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education, Harlow UK, second edition, 2001.

[114] G. M. Torrie and J. P. Valleau. Nonphysical sampling distribution in monte carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23:187–199, 1977.

[115] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, 96:1776–1783, 1992.

[116] E. Marinari and G. Parisi. Simulated tempering: A new monte carlo scheme. *Europhys. Lett.*, 19:451–458, 1992.

[117] Ruhong Zhou and B. J. Berne. Smart walking: A new method for boltzmann sampling of protein conformations. *J. Chem. Phys.*, 107:9185–9196, 1997.

[118] Yuji Sugita and Yuko Okamoto. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.*, 329:261–270, 2000.

[119] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.*, 116:9058–9067, 2002.

[120] Michael K. Fenwick and Fernando A. Escobedo. Expanded ensemble and replica exchange methods for simulation of protein-like systems. *J. Chem. Phys.*, 119: 11998–12010, 2003.

[121] Soonmin Jang, Seokmin Shin, and Youngshang Pak. Replica exchange method using the generalized effective potential. *Phys. Rev. Lett.*, 91:058305, 2003.

[122] Ayori Mitsutake and Yuko Okamoto. Replica-exchange extensions of simulated tempering method. *J. Chem. Phys.*, 121:2491–2504, 2004.

[123] P. Liu, Byungchan Kim, Richard A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA*, 102:13749–13754, 2005.

[124] Steven W. Rick. Replica exchange with dynamical scaling. *J. Chem. Phys.*, 126: 054102, 2007.

[125] Alexis J. Lee and Steven W. Rick. Improving replica exchange using driven scaling. *J. Chem. Phys.*, 131:174113, 2009.

[126] Robert W. Zwanzig. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.*, 22:1420–1426, 1954.

[127] John G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3: 300–313, 1935.

# Chapter 2

# Water at the DNA Gyrase-Inhibitor Interface

## 2.1 Introduction

A recent analysis of the crystal structures reveals that over 85% of the structures have at least one water molecule at the protein-ligand interface [1]. These water molecules have significant but not completely understood influences on inhibitor binding thermodynamics. The introduction of ordered, relative to the liquid, water molecules is generally considered to have a significant entropic cost, decreasing the binding affinity of the ligand [2, 3]. This influence can be isolated from other factors by considering changes in the ligand or protein which modify the water structure. Differences in binding affinities between different ligands [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] or the same ligand with mutated proteins [16, 17, 18] have been attributed to the addition of one or more tightly bound water molecules. Both these effects are demonstrated in the DNA gyrase/novobiocin complex.

DNA gyrase is a bacterial enzyme which is a target for several antibiotics [19]. The enzyme from *Escherichia coli* is an $A_2B_2$ tetramer made up of the two subunits A and B. The antibiotic novobiocin is one in a class of coumarin inhibitors which inhibit gyrase by preventing dimerization of the two B subunits [20, 21, 22]. For this system, bound water molecules have been proposed to change the thermodynamics of binding, both by changes

This chapter has been published previously as a paper in the Journal of the American Chemical Society: "*Free Energies and Entropies of Water Molecules at the Inhibitor-Protein Interface of DNA Gyrase*".

Table 2.1: Thermodynamic parameters for the binding of novobiocin and clorobiocin to the 24 kDa fragment of the DNA gyrase B protein for the wild type and Arg 136 His mutant.

| complex | $K_a$ (x $10^6$ $M^{-1}$) | $\Delta G^0$ (kcal/mol) | $\Delta H^0$ (kcal/mol) | $T\Delta S^0$ (kcal/mol) | number of water molecules |
|---|---|---|---|---|---|
| WT-clorobiocin[a] | 860±220 | -12.2±0.1 | -9.5±0.6 | 2.7±0.2 | 1 |
| WT-novobiocin[b] | 23±4 | -10.1±0.1 | -12.2±0.1 | -2.1±0.2 | 3 |
| R136H-novobiocin[b] | 0.83±0.03 | -8.1±0.1 | -14.3±0.1 | -6.1±0.1 | 4 |

[a] Reference [14]
[b] Reference [16]

in the protein through mutations [16] or by changes in the inhibitor [14]. A mutation of Arg-136 to histidine on the B fragment is one of the naturally occurring resistant mutants to coumarin inhibitors [23]. The B fragment with the R136H mutation has an association constant for novobiocin over an order of magnitude smaller than the wildtype (Table 2.1) [16]. The binding has a more favorable enthalpy change, $\Delta H_a$, for the mutant but a much less favorable entropy change, $\Delta S_a$. The structures of the R136H and wildtype of the 24 kDa N-terminal fragment of the B subunit with novobiocin show that the space created by the absence of the guanidinium group of the arginine residue is occupied by a water molecule (water 12 in the 1AJ6 pdb structure and in Figure 2.1) [16]. The large change in $K_a$ is attributed to the presence of this water molecule, which is not present in the wildtype structure. This water molecule is in contact with the solvent.

The water molecules 1 and 11 are located between the carbonate nitrogen attached to the sugar ring on novobiocin and polar atoms (on Val 43 and Asp 73) of the protein. This side of the inhibitor is away from he solvent. For the inhibitor clorobiocin, a pyrrole ring replaces the $NH_2$ group (Figure 2.2). The bulkier group fills the space occupied by W 1 and W 11 in the complex with novobiocin [14]. Clorobiocin binds (wildtype) gyrase over an order of magnitude better than novobiocin, with the improved binding not due to enthalpy, which favors novobiocin, but entropy. The entropic contribution, $-T\Delta S$, is

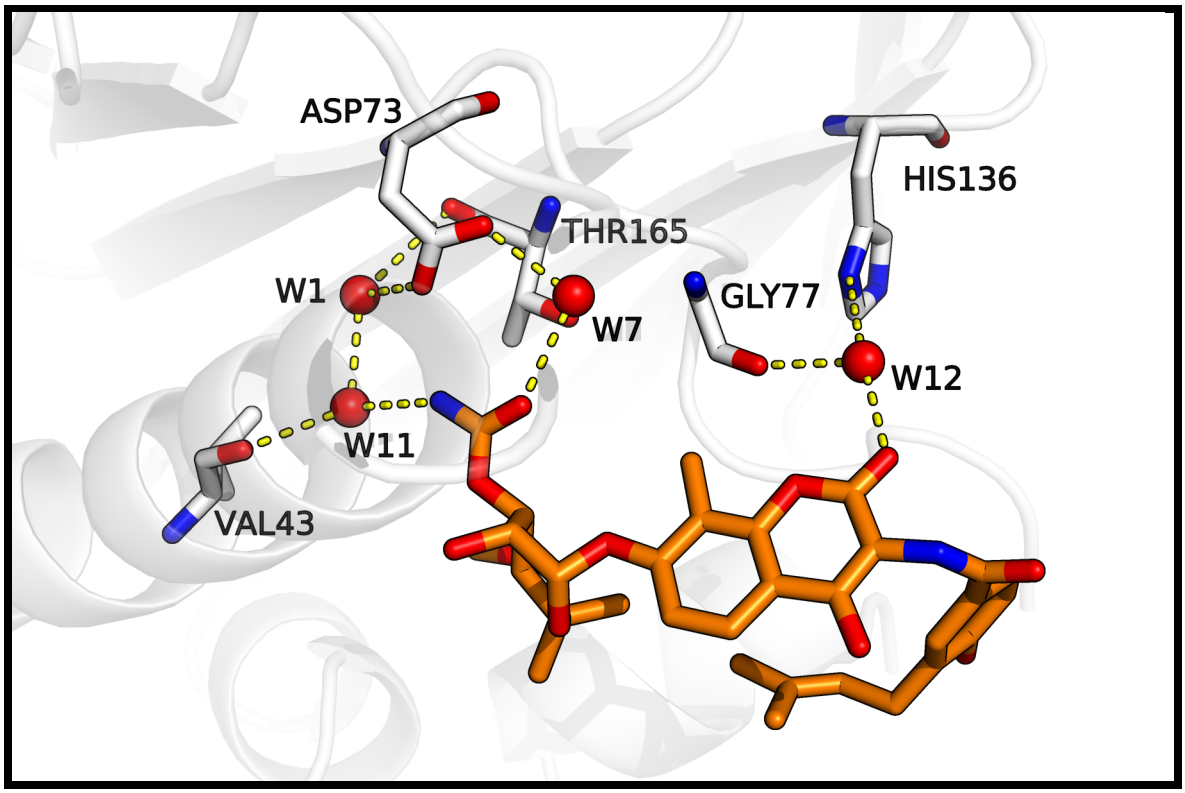


Figure 2.1: PDB structure 1AJ6 of the complex of novobiocin with the R136H mutant of DNA gyrase.

4.8 kcal/mol more favorable for clorobiocin than novobiocin (see Table 2.1). The large thermodynamic changes between clorobiocin and novobiocin, as well as the between the wildtype and mutant proteins, appear to be largely due to only a few water molecules [14, 16].

There are exceptions to the "less water, better binding" heuristic suggested by these studies, in which compounds with more water molecules at the interface bind with greater affinity [15]. In addition, several computational studies using free energy perturbation [24] and inhomogeneous fluid solvent theory [25, 26, 27] have shown that the entropy of the bound water molecules varies considerably and in some cases can be greater than that of bulk water. For the related, but distinct, problem of water displacement from the binding site by the ligand, inhomogeneous fluid solvent theory also finds that the entropy of water molecules in the empty binding site can vary considerably [28]. This

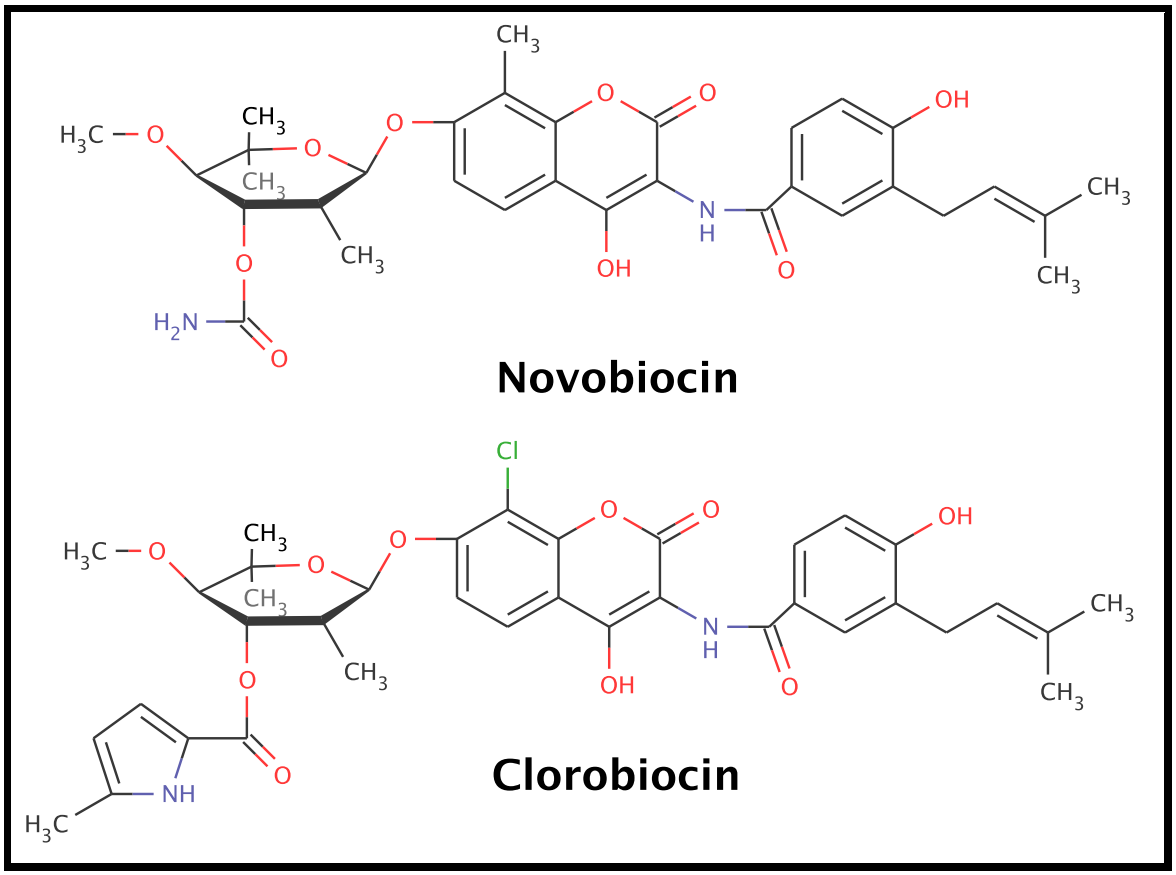


Figure 2.2: Chemical structures of DNA gyrase inhibitors.

variability makes assessing the role of water difficult. A number of studies have evaluated the importance of water molecules using protein-ligand docking and scoring models, with a general, but not universal, consensus that the inclusion of water improves accuracy [29, 30, 31, 32, 33, 34, 35, 36, 37, 38]. The variability of the entropic contribution from the bound water presents a challenge for empirical scoring models [31, 34]. For example, the GOLD scoring model adds a constant entropic penalty term (of 0.5 kcal/mol) for each bound water [34].

The entropic contribution of the bound water molecules appears to be the key to understanding the binding thermodynamics for the novobiocin/clorobiocin-DNA gyrase complex. In this study we use computational methods to calculate the entropy change, as well as the free energy change, for adding water molecules to the three sites which change for the various complexes (W 1, 11, and 12). A number of studies have calculated free

energies for water molecules at protein-ligand interfaces [39, 40, 41, 42] and in protein cavities [24, 43, 44, 45, 46]. These studies have not calculated the entropy change, except for our own study, which calculated free energy changes as a function of temperature to extract the entropy and enthalpy changes [24]. In this study, we will use the same approach. This calculates the exact thermodynamics, depending only of the quality of the potential energy function used.

## 2.2 Methods

### 2.2.1 Free Energy Calculations

The free energy calculations were done in one of two ways, depending on the proximity of the water molecules to the bulk solvent. These two approaches, the alchemical "double decoupling method", in which interactions are turned on, while a restraining potential is turned off, and a potential of mean force (PMF) approach, in which the molecule is made to move from the binding site, have both been applied to many binding free energy calculations, as reviewed recently [47]. The free energy calculations for the water molecules in the 1 and 11 positions are done with the double decoupling method, as described previously, except using thermodynamic integration rather than free energy perturbation theory [24]. In this method, interactions are turned off between the one bound water molecule and all other atoms in the system. The non-interacting molecule is localized in the site with a harmonic potential with a force constant, $k_{harm}$, equal to 3 kcal/mol/Å$^2$. A correction for the harmonic restraint equal to $-kT\ln[\rho(\pi kT/k_{harm})^{3/2}]$ is added to the calculated free energy [45, 46, 48]. This method is similar to the methods developed for other free energy calculations [45, 46, 48] except for the addition of a short-ranged interaction is added to keep other water molecules out of that position. This interaction is of the form $\epsilon(r_{OX}/\sigma)^{-12}$, where $r_{OX}$ is the distance between the position the water molecule is being added, $r_x$, and the oxygen atoms on all other water molecules. The parameters $\sigma$ and $\epsilon$ are set equal to 2.0 Å and 0.143 kcal/mol, respectively. To calculate

the free energy of adding a water molecule to this position relative to the bulk liquid, the free energy of adding a water molecule to the liquid must be calculated. The free energy for this process is -7.04±0.04 kcal/mol and the entropy change $T\Delta S$ is -4.03±0.04 kcal/mol. For the water molecule in site 12, which is in contact with solvent water molecules, the PMF approach was used. In this approach, the interactions of the specified bound water are not turned off, but rather the water molecule is reversibly forced out of the bound water position using the potential $E_{\lambda} = \lambda \sum_i 4\epsilon[(r_{ix}^2 + (1-\lambda)\delta)/\sigma^2]^{-6}$, where the sum is over all water molecule oxygen atoms and $\lambda$ is a free energy variable, varying from 0 to 1. This method uses separated-shifted scaling method, adding the term $(1-\lambda)\delta$ to avoid singularities as $r_{ix}$ goes to zero [49]. The parameter $\delta$ was set equal to 7.0 Å$^2$ and $\sigma$ and $\epsilon$ have the same values as given above. This approach was hoped to be more efficient for calculating hydration free energy changes for water molecules which can move to the solvent easily, but this turned out to be not the case. It was better to turn off the interactions in the protein/ligand site and then, in a separate calculation, turn them on again in the liquid, rather than to force the molecule to exit to the solvent. Free energy calculations for all the water sites were done at three temperatures, so the entropy could be calculated from the temperature dependence, using $\Delta S = -(\Delta G(T + \Delta T) - \Delta G(T + \Delta T)/(2\Delta T)$. The enthalpy change can be found from $\Delta H = \Delta G + T\Delta S$.

### 2.2.2 DNA Gyrase/Novobiocin Structure

All calculations were done with the R136H mutant of the 24 kDa B subunit fragment of DNA gyrase from *Escherichia coli* using the 1AJ6 structure for the gyrase/novobiocin complex [16]. In this structure, there are two loop regions, from residues 83 to 85 and 105 to 111, that are unresolved. These regions were reconstructed using the coordinates of the 1kij structure on the 43 kDa fragment of DNA gyrase from *Thermus thermophilus* in complex with novobiocin [50].

Protonation states of various groups must also be assigned. Novobiocin is an acid with a pKa equal to 4.3 [51], so the acidic proton, on the phenolic oxygen connected to the coumarin double ring, is taken to be absent. In addition to the mutant histidine at position 136, the 24 kDa subunit of *e. coli* DNA gyrase has 11 histidine residues. Based on the pKa calculations for the wildtype of Schechner et al. [52], we assigned the following protonation states (residue number, protonation state): 37 HID, 38 HIP, 55 HIE, 64 HIE, 83 HIE, 99 HIP, 116 HIP, 141 HIE, 147 HIP, 215 HIE, and 217 HIP, where HID is the N$\delta$ tautomer, HIE is N$\epsilon$ tautomer, HIP is doubly protonated. This gives a charge for DNA gyrase equal to -5 and novobiocin adds another negative charge. Due to the overall negative charge, including the nearby novobiocin negative charge, we took the histidine 137 residue to be the doubly protonated positively charged form. In addition, in the 1aj6 structure, there are oxygen atoms (the main chain oxygen on Arg 76 for the N$\delta$ atom and the O$\gamma$ on Thr 80 for the N$\epsilon$ atom) close to both nitrogens on the histidine ring. Also given that it replaces an arginine residue, with a positive one charge, a positively charge histidine seemed the simplest assumption to make. Isothermal titration calorimetry (ICT) measurements with different buffers find that the enthalpy of binding, $\Delta H^0$ for the wildtype/novobiocin complex is independent of the buffer [16]. The two buffers used (20 mM phosphate buffer with pH 7.4 and 69 mM Tris-HCl at pH 7.4) have enthalpies of ionization that are different by 10.5 kcal/mol, so if binding involved proton movement, it will have different $\Delta H^0$ values for the two buffers. For R136H mutant/novobiocin binding, $\Delta H^0$ is different for the two buffers by 3.2 kcal/mol (the values on Table 2.1 are with the Tris buffer) so any proton changes upon binding must involve a partially ionized group [16]. For all these reasons—the overall negative charge, the nearby hydrogen bond acceptors, and the fact that the histidine replaces the positively charged arginine residue—the simplest assumption is that His-136 is doubly protonated and remains so during binding.

### 2.2.3 Simulation Details

All protein molecular dynamics simulations were performed using the Amber7 suite of programs [53]. Charges for novobiocin were generated from a RESP [54] charge fitting procedure with input from Hartree-Fock calculations at the 6-31G* level using the Gaussian03 program [55]. Additional parameters were generated using the gaff parameter set [56]. The charges and gaff parameter type for each atom are given in Supplementary material. The TIP4P-Ew model was used for water [57]. Charge neutrality of the system was created by adding five sodium ions, using the Amber 99 parameter set [58]. The system contain 7891 water molecules. Simulations were ran in the T,P,N ensemble at a pressure of 1 atm and temperatures of 283, 298, and 313 K. All bonds containing hydrogen atoms were constrained with SHAKE, a 1 fs time step was used, and long-ranged electrostatics were treated with particle mesh Ewald. For the water molecules at positions 1 and 11, 15 $\lambda$ values were used, each simulated, on average, for 500 ps. For water molecule at site 12, using the different method, 17 $\lambda$ values were used, each simulated, on average, for 1200 ps.

## 2.3 Results

Tables 2.2 and 2.3 give the calculated free energy changes for the addition of the water molecules to the positions labelled in Figure 2.1. The value for W11 is calculated twice, once with the water at position W1 and once without. The value without the W1 water is listed as W11' and this value plus that for W1 gives the free energy for the addition of the two water molecules to the empty cavity. Table 2.2 gives the results for the free energy calculations at the three temperatures. To the values for the W1, W11, and W11' sites have been added the harmonic restraint correction [45, 46, 48]. To get the hydration free energy (the difference between the free energy of a water molecule in the bulk liquid and in the specified site) for these three sites, the free energy of a water molecule in the

Table 2.2: The free energies for the addition of a water molecule to the various positions at different temperatures.

| | $\Delta G$ (kcal/mol) | | |
|---|---|---|---|
| T (K) | 283 | 298 | 313 |
| W1 | -15.12±0.08 | -14.73±0.10 | -14.44±0.09 |
| W11 | -10.26±0.11 | -9.87±0.14 | -9.67±0.15 |
| W11' | -7.76±0.12 | -7.42±0.14 | -7.32±0.11 |
| W12 | -2.88±0.14 | -2.81±0.14 | -2.76±0.12 |

liquid has to be subtracted. For the water in position W12, the method used finds the free energy difference between that site and bulk water directly. For all water molecules, the free energy of hydration is negative, indicating that water is stable in that position, the entropy change is negative, and the enthalpy change is negative. One possible exception is W11' water, for which $\Delta G$, $\Delta H$, and $\Delta S$ are about zero, and so this water is only stable if there is a neighboring water at position W1. The large enthalpic change, which outweighs the unfavorable entropy change, leads to the stability of the water molecules in these positions.

The average number of hydrogen bonds each water molecule makes to the protein, the inhibitor, or other water molecules from our simulations (at the endpoint of the free energy calculation when the water is fully interacting with the rest of the system) is given on Table 2.3. The Mancera and Buckingham [59] definition of a hydrogen bond is used in which a hydrogen bond is taken to exist if the oxygen-oxygen distance is less than 3.6Å and the angle between the O-H vector on the hydrogen bonding donor and the O-O vector is between 130° and 180°. With this criteria, the water at W1 forms 2.2 hydrogen bonds on average. One hydrogen bond is made to the water at W11 (W11 acts as the hydrogen bond donor) and another is made to the O$\delta$ atom of Asp-73 (W1 is the donor). Another

Table 2.3: Thermodynamic properties for the transfer of a water molecule from the solvent to the various positions.

| | ΔG (kcal/mol) | ΔH (kcal/mol) | TΔS (kcal/mol) | number of hydrogen bonds | change in hydrogen bonds |
|---|---|---|---|---|---|
| W1 | -7.7±0.1 | -10.5±1.2 | -2.8±1.2 | 2.2±0.1 | 2.2±0.1 |
| W11 | -2.8±0.1 | -4.7±1.9 | -1.9±1.9 | 2.9±0.1 | 1.8±0.1 |
| W11' | -0.4±0.1 | -0.1±1.6 | -0.3±1.6 | 1.4±0.2 | 0.8±0.4 |
| W1+W11' | -8.1±0.2 | -10.6±2.0 | -3.1±2.0 | 3.6±0.2 | 3.0±0.4 |
| W12 | -2.8±0.1 | -3.9±1.8 | -1.2±1.8 | 2.7±0.1 | 1.9±0.1 |

hydrogen bond is made a fraction of the time (0.2) to the Thr-165 O atom. The water at W11 makes about 3 hydrogen bonds. In addition to the hydrogen bond to the W1 water, it makes a hydrogen bond to the Val-43 O atom (W11 is the donor) and the another with one of the amide hydrogens connected to the noviose sugar of the Novobiocin molecule. The W12 water forms about 3 hydrogen bonds, one with the Gly-77 O atom, one with the phenolic oxygen on the coumarin double ring on the Novobiocin molecule (which is taken to be unprotonated, see Methods), both of these as a donor, and a third (made a fraction 0.7 of the time) with a solvent water molecule, as an acceptor. A hydrogen bond with the N$\delta$ atom on His-136 is rarely made. These hydrogen bonds are indicated in Figure 2.1.

The formation of hydrogen bonds with the water molecules is consistent with the negative $\Delta H$. Less consistent is the fact that W11 and W12 form the most hydrogen bonds, about 3, but have a less favorable $\Delta G$ than W1, which forms less hydrogen bonds. More strongly correlated to the free energy is the change in the number of hydrogen bonds, or the number of hydrogen bonds that form as the water is added minus the number that are lost. To do this, we identified the atoms which formed hydrogen bonds to the specific water molecule (those atoms are mentioned in the previous paragraph) and calculated the number of hydrogen bonds those atoms form with and without the water in that position.

In all cases, the hydrogen bonds change by fractional amount, rather than through the formation of new contacts which occur only when the water molecule is absent. For example, one of the W12 water's hydrogen bond partner, the Gly-77 O atom, decreases the number of hydrogen bonds it forms upon the addition of the water by 0.4±0.2, mostly due to a decrease in the fraction of time a hydrogen bond is made with the His-136 N$\delta$ atom (from 0.8±0.1 without the water to 0.5±0.1 with water). The Novobiocin phenolic oxygen atom does not changes its hydrogen bond structure noticeably. The third significant hydrogen bond partner with the water at site W12 is a solvating water molecule. A water molecule closest to the position of W12 decreases its number of hydrogen bonds to other water molecules, not counting W12, from 2.8±0.3 to 2.4±0.1, upon addition of W12. This means that the gain of hydrogen bonds between W12 and its closest neighbor (made a fraction 0.7 of the time) is partially offset (by 0.4) by a loss in hydrogen bonds with other neighboring water molecules. The net effect is that the addition of the W12 water leads to an increase in hydrogen bonds by only 1.9 rather than 2.7.

The hydrogen bond neighbors of water W11 also change the number of hydrogen bonds with other atoms as this water is added. W11's neighbor Val-43 O shows a decrease in the fraction of the time a hydrogen bond is made to other protein atoms (a hydrogen bond to the Asn-46 N-H changes from 0.46±0.08 to 0.10±0.04 and to the Ala-47 N-H changes from 0.5±0.1 to 0.19±0.07). The amide N atom on novobiocin, which also hydrogen bonds to W11, does not show any appreciable change in hydrogen bonds upon addition of W11. The third hydrogen bond partner of W11 is W1. This water shows a decrease of 0.5±0.3 in hydrogen bonds with other atoms as W11 is added. This decrease is mostly due to a loss in a hydrogen bond with the Thr-165 O atom, which changes from 0.4±0.2 to 0.09±0.04. This all leads to a net change in hydrogen bonds equal to 1.8±0.1. When the W11 is added without the water W1 there (in the change labelled W11'), a similar analysis reveals that the hydrogen bonded neighbors lose 0.6±0.3 as the 1.4±0.2 hydrogen bonds are made to give a net change in hydrogen bonds equal to 0.8±0.4 hydrogen bonds.

For the W1 water, no hydrogen bonds are lost as that water is added so that the change in hydrogen bonds is simply 2.2±0.1.

These hydrogen bond changes are shown on Table 2.3 and show a stronger correlation with $\Delta G$ than simply the number of hydrogen bonds made. This analysis helps to explain why the W1 position has more favorable $\Delta G$ than the W11 and W12 positions, because while water molecules at both those positions make more hydrogen bonds with their neighbors, they also disrupt more hydrogen bonds, leading to a smaller change in hydrogen bonds than at the W1 position. (In addition, the water at position W1 makes a hydrogen bond with the negatively charged Asp-73 O$\delta$ atom, so this hydrogen bond is stronger than average, with a more favorable $\Delta H$.) Both the entropic contribution to the free energy, $T\Delta S$, which increases, and the enthalpy, which decreases, show a correlation with the change in the hydrogen bonds.

## 2.4 Conclusion

The calculations find that the addition of the water molecules to the protein-ligand interface is entropically unfavorable. The resulting $T\Delta S$ values are -1.1±1.8 kcal/mol for the water at position W12 and -3.1±2.0 kcal/mol for the addition of the two water molecules at sites W1 and W11 (see Figure 2.1). These can be compared to the difference in the entropy changes for the binding of novobiocin or clorobiocin to the wildtype or R136H mutant of DNA gyrase. The R136H mutation introduces the W12 water to the novobiocin/gyrase interface and results in a decrease in $T\Delta S$ equal to 4.0±0.2 kcal/mol. Our calculations suggest that 1 kcal/mol of this is due to the ordered water. The binding of clorobiocin, which eliminates the W1 and W 11 water molecules, results in a decrease in $T\Delta S$ equal to 4.8±0.2 kcal/mol, of which our calculations would suggest that 3.1 kcal/mol of that is due to the two water molecules. Taken together, our results indicate that a sizable fraction, but not all, of the entropic differences in the binding of ligands involving different numbers of water molecules is directly due to the water molecules.

The range of values of the entropy changes show that not all water molecules would have the same entropic penalty to the binding thermodynamics of ligands. This variability is in agreement with earlier theoretical studies using free energy perturbation [24] and inhomogeneous fluid solvent theory [25, 26, 27] as well as conclusions drawn from experimental data on protein stability [60]. The water molecules considered in this study form different numbers of hydrogen bonds with neighboring atoms, including other water molecules, the protein, and the inhibitor. However, the free energy, entropy, and enthalpy changes correlate more strongly with the *change* in the hydrogen bonds made as the water molecule is added, rather than the number of hydrogen bonds the water makes (see Table 2.3). The number of hydrogen bonds change by an amount different from that directly involving the added water because the added water disrupts the local hydrogen bonds that are formed in the absence of this water molecule. This disruption is due to hydrogen bonds, made both with and without the water, that are made less often when the water is there. The presence of the water gives the neighboring atoms another hydrogen bond partner and this appears to decrease the probability that other hydrogen bonds are made. This change in the local hydrogen bond structure has previously been reported for water molecules in protein interiors [24] and indicates that the entropy change for localizing a water molecule at a specific site, with a loss of conformational freedom, is partially offset by a gain in conformation freedom of the nearby atoms involved in the hydrogen bonds.

For the design of drugs, having a water molecule with more than two hydrogen bonds is entropically unfavorable and it would be better to modify the ligand to eliminate the water or to reduce the number of hydrogen bonds that water can make. A fit to our data suggests that $T\Delta S$ should decrease by an amount of 1.7 kcal/mol for each hydrogen bond made by the addition of the water molecule. This is very close to the value of 1.6 kcal/mol by Cooper, an estimate based on the loss of degrees of freedom of a bound water for each hydrogen bond made [61]. Although there is no reason for the entropy change to

have a linear dependence on the number of hydrogen bonds, each hydrogen bond will limit rotational and translational freedom, consistent with a decrease in entropy. Analyses of crystal structure B factors for bound waters in protein interiors [62] and at protein-ligand interfaces [1] reveal that the water molecules become more localized as hydrogen bonds are made, but the decrease is not linear. The decrease is large as each of the first three hydrogen bonds are made and adding the forth has little or no effect on the B factor, suggesting that there is little entropic penalty for forming the fourth hydrogen bond. For the range of hydrogen bonds made for by water molecules in the present study (one to two), the $T\Delta S$ changes from about 0 to -2 kcal/mol. The hydrogen bond numbers of these three water molecules are fairly typical of water molecules found at protein/ligand interfaces. The analysis of crystal structures by Wang, *et al.*, reveals that water molecules at protein/ligand interfaces form on average three hydrogen bonds, with two being almost equally likely [1]. The entropies, as well as the free energies, of the water molecules in this study may be therefore fairly representative of water molecules commonly found in protein/inhibitor complexes, but the values for specific molecules will depend on their local environment.

## Bibliography


[1] Yipin Lu, Renxiao Wang, Chao-Yie Yang, and Shaomeng Wang. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.*, 47:668–675, 2007.

[2] Jack D. Dunitz. The entropic cost of bound water in crystals and biomolecules. *Science*, 264:670, 1994.

[3] John E. Ladbury. Just add water! the effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.*, 3: 973–980, 1996.

[4] Patrick Y. S. Lam, Prabhakar K. Jadhav, Charles J. Eyermann, C. Nicholas Hodge, Yu Ru, Lee T. Bacheler, James L. Meek, Michael J. Otto, Marlene M. Rayner, Y. Nancy Wong, Chong-Hwan Chang, Patricia C. Weber, David A. Jackson, Thomas R. Sharpe, and Susan Erickson-Viitanen. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, 263:380–384, 1994.

[5] Vincent Mikol, Christos Papageorgiou, and Xaver Borer. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: Synthesis, biological activity, and crystallographic analysis with cyclophilin a. *J. Med. Chem.*, 38: 3361–3367, 1995.

[6] P. C. Weber, M. W. Pantoliano, D. M. Simons, and F. R. Salemme. Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.*, 116: 2717–2724, 1994.

[7] J. V. N. Vara Prasad, Elizabeth A. Lunney, Donna Ferguson, Peter J. Tummino, J. Ronald Rubin, Eric. L. Reyner, Barbra H. Stewart, Robert J. Guttendorf, John M. Domagala, Leonid I. Suvorov, Sergei V. Gulnik, Igor A. Topol, T. N. Bhat, and John W. Erickson. HIV protease inhibitors possessing a novel, high-affinity, and achiral $p_1$'/$p_2$' ligand with a unique pattern of *i*n vitro resistance. importance of a conformationally-restricted template in the design of enzyme inhibitors. *J. Am. Chem. Soc.*, 117:11070–11074, 1995.

[8] William E. Royer, Jr., Animesh Pardanani, Quentin H. Gibson, Eric S. Peterson, and Joel M. Friedman. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. USA*, 93:14526–14531, 1996.

[9] Christian Heiss, Maris Laivenieks, J. Gregory Zeikus, and Robert S. Phillips. The stereospecificity of secondary alcohol dehydrogenase from *t*hermoanaerobacter

ethanolicus is partially determined by active site water. *J. Am. Chem. Soc.*, 123: 345–346, 2001.

[10] Christopher Clarke, Robert J. Woods, John Gluska, Alan Cooper, Margaret A. Nutley, and Geert-Jan Boons. Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.*, 123:12238–12247, 2001.

[11] Mark J. Snider and Richards Wolfenden. Site-bound water and the shortcomings of a less than perfect transition state analogue. *Biochemistry*, 40:11364–11371, 2001.

[12] David L. Zechel, Alisdair B. Boraston, Tracey Gloster, Catherine M. Boraston, James M. Macdonald, D. Matthew G. Tilbrook, Robert V. Stick, and Gideon J. Davies. Iminosugar glycosidase inhibitors:â‰ structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to Î²-glucosidases. *J. Am. Chem. Soc.*, 125:14313–14323, 2003.

[13] Tracey M. Gloster, Spencer J. Williams, Shirley Roberts, Chris A. Tarling, Jacqueline Wicki, Stephen G. Withers, and Gideon J. Davies. Atomic resolution analyses of the binding of xylobiose-derived deoxynojirimycin and isofagomine to xylanase Xyn10A. *Chem. Commun.*, pages 1794–1795, 2004.

[14] Daniel Lafitte, Valérie Lamour, Philippe O. Tsvetkov, Alexander A. Makarov, Michel Klich, Pierre Deprez, Dino Moras, Claudette Briand, and Robert Gilli. DNA gyrase interaction with coumarin-based inhibitors: The role of the hydroxybenzoate isopentenyl moiety and the 5'-methyl group of the noviose. *Biochemistry*, 41:7217–7223, 2002.

[15] Antonio Hernandez Daranas, Hiroki Shimizu, and Steve W. Homans. Thermodynamics of binding of d-galactose and deoxy derivatives thereof to the l-arabinose-binding protein. *J. Am. Chem. Soc.*, 126:11870–11876, 2004.

[16] Geoffrey A. Holdgate, Alan Tunnicliffe, Walter H. J. Ward, Simon A. Weston, Gina Rosenbrock, Peter T. Barth, Ian W. F. Taylor, Richard A. Pauptit, and David Timms. The entropic penalty of ordered water acccounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: A thermodynamic and crystallographic study. *Biochemistry*, 36:9663–9673, 1997.

[17] Karen M. Peterson, K. V. Gopalan, Andreas Nandy, and D. K. Srivastava. Influcence of Glu-376 → Gln mutation on enthalpy and heat capacity changes for the binding of slightly altered ligands to medium chain acyl-CoA dehydrogenase. *Protein Sci.*, 10: 1822–1834, 2001.

[18] Patrick R. Connelly, Robert A. Aldape, Frank J. Bruzzese, Stephen P. Chambers, Matthew J. Fitzgibbon, Mark A. Fleming, Susumu Itoh, David J. Livingston, Manuel A. Navia, John A. Thomson, and Keith P. Wilson. Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc. Natl. Acad. Sci. USA*, 91: 1964–1968, 1994.

[19] Richard J. Reece and Anthony Maxwell. The C-terminal domain of the *Escherichia coli* DNA gyrase: A subunit is a DNA-binding proteins. *Nucleic Acids Res.*, 19:1399–1405, 1991.

[20] Janid A. Ali, Andrew P. Jackson, Alison J. Howells, and Anthony Maxwell. The 43-kilodalton N-terminal fragment of the DNA gyrase B protein hydrolyzes ATP and binds coumarin drugs. *Biochemistry*, 32:2717–2724, 1993.

[21] E. Jane Gilbert and Anthony Maxwell. The 24 kDa N-terminal sub-domain of the DNA gyrase B protein binds coumarin drugs. *Mol. Microbiol.*, 12:365–373, 1994.

[22] Janid A. Ali, George Prphanides, and Anthony Maxwell. Nucleotide binding to the 43-kilodalton N-terminal fragment of the DNA gyrase B protein. *Biochemistry*, 34: 9801–9808, 1995.

[23] Asunción Contreras and Anthony Maxwell. *gyrB* mutations which confer coumarin resistance also affect DNA supercoiling and ATP hydrolysis by *Escherichia coli* DNA gyrase. *Mol. Microbiol.*, 6:1617–1624, 1992.

[24] L. Renee Olano and Steven W. Rick. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.*, 126:7991–8000, 2004.

[25] Zheng Li and Themis Lazaridis. Thermodynamics contribution of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.*, 125:6636–6637, 2003.

[26] Zheng Li and Themis Lazaridis. The effect of water displacement on binding thermodynamics: Concanavalin A. *J. Phys. Chem. B*, 109:662–670, 2005.

[27] Zheng Li and Themis Lazaridis. Thermodynamics of buried water clusters at a protein-ligand binding interface. *J. Phys. Chem. B*, 110:1464–1475, 2006.

[28] Robert Abel, Tom Young, Ramy Farid, Bruce J. Berne, and Richard A. Friesner. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, 120:2817–2831, 2008.

[29] Louise Birch, Christopher W. Murray, Michael J. Hartshorn, Ian J. Tickle, and Marcel L. Verdonk. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput.-Aided Mol. Des.*, 16:855–869, 2002.

[30] Renxiao Wang, Yipin Lu, and Shaomeng Wang. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.*, 46:2287–2303, 2003.

[31] Micaela Fornabaio, Francesca Spyrakis, Andrea Mozzarelli, Pietro Cozzini, Donald J. Abraham, and Glen E. Kellogg. Simple, intuitive calculations of free energies of binding for protein-ligand complexes. 3. the free energy contribution of structural water molecules in HIV-1 protease complexes. *J. Med. Chem.*, 47:4507–4516, 2004.

[32] Richard A. Friesner, Jay L. Banks, Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, Eric H. Knoll, Mee Shelley, Jason K. Perry, David E. Shaw, Perry Francis, and Peter S. Shenkin. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.*, 47:1739–1749, 2004.

[33] David G. Lloyd, Alfonso T. García-Sosa, Ian L. Alberts, Nikolay P. Todorov, and Ricardo L. Mancera. The effect of tightly bound water molecules on the structural

interpretation of ligand-derived pharmacophore models. *J. Comput.-Aided Mol. Des.*, 18:89–100, 2004.

[34] Marcel L. Verdonk, Gianni Chessari, Jason C. Cole, Michael J. Hartshorn, Christopher W. Murray, J. Willem M. Nissink, Richard D. Taylor, and Robin Taylor. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.*, 48: 6504–6515, 2005.

[35] Alfonso T. García-Sosa, Stuart Firth-Clark, and Ricardo L. Mancera. Including tightly-bound water molecules in de novo drug design. exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands. *J. Chem. Inf. Model.*, 45: 624–633, 2005.

[36] Chris de Graaf, Pavel Pospisil, Wouter Pos, Gerd Folkers, and Nico P. E Vermeulen. Binding mode prediction of cytochrome P450 and thymidine kinese protein-ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.*, 48:2308–2318, 2005.

[37] Chris de Graaf, Chris Oostenbrink, Peter H. Keizers, Tushar van der Wijst, Aldo Jongejan, and Nico P. E. Vermeulen. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.*, 49:2417–2430, 2006.

[38] Benjamin C. Roberts and Ricardo L. Mancera. Ligand-protein docking with water molecules. *J. Chem. Inf. Model.*, 48:397–408, 2008.

[39] Volkhard Helms and Rebecca C. Wade. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: A molecular dynamics study. *Biophys. J.*, 69:810–824, 1995.

[40] Donald Hamelberg and J. Andrew McCammon. Standard free energy of releasing a localized water molecule from the binding of pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc.*, 126:7683–7689, 2004.

[41] Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.*, 128: 11830–11839, 2006.

[42] Caterina Barillari, Justine Taylor, Russell Viner, and Jonathan W. Essex. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.*, 129:2577–2587, 2007.

[43] Rebecca C. Wade, Michael H. Mazor, J. Andrew McCammon, and Florante A. Quiocho. Hydration of cavities in proteins: A molecular dynamics approach. *J. Am. Chem. Soc.*, 112:7057–7059, 1990.

[44] Rebecca C. Wade, Michael H. Mazor, J. Andrew McCammon, and Florante A. Quiocho. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers*, 31:919–931, 1991.

[45] Li Zhang and Jan Hermans. Hydrophilicity of cavities in proteins. *Proteins: Struct. Funct. Bioinf.*, 24:433–438, 1996.

[46] Benoît Roux, Mafalda Nina, Régis Pomès, and Jeremy C. Smith. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.*, 71:670–681, 1996.

[47] Yuqing Deng and Benoît Roux. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B*, 113:2234–2246, 2009.

[48] Michael K. Gilson, James A. Given, Bruce L. Bush, and J. Andrew McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, 72:1047–1069, 1997.

[49] M. Zacharias, T. P. Straatsma, and J. A. McCammon. Separation-shifted scaling, a new scaling method for lennard-jones interaction in thermodynamic intergration. *J. Chem. Phys.*, 100:9025–9031, 1994.

[50] Valérie Lamour, Laurence Hoermann, Jean-Marc Jeltsch, Pierre Oudet, and Dino Moras. An open conformation of the *thermus thermophilus* gyrase B ATP-binding domain. *J. Biol. Chem.*, 277:18947–18953, 2002.

[51] Susan Budavari, Maryadele J. O'Neil, Ann Smith, Patricia E. Heckelman, and Joanne F. Kinneary, editors. *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*. Merck & Co., Inc., Whitehouse Station, NJ, 1996. p 1155, no. 6818.

[52] Martina Schechner, Annick P. Dejaegere, and Roland H. Stote. Effects of loop conformation on *pka* and ligand binding in DNA gyrase B. *Int. J. Quantum Chem.*, 98: 378–387, 2004.

[53] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Growley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. Amber7, 2002. University of California, San Francisco.

[54] Christopher I. Bayly, Piotr Cieplak, Wendy D. Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, 97:10269–10280, 1993.

[55] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford,

J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian03, 2003. Gaussian Inc., Pittsburgh, PA.

[56] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25: 1157–1174, 2004.

[57] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffry D. Madura, Thomas J. Dick, Greg L. Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, 120:9665–9678, 2004.

[58] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21:1049–1074, 2000.

[59] R. L. Mancera and A. D. Buckingham. Temperature effects on the hydrophobic hydration of ethane. *J. Phys. Chem.*, 99:14632–14640, 1995.

[60] Kazufumi Takano, Jun Funahashi, Yuriko Yamagata, Satoshi Fujii, and Katsuhide Yutani. Contribution of water molecules in the interior of a protein to the conformational stability. *J. Mol. Biol.*, 274:132–142, 1997.

[61] Alan Cooper. Heat capacity effects in protein folding and ligand binding: A re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.*, 115:89–97, 2005.

[62] Sheldon Park and Jeffery G. Saven. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins: Struct. Funct. Bioinf.*, 60:450–463, 2005.

# Chapter 3

# Water in Various Protein Cavities

## 3.1 Introduction

Space due the packing arrangements of folded proteins or the binding geometries of protein-ligand or protein-protein complexes can be occupied by water molecules [1, 2, 3, 4]. Protein interiors contain on average one water molecule for every 27 residues, as identified by X-ray crystallography, nuclear magnetic resonance, and neutron diffraction experiments [1]. Water molecules are about as common inside membrane proteins as inside water soluable proteins [5] and at the protein-ligand interface, there is at least one water molecule in over 85% of crystal structures [4]. These water molecules have important influences on protein function and stability. For example, in a recent study mutations which change the hydrogen bonding of one internal water molecule lead to structural changes in a distant binding pocket [6]. In ligand binding, the addition of one water molecule can increase binding affinity by a factor of 10 or more [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. The water molecules are found in a wide variety of local environments. The number of polar contacts a water makes, either with other water molecules, the protein, or a ligand, varies from zero to about five, with most having two or three [1, 2, 4]. Most water molecules then form less than the four hydrogen bonds possible. This is reflective of the types of cavities available to water molecules but also,

This chapter has been published previously as a paper in the Journal of Physical Chemistry B: "*Free Energy, Entropy, and Enthalpy of a Water Molecule in Various Protein Environments*".

possibly, indicative of the way proteins have evolved to include water. It may be that having fewer than four hydrogen bonds to interior water molecules is beneficial to function or to binding.

A number of studies have calculated free energies for water molecules in protein cavities [21, 22, 23, 24, 25] and at protein-ligand interfaces [26, 27, 28, 29, 30, 31] using free energy perturbation, thermodynamic integration or similar methods. From these studies, some ideas about what stabilizes bound water have emerged. Zhang and Hermans concluded that the water-protein energy must be lower than -12 kcal/mol in order for the site to be occupied [23]. Olano and Rick calculated the free energies for two water molecules, one which make only one hydrogen bond and another which can make four, and found that only the later should be stable [25]. This is consistent with the results of Barillari et al., who calculated free energies for a variety of different water molecules at the ligand-protein interface [29]. That study found that water molecules which formed only two hydrogen bonds only had a weakly negative free energy (of about -1 to -2 kcal/mol) while those with more hydrogen bonds were much more stable. In another study, the free energies of bound water at the protein-ligand interface were found to depend on the change in the number of hydrogen bonds due the presence of the water, rather than just the number of hydrogen bonds made with the water [30]. The protein can adjust to the loss of the water and increase by a small amount the number of protein-protein hydrogen bonds.

Note that a protein-water energy of -12 kcal/mol is less than the energy to remove a water molecule from the liquid phase (-10 kcal/mol) implying that there is an entropic contribution of about 2 kcal/mol to be overcome. Estimates entropic cost based on the entropy difference between ice or crystalline hydrate salts and water also give a value of 2 kcal/mol [32]. The entropy change, like the free energy change, is dependent on the local environment. Simulation results find that the entropy contribution to the free energy, $-T\Delta S$, varies from 3 kcal/mol in a hydrophilic, hydrogen bonding, environment

to -5 kcal/mol in a hydrophobic environment [25, 30]. This increase in entropic cost as the number of hydrogen bonds increases suggests that a balance between entropy and enthalpy may help make water molecules stable in environments with less than the optimal number of hydrogen bonds. These free energy calculations using free energy perturbation or thermodynamic integration represent exact free energies, subject only the accuracy of the potential models. Other more approximate methods have been used to understand the entropy of the bound water molecules. Simple theoretical estimates using the entropy of melting of ice give that $-T\Delta S$ should increase by 1.6 kcal/mol for each hydrogen bond [33]. Calculations using inhomogeneous fluid solvent theory find that $-T\Delta S$ of bound water molecules can vary from 3, greater than the Dunitz estimate, [32] to -0.5 kcal/mol [34, 35, 36]. These results show a good correlation with the number of hydrogen bonds the water molecule forms, although the correlation is not perfect [36]. In general, the water molecules with more hydrogen bonds have a lower entropy, but there are some exceptions and some water molecules with the same number of hydrogen bonds can have much different entropies.

A single water molecule does not appear to occupy cavities that are purely hydrophobic with no possibilities for hydrogen bonds [23, 25, 26]. Only as the cavities get larger and can hold more that one water molecule, can non-polar cavities be occupied by water [37, 38]. The free energy calculations of Vaitheeswaran et al. for water in non-polar, graphine-like spherical cavities of various diameters found that the smallest cavity containing water is 1.0 nm in diameter and contains three water molecules [37]. The stability of the water molecules in these cavities is sensitive to the interactions between the water molecules and the atoms comprising the cavity wall. In addition, there is a change from being entropically favorable to entropically unfavorable as the number of water molecules increases. The minimum cavity size for a non-polar cavity is consistent with the results of Roux and coworkers, who found that a hydrophobic cavity in the bacteriorhodopsin proton channel contains four water molecules [24]. Other studies have revealed that large

hydrophobic pores in proteins are apparently unoccupied by water [39, 40, 41]. A narrow, "sock" shaped, pocket in the mouse urine protein large enough to hold five water molecules was proposed, based on NMR, molecular dynamics, and isothermal titration calorimetry results, to contain no water [39]. Water inside this pocket was found to be enthalpically unfavorable but entropically favorable. In another study using molecular dynamics, water was found to be absent from a narrow, hydrophobic tube-shaped pore, large enough to hold seven water molecules, in the COX-2 protein [40]. Adding water to this site was also predicted to be enthalpy unfavorable. Similarily, NMR and simulation data reveal that a large 315 $Å^3$ binding pore, big enough to hold five water molecules, in bovine beta-lactoglobulin is completely empty of water [41]. Other comparably sized hydrophobic pockets may contain water. The bowl-shaped deep-cavity cavitand molecule contains about four water molecules, as revealed by MD simulations [42]. In this case, adding water is enthalpically favorable and entropically unfavorable. Shape, as well as size and the strength of the interactions, appears to have a strong influence on the stability of the water molecules.

In the present study, we are considering only cavities large enough to hold one water molecule, with the goal of understanding how the free energy, entropy, and enthalpy change as the cavity changes in hydrophilicity and size. To allow for control over the details of the cavity, model cavities are made of the regions in direct contact with the water. Atoms beyond this local region will be treated approximately, taking into account electrostatic and Lennard-Jones interactions. Everything specific about a particular cavity environment in this model is in the local region, the non-local region is described using average properties of proteins. This allows for predictions of the thermodynamics of bound waters based on local details of the cavity, such as the number of hydrogen bonds or the volume. It has been demonstrated, however, that non-local interactions can influence the free energy of binding [28]. The free energy calculations of Lu et al. on the binding of a water molecule at the interface between HIV-1 protease and an inhibitor find that the

protonation state of an aspartic acid residue not in direct contact with the water changes the free energy of the water by about 1 kcal/mol [28]. The deprotonated, charged, state gives a more favorable free energy. Other non-local contributions to the free energy are purely entropic, coming from changes in the vibrational entropy. The addition of a water molecule appears to change the flexibility of the protein, either increasing or decreasing it [25, 43, 44, 45], but this change is not limited to parts of the protein near the water [25]. The binding of ligands other than water also appears to change the flexibility of a protein [46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56] and these changes also involve non-local regions of the protein. Studies using MD and NMR have shown that local regions may get less flexible and some non-local regions more flexible upon binding ligands, to give partially compensating contributions to the entropy change [46, 47, 48, 50, 51, 52, 53, 54, 56]. Despite the limitations from our treatment of the non-local regions of the protein, the protein cavity models should provide insight into the various environments surrounding water molecules.

The protein cavity models will allow us to determine how the free energy, entropy, and enthalpy change as the number of hydrogen bond partners changes and as the size of the cavity changes. We will examine if hydrogen bond donors and hydrogen bond acceptors affect the thermodynamics differently. We will also use different potential models to determine how this influences the results.

## 3.2 The Cavity Model

In this study, we assume that the interactions of a bound water molecule with the atoms making up the cavity are the most important, while the interactions with the rest of the protein and the solvent are less important. This greatly reduces the size of the calculations from thousands of atoms to less than a hundred. In addition, we can construct environments that represent average protein environments which are only

specific to a particular protein at the local level. Following this assumption, the protein is represented only by the cavity residues, as shown in Figure 3.1.

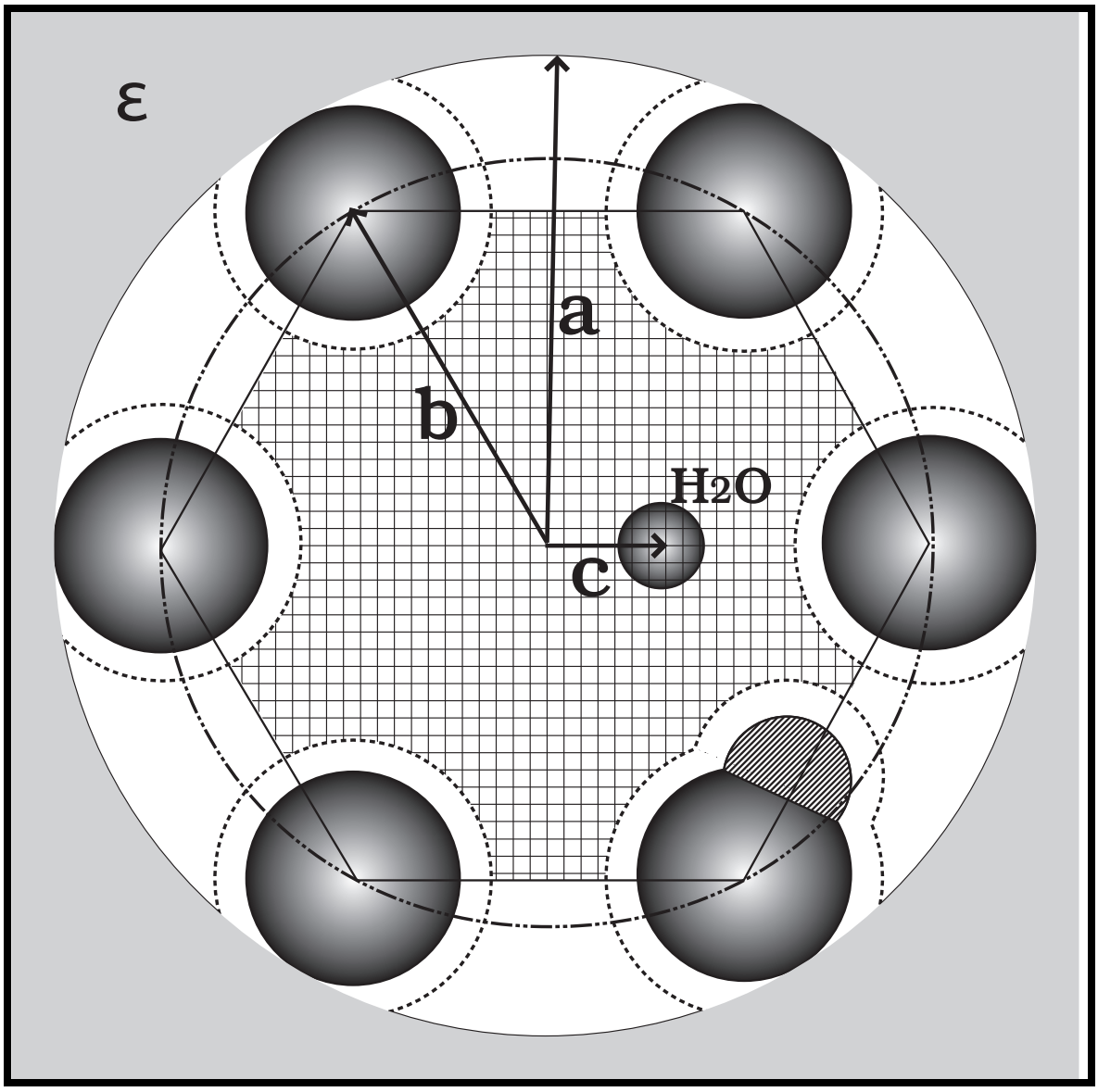


Figure 3.1: The protein cavity model: $a$ is the radius of the local region, $\mathbf{b} = \mathbf{r}_i^C - \mathbf{r}_{cm}$, and $\mathbf{c} = \mathbf{r}_W^O - \mathbf{r}_{cm}$, where $\mathbf{r}_i^C$ is the position of carbon atom in the $i$th molecule, $\mathbf{r}_W^O$ is the position of oxygen atom in water molecule, and $\mathbf{r}_{cm}$ is the center of the cavity. The grid shows the regions that are inside the cavity. Note the sphere representing the water molecule is not to scale.

In our simulation, the cavity residues are further simplified and represented by small molecules. For an absolute hydrophobic environment, surrounded by hydrophobic residues, the cavity is modeled by 12 methane molecules. In the case of alanine residues,

the charges on the atoms of methane molecule are the same as those of alanine atoms in the OPLS-AA force field. More hydrophilic cavities are constructed by replacing one or more methane molecule(s) with formaldehyde or methanol molecule(s). These mimic the alcohol containing residues serine, threonine, and tyrosine, and the carboxyl group in the amide on the peptide backbone as well as the glutamine and asparagine resides. These two molecules were chosen to give both hydrogen bond donors and acceptors. Formaldehyde acts only as a hydrogen bond acceptor and methanol can be a donor or acceptor.

To maintain the shape of the cavity, a constraint potential is applied, $u_{cav} = k_{cav} \sum_i (|\mathbf{r}_i^C - \mathbf{r}_{cm}| - R)^2$, where $k_{cav} = 3.6$ kcal/mol/Å$^2$ is the force constant, $\mathbf{r}_i^C$ is the coordinate of the carbon atom in $i$th residue molecule, the sum over $i$ is for all the cavity residues, $\mathbf{r}_{cm} = \sum_i \mathbf{r}_i^C$ is the center of the cavity, and $R$ is the designated radius of the cavity. In simulations, a radius of 3.614 Å is used, since this gives a volume similar to the volumes of protein cavities containing a single water molecule. Other values of $R$ are used as well to examine the effects of cavity size. Since methane molecules are roughly spherical, no orientational restraints are necessary, but to facilitate the formations of hydrogen bonds with the formaldehyde and methanol molecules, and to represent the rigidity of the corresponding groups in proteins, additional harmonic constraints are applied on the carboxyl or hydroxyl oxygens. This is of the form $u_{dir} = k_{dir} \sum_j (|\mathbf{r}_{cm} - \mathbf{r}_j^O| - R_j^O)^2$, where $k_{dir} = 3.0$ kcal/mol/Å$^2$ is the force constant, the sum over $j$ is over all oxygen atoms in the cavity residues, $\mathbf{r}_j^O$ is the position of the $j$th oxygen atom, and $R_j^O$ is the equilibrium distance of $j$th oxygen atom to the center of the cavity. The value of $R_j^O$ is chosen so that the residue can make hydrogen bond with the water molecule in a conformable pose. The restraints are similar to the restraints on orientational motion that the hydrogen bonding groups would feel as part of a protein. Another restraining potential is added to the water oxygen, $u_{res} = k_{res}(|\mathbf{r}_W^O - \mathbf{r}_{cm}|/R)^{12}$, ensuring that the water molecule remains inside the cavity during the course of simulation, where $k_{res} = 3.0$ kcal/mol/Å$^2$ is the force constant and $\mathbf{r}_W^O$ is the position of water oxygen.

The non-local environment (the shaded region in Figure 3.1) is treated separately, with contributions to the free energy that are added after the free energy calculations. These contributions come from the electrostatic and Lennard-Jones interactions. The electrostatic contribution is treated using a dielectric continuum model for the non-local region. The contribution to the free energy is given by the Kirkwood-Onsager equation

$$\Delta G_{KO} = -\frac{1}{2}\left[\frac{2(\epsilon - 1)}{(2\epsilon + 1)}\right]\frac{\mu^2}{a^3} \tag{3.1}$$

where $\epsilon$ is the dielectric constant of the protein, taken to be 20, $\mu$ is the dipole moment of the water molecule, and $a$ is the radius of the entire local region, which is given by the "real" radius of the cavity $\langle r \rangle$ plus 1.88 Å, the radius of the methyl group [57]. This value of the dielectric constant is in the middle of the range of dielectric constants calculated from protein simulations and estimated from $pK_a$ and other measurements [58, 59, 60, 61, 62, 63]. Our results are not too sensitive to the dielectric constant used. Using the dipole moment for the TIP4P model (2.18 Debye) and a radius of 5.80 Å, equation 3.1 gives -0.07 kcal/mol with an $\epsilon$ of 2 and -0.17 kcal/mol with an $\epsilon$ of 50. The real radius $\langle r \rangle$ is taken to be the average of the distances of the carbon atoms in the cavity residues to the center of the cavity. It should be noted that $\langle r \rangle$ is a little larger than the designed radius $R$, implying a repulsive interaction between residues, and it varies for different types of cavities even though they may have the same designed radius. For the polarizable model, the average value of $\mu$ from the simulations is used. Rigorously, because there is an interaction between the dielectric continuum and the polarizable molecules, there will be coupling between the local and non-local free energies. The dielectric continuum will polarize the water and the cavity molecules, changing the interactions with each other. This coupling is not present for non-polarizable models. Our treatment neglects this coupling, but it turns out the coupling is weak and $\Delta G_{KO}$ is sufficiently small that neglecting this coupling will not have a noticeable effect on the free energies. This correction is taken to be purely enthalpic and so contributes to the $\Delta H$ only.

The non-local Lennard-Jones contribution to the free energy is treated by adding a long-ranged correction to the energy [64, 65],

$$E_{LJ} = 2\pi\rho \sum_i w_i \int_a^\infty 4\epsilon_{oi}\left[\left(\frac{\sigma_{oi}}{r}\right)^{12} - \left(\frac{\sigma_{oi}}{r}\right)^{6}\right] r^2 dr$$
$$= 8\pi\rho \sum_i w_i \epsilon_{oi}\sigma_{oi}^3\left[\frac{1}{9}\left(\frac{\sigma_{oi}}{a}\right)^{9} - \frac{1}{3}\left(\frac{\sigma_{oi}}{a}\right)^{3}\right] \quad (3.2)$$

where $\rho$ is the average number density for proteins, $w_i$ is the probability of finding a atom of type $i$ in proteins, $\epsilon_{oi}$ and $\sigma_{oi}$ are the Lennard-Jones interaction parameters for atom of type $i$ with the oxygen atom of the water molecule, $a$ is the radius of local region, and the sum over $i$ is over all protein atom types. The probabilities $w_i$ are found from the average occurrence of each of the twenty amino acids[66] and the composition of each amino acid in terms of the 27 atom types. The equation $w_i = \sum_j P_j P(i|j)$ is used to calculate this weighting probability, where $P_j$ is the occurrence of amino acid type $j$ and $P(i|j)$ is the probability of finding atom type $i$ in amino acid type $j$, and the sum over $j$ is over all twenty amino acids. Values of $w_i$ are given in the Supplementary Information. The OPLS-AA force filed parameters are used [67] and the interaction parameters between different atom types are calculated from the combining rules ($\epsilon_{oi} = \sqrt{\epsilon_o \epsilon_i}$ and $\sigma_{oi} = \sqrt{\sigma_o \sigma_i}$). The number density of proteins is calculated from equation $\rho = \rho_m / \sum_i m_i w_i$, where $\rho_m = 1.35$ g/cm$^3$ is the average mass density of proteins [68], and $m_i$ is the mass of atom type $i$. This correction contributes to the free energy as $\Delta G_{LJ}$=$E_{LJ}$, so this correction is taken to be purely energetic.

## 3.3 Methods

A three-step method was used for the hydration free energy simulations, as described elsewhere [23, 24, 25, 30, 69]. In this method, the free energies for three processes are calculated: (1) the free energy $\Delta G_{wat}$ for transferring a water molecule from the liquid to the gas phase, (2) the localization free energy $\Delta G_{loc}$ for constraining a gas phase water

molecule in the binding site in the absence of interactions with the cavity atoms, and (3) the free energy $\Delta G_{int}$ for the localized, non-interacting water molecule to become fully interacting with the protein cavity. Gas phase for these processes means non-interacting and free to move throughout the system. The hydration free energy of the protein interior cavity, $\Delta G_{hyd}$, is the sum of these three terms, plus the non-local correction. Both a non-polarizable model with OPLS all-atom force field [67] and TIP4P water [70] and polarizable model with fluctuating charge(FQ) force field [71, 72] and TIP4P-FQ water [73] were used in our simulations.

The simulations for free energies $\Delta G_{wat}$ were performed in T, P, N ensembles at a pressure of 1 atm and temperature of 298 K. The Anderson barostat and Nosé-Hoover thermostat were employed to control the pressure and temperature [74, 75, 76]. The calculations were run in cubic boxes with 256 TIP4P or TIP4P-FQ water molecules. Lang range electrostatic interactions were treated with Ewald summation [64]. All bonds were constrained by the SHAKE algorithm with a 1 fs time step [77]. The separated-shifted scaling method was used to avoid singularities.[78, 79] The free energies were calculated by using of the thermodynamic integration with 16 $\lambda$ values ranging from 0.001 to 1.0 and 2.5 ns simulation for each $\lambda$. The enthalpy changes were calculated from $\Delta H_{wat} = \langle E + PV \rangle / N$ at point $\lambda = 1.0$, where $N$ is the number of water molecules in the primary simulation box, $E$ is the total energy, $P$ and $V$ are pressure and volume of the system. The enthalpy changes are 9.866±0.003 for TIP4P and 9.909±0.005 for TIP4P-FQ. The entropy changes were found from $-T\Delta S_{wat} = \Delta G_{wat} - \Delta H_{wat}$. In our simulations, the free energy changes, $\Delta G_{wat}$, are 6.14 ± 0.03 kcal/mol for TIP4P water and 5.97 ± 0.08 kcal/mol for TIP4P-FQ water. The corresponding entropy changes, $-T\Delta S_{wat}$, are $-3.72 \pm 0.03$ kcal/mol and −3.94 ± 0.08 kcal/mol, respectively. These results are in good agreement with the experimental values of $\Delta G_{wat}$ = 6.3 kcal/mol and $-T\Delta S_{wat}$ = −3.6 kcal/mol [80], and with other free energy calculations of 6.5 ± 0.4 kcal/mol and 6.1 ± 0.3 kcal/mol for TIP4P water [29, 81].

For the localization free energy, if a gas phase water molecule is localized in the binding site with a harmonic potential, the localization free energy is given by $\Delta G_{loc} = -kT \ln[\rho (\pi kT/k_{harm})^{3/2}]$, where $k$ is Boltzmann's constant, $T$ is temperature, $\rho$ is the bulk density of water, and $k_{harm}$ is the force constant [23, 24, 69]. A harmonic potential with a force constant equal to 3.0 kcal/mol/Å$^2$ gives a free energy $\Delta G_{loc}$ of 2.44 kcal/mol and a entropy change $-T\Delta S_{loc}$ of 1.59 kcal/mol at temperature 298 K. The value of $k_{harm}$ is chosen for numerical convenience and results for the total free energy of adding a water molecule to the cavity will be independent of $k_{harm}$. Any changes in $k_{harm}$ will have completely compensating effects on $\Delta G_{loc}$ and $\Delta G_{int}$.

To calculate the free energy $\Delta G_{int}$, the alchemical free energy calculation method was used, with a $\lambda$ scaled potential given by

$$\begin{aligned} U_{\lambda} \;=\; & k_{cav} \sum_i \left( \left| \mathbf{r}_i^C - \mathbf{r}_{cm} \right| - R \right)^2 + (1-\lambda)\, k_{harm} \left( \mathbf{r}_W^O - \mathbf{r}_{cm} \right)^2 \\ & + k_{res} \left( \frac{\mathbf{r}_W^O - \mathbf{r}_{cm}}{R} \right)^{12} + k_{dir} \sum_j (|\mathbf{r}_{cm} - \mathbf{r}_j^O| - R_j^O)^2 \\ & + \lambda \left( \sum_i 4\epsilon_{oi} \left[ \left( \frac{\sigma_{oi}}{r_{oi}} \right)^{12} - \left( \frac{\sigma_{oi}}{r_{oi}} \right)^{6} \right] + \sum_i \sum_k \frac{q_i q_k}{r_{ik}} \right) \\ & + \sum_i \sum_l \left[ \left( \frac{\sigma_{il}}{r_{il}} \right)^{12} - \left( \frac{\sigma_{il}}{r_{il}} \right)^{6} \right] + \sum_i \sum_l \frac{q_i q_l}{r_{il}}, \end{aligned} \tag{3.3}$$

where the subscript $o$ in 5th term denotes the water oxygen, the sum over $i$ and $l$ are for all cavity atoms, the sum over $j$ is for all cavity oxygen atoms, and the sum over $k$ is for water sites, except the oxygen atom (which does not have a charge in the models used here). The parameter $\lambda$ therefore scales the interactions of the hydration water molecule with cavity residues while simultaneously scaling the harmonic localizing term. As $\lambda$ goes from 0.0 to 1.0, the interactions between the water molecule and the cavity residues are turned on while the localizing harmonic potential is turned off, and a fully hydrated state is achieved at $\lambda = 1.0$. The free energy $\Delta G_{int}$ and entropy $-T\Delta S_{int}$ are evaluated from the

thermodynamic integration equations

$$\Delta G_{int} = \int_0^1 \left\langle \frac{\partial U_{\lambda}}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{3.4}$$

and

$$-T\Delta S_{int} = \frac{1}{kT} \int_0^1 \left( \left\langle U_{\lambda} \frac{\partial U_{\lambda}}{\partial \lambda} \right\rangle - \langle U_{\lambda} \rangle_{\lambda} \left\langle \frac{\partial U_{\lambda}}{\partial \lambda} \right\rangle_{\lambda} \right) d\lambda \tag{3.5}$$

where $T$ is the temperature, and

$$\frac{\partial U_{\lambda}}{\partial \lambda} = \sum_i 4\epsilon_{oi} \left[ \left( \frac{\sigma_{oi}}{r_{oi}} \right)^{12} - \left( \frac{\sigma_{oi}}{r_{oi}} \right)^{6} \right] + \sum_i \sum_k \frac{q_i q_k}{r_{ik}} - k_{harm} \left( \mathbf{r}_W^O - \mathbf{r}_{cm} \right)^2 \tag{3.6}$$

The simulations where preformed at a constant temperature of 298 K, without periodic boundary conditions. The temperature was regulated by Nosé-Hoover chain thermal bath with 3 chains for cavity residues and 3 chains for the hydration water molecule to avoid the “hot solvent/cold solute" problem [82, 83]. All bonds were constrained with SHAKE [77]. A 1 fs time step was used for nonpolarizable model and a 0.5 fs time step was used for polarizable model. For each cavity, from 14 to 18 $\lambda$ values ranging from 0.0001 to 1.0 was used, with 2.5 ns simulation for each $\lambda$ value.

Solvent accessible volumes of the cavities were calculated with a probe of radius 1.4Å. In a volume calculation, all atoms in the residue molecules except hydrogens are first mapped onto a 3D grid with spacing of $0.1 \times 0.1 \times 0.1$ Å$^3$. Volumes for these atoms are found by adding the radius of the probe molecule to the atomic radii of the atom. Subsequently, the grid sites within the convex polyhedron formed by the cavity-defining atoms are identified as cavity sites if they are not within the extended volume of the atoms. For our model cavities, the cavity-defining atoms used to calculate the volume are the carbon atoms. For comparison we also calculated volumes for proteins, which used all heavy atoms within 5 Å distance of the binding waters oxygens. The atomic group radii used in our calculation are taken from VOIDOO [84] and listed in Table 3.1.

Table 3.1: Atomic group radii used for cavity volume calculations.

| Atomic Group | CH/C | CH2 | CH3 | N/NH | O | OH | S |
|---|---|---|---|---|---|---|---|
| Radius (Å) | 1.85 | 1.925 | 2.00 | 1.75 | 1.60 | 1.65 | 2.00 |

## 3.4 Results

The cavities are changed by introducing more hydrogen bonding molecules, keeping the size roughly constant, and by changing size, keeping the molecules the same. These variations are done with both non-polarizable and polarizable models. The results for the cavities with a radius of about 3.9 Å are shown in Table 3.2, using a non-polarizable model and Table 3.3 using a polarizable model. The cavities all contain twelve total molecules (in addition to the water molecule). The composition of the cavities is indicated in the first column, where "F" stands for formaldehyde and "M" stands for methanol and the the number before each letter indicates how many methane molecules have been replaced by a formaldehyde or a methanol molecule. For example, "1F3M" is a cavity with 8 methane, 1 formaldehyde, and 3 methanol molecules and "0F0M" is made up of only 12 methane molecules. The non-local corrections to the free energies are given in the Supplementary Information. These corrections depend on the size of the cavity and the dipole moment of the water molecule, which depends on the type of cavity for the polarizable model. The corrections are small for all cavities, about -0.55 kcal/mol. In the hydrogen bond analysis, the Mancera and Buckingham definition is used in which the oxygen-oxygen distance must be less than 3.6Å and the angle between the O-H vector on the hydrogen bonding donor and the O-O vector must be between 130° and 180° [85]. For most of the cavities considered, the change in the number of hydrogen bonds is simply the number of hydrogen bonds made between the water molecule and the molecules making up the cage. It is also important to consider how the hydrogen bonds between cage molecules change as the central water is added [30]. Only for the smallest cavities and for cavities

Table 3.2: Hydration thermodynamics for cavities with different hydrogen bond forming molecules, calculated from the non-polarizable model.[a]

| | change in hydrogen bonds | $R$ Å | $\langle r \rangle$ Å | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) | $\langle \delta \mathbf{r}_W^2 \rangle$ (Å$^2$) |
|---|---|---|---|---|---|---|---|---|
| 0F0M | 0.00(0) | 3.614 | 3.931(3) | 6.08(3) | -2.17(3) | 8.25(2) | 96.6(3) | 0.36(1) |
| 1F0M | 0.91(1) | 3.614 | 3.929(4) | 4.03(3) | -0.45(4) | 4.48(4) | 88.2(6) | 0.30(1) |
| 0F1M | 0.89(1) | 3.614 | 3.946(5) | 3.42(3) | 0.20(5) | 3.22(4) | 90.2(6) | 0.33(1) |
| 2F0M | 1.86(2) | 3.614 | 3.935(4) | 2.16(3) | 1.64(7) | 0.52(7) | 80.4(4) | 0.28(1) |
| 1F1M | 1.84(4) | 3.614 | 3.946(6) | 1.85(3) | 1.61(6) | 0.24(5) | 82.2(2) | 0.28(1) |
| 0F2M | 1.84(3) | 3.614 | 3.960(4) | 1.18(3) | 1.91(6) | -0.73(6) | 83.6(6) | 0.29(2) |
| 2F1M | 2.91(1) | 3.614 | 3.939(4) | -0.66(4) | 4.3(1) | -5.0(1) | 74.5(5) | 0.29(1) |
| 1F2M | 2.88(1) | 3.614 | 3.960(3) | -1.03(4) | 4.1(1) | -5.1(1) | 76.6(5) | 0.28(2) |
| 0F3M | 2.79(2) | 3.614 | 3.983(3) | -1.38(4) | 3.92(9) | -5.30(9) | 78.7(4) | 0.27(1) |
| 2F2M | 3.88(3) | 3.614 | 3.956(2) | -1.69(5) | 7.1(1) | -8.7(1) | 68.9(3) | 0.27(3) |
| 1F3M | 3.84(1) | 3.614 | 3.978(3) | -2.50(4) | 6.0(1) | -8.5(1) | 71.0(4) | 0.27(1) |
| 0F4M | 3.78(1) | 3.614 | 3.998(2) | -3.21(4) | 5.51(8) | -8.72(8) | 73.8(3) | 0.26(1) |

[a] The letter "F" indicates a formaldehyde molecule and the letter "M" indicates a methanol molecule. The numbers before the letters indicate how many of twelve methane molecules have been replaced by the corresponding molecule.

which can hydrogen bond with themselves (that contain at least one methanol and at least one other methanol or formaldehyde molecule) are changes in cage-cage hydrogen bonds seen. For the 1F3M cavity with a designed radius $R$ equal to 3.346 Å introducing a water reduces the cage-cage hydrogen bonds by -0.34±0.03. This is the largest change in hydrogen bonds seen.

The results from Tables 3.2 and 3.3 show a strong correlation with the number of hydrogen bonds. The free energies are positive for those cavities forming less than two hydrogen bonds and negative for those with more. The entropic contribution is negative for the hydrophobic cavities and becomes larger as the number of hydrogen bonds increases. The enthalpy changes from being positive to negative as the hydrogen bonds increase. These results are consistent with earlier studies.[25, 30] The free energies, enthalpies, and entropies of hydration as a function of the number of hydrogen bonds the water makes with the cavity are shown in Fig. 3.2. This is the data for the non-polarizable model

Table 3.3: Hydration thermodynamics for cavities with different hydrogen bond forming molecules, calculated from the polarizable model.

| | change in hydrogen bonds | $R$ Å | $\langle r \rangle$ Å | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) | $\langle \delta \mathbf{r}_W^2 \rangle$ (Å$^2$) |
|---|---|---|---|---|---|---|---|---|
| 0F0M | 0.00(0) | 3.614 | 3.943(5) | 4.57(8) | -1.85(8) | 6.42(2) | 98.7(7) | 0.32(1) |
| 0F1M | 0.86(2) | 3.614 | 3.955(3) | 2.96(8) | 0.03(9) | 2.93(4) | 91.8(5) | 0.29(1) |
| 0F2M | 1.83(3) | 3.614 | 3.967(3) | 1.31(8) | 2.1(1) | -0.78(6) | 85.3(5) | 0.27(1) |
| 0F3M | 2.73(2) | 3.614 | 3.979(2) | -0.72(8) | 3.6(1) | -4.3(1) | 78.9(2) | 0.26(1) |
| 0F4M | 3.82(1) | 3.614 | 3.988(1) | -2.66(8) | 6.2(1) | -8.8(1) | 71.8(1) | 0.23(1) |

from Table 3.2. For comparison, results from previous calculations for water in protein interiors and at the protein-ligand interface are shown. We chose the systems closest to our model cavities, with cavities that contain a single water molecule, not in direct contact with other water molecules, and also not in direct contact with charged groups. We also chose systems for which the number of hydrogen bonds was well characterized. We used the BPTI and I76A mutant of barnase systems previously studied in our lab [25], for which we calculated the number of hydrogen bonds using the Mancera and Buckingham definition, as we did for the model cavities. The other systems are for the same water at the interface between HIV-1 protease and two different inhibitors [27, 28]. As mentioned in the introduction, the free energy of one of those water molecules depends on the protonation state of aspartic acid residues not in direct contact with the water. Both those values are shown. For the HIV-1 protease systems, Lu et al. reported the average number of hydrogen bonds, although they did not describe how a hydrogen bond was defined [28]. These values are given in Table 3.4. Only the study of ref [25] calculated the separate entropy and enthalpy contributions.

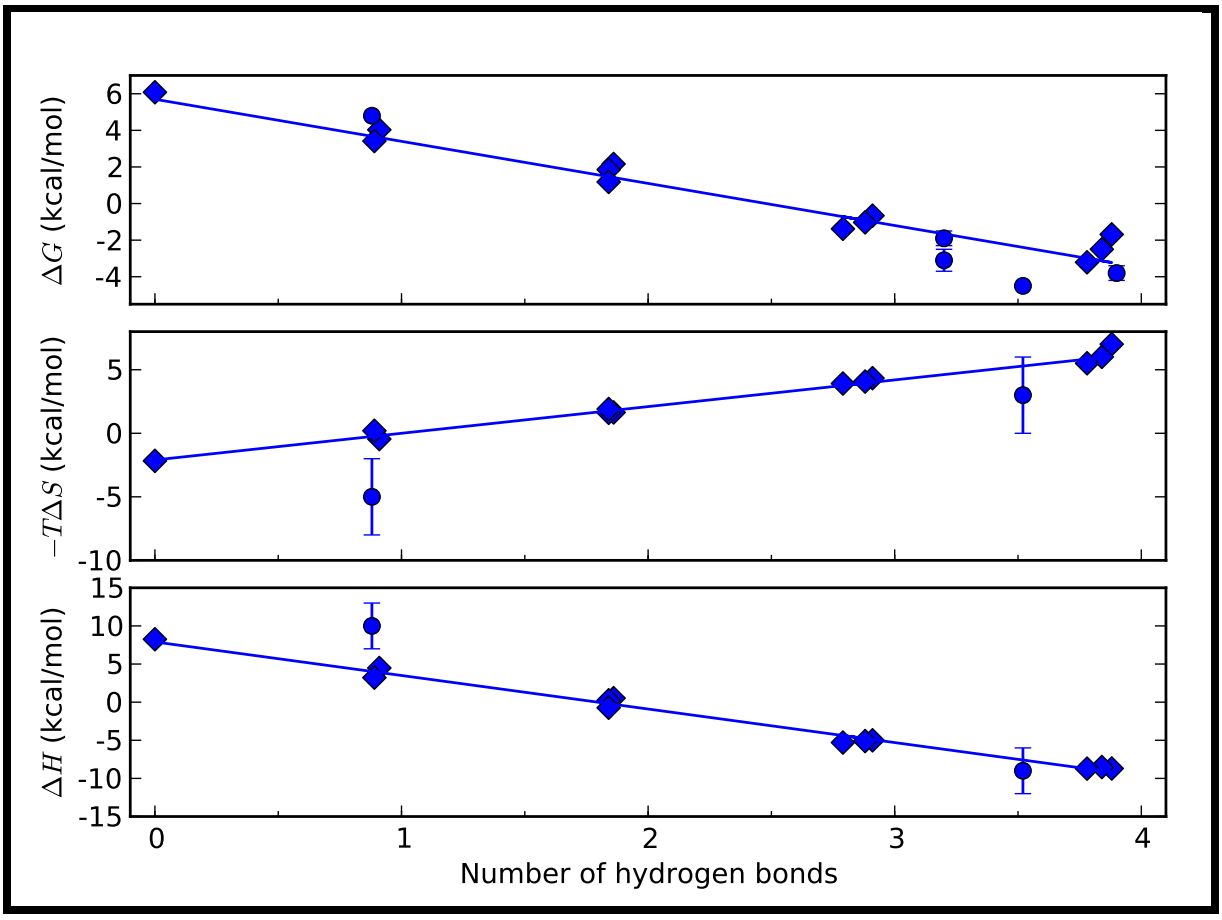


Figure 3.2: Values of $\Delta G$, $-T\Delta S$, and $\Delta H$ for the model cavities as a function of the number of hydrogen bonds formed (blue diamonds) using the non-polarizable potential model. The lines are linear fits to the data. The blue circles show previously calculated values for specific protein environments: bpti [25], barnase [25] HIV-1 protease/KNI-272 [27, 28], and HIV-1 protease/ABT-538 [28]. Except as shown, the error bars are smaller than the data symbols.

The free energies, entropies, and enthalpies all show a strong correlation with the number of hydrogen bonds. The linear fits are given by

$$\Delta G(\mathrm{x_{hb}}) = 5.7\ \mathrm{kcal/mol} - (2.3\ \mathrm{kcal/mol})\ \mathrm{x_{hb}} \tag{3.7}$$

$$-T\Delta S(\mathrm{x_{hb}}) = -2.1\ \mathrm{kcal/mol} + (2.1\ \mathrm{kcal/mol})\ \mathrm{x_{hb}} \tag{3.8}$$

$$\Delta H(\mathrm{x_{hb}}) = 7.9\ \mathrm{kcal/mol} - (4.4\ \mathrm{kcal/mol})\ \mathrm{x_{hb}} \tag{3.9}$$

Table 3.4: Hydration free energy data from previous studies, with cavity volumes calculated in the present study.

| PDB code | change in hydrogen bonds | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) |
|---|---|---|---|---|---|
| 1BRI[a] | 0.88 | 4.8(1) | -5(3) | 10(3) | 100.4 |
| 5PTI[a] | 3.52 | -4.5(1) | 3(3) | -9(3) | 61.3 |
| 1HPX[b] | 3.2 | -1.9(4) | | | 78.9 |
| 1HPX[b] | 3.2 | -3.1(6) | | | 78.9 |
| 1HXW[c] | 3.9 | -3.8(4) | | | 82.3 |

[a] Reference [25]
[b] Reference [28]
[c] Reference [27]

where $x_{nb}$ is the number of hydrogen bonds. The free energy also appears to depend not just on the number of hydrogen bonds, but, to a lesser degree, the type of hydrogen bonds. Among the cavities with roughly equal numbers of hydrogen bonds, the free energy is lowest for cavities with relatively more methanol and fewer formaldehyde molecules (Table 3.2). For methanol neighbors, the water molecule can act as both a hydrogen bond donor and a hydrogen bond acceptor (see Supplementary material) and the entropy for the cavities containing methanol is higher than those containing formaldehyde.

The results for the polarizable model are plotted on Fig. 3.3. The linear fits to the polarizable results are

$$\Delta G(x_{hb}) = 4.6\,\text{kcal/mol} - (1.9\,\text{kcal/mol})\,x_{hb} \quad (3.10)$$

$$-T\Delta S(x_{hb}) = -1.8\,\text{kcal/mol} + (2.1\,\text{kcal/mol})\,x_{hb} \quad (3.11)$$

$$\Delta H(x_{hb}) = 6.4\,\text{kcal/mol} - (4.0\,\text{kcal/mol})\,x_{hb}. \quad (3.12)$$

The results are similar to the non-polarizable model, with the largest difference for the purely hydrophobic cavity with no hydrogen bonds. The smaller $\Delta G$ is consistent other

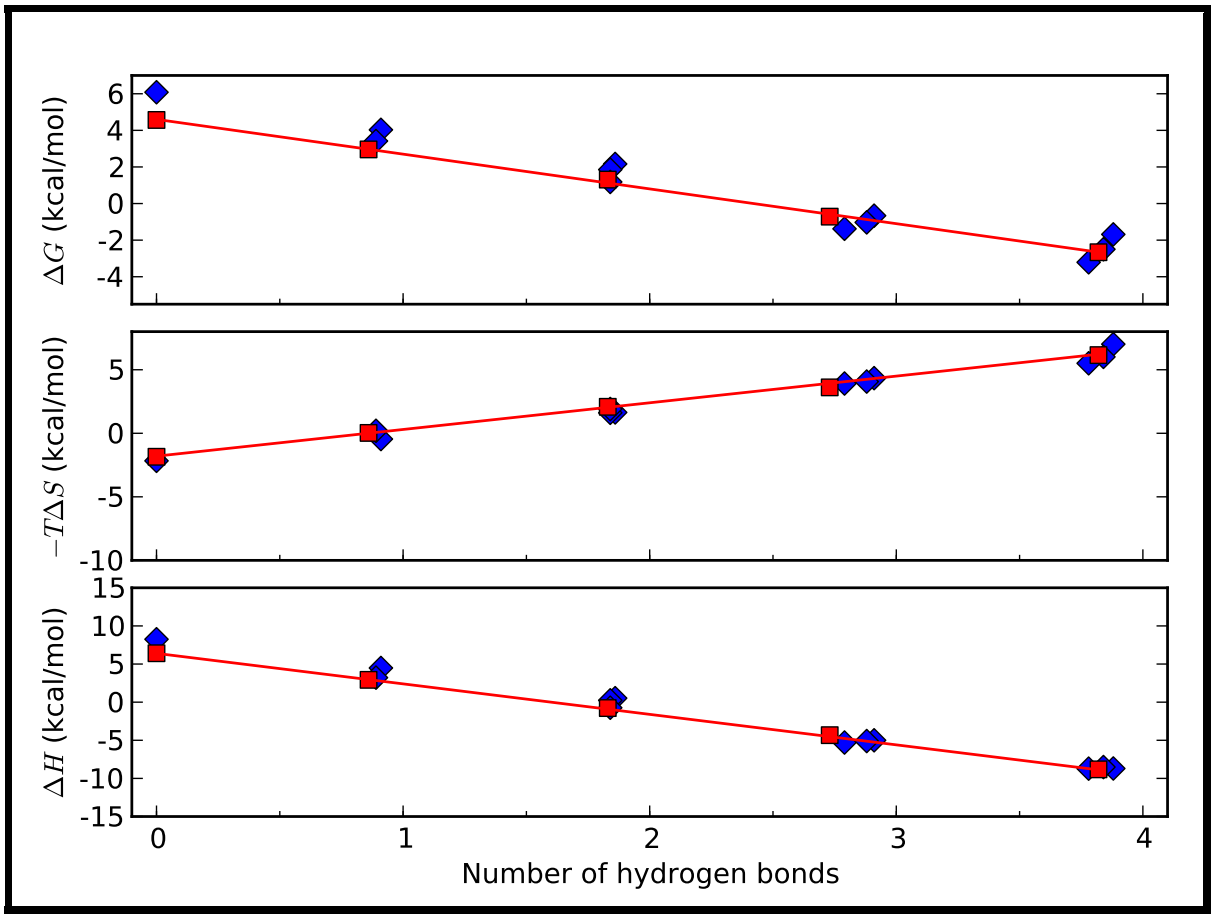


Figure 3.3: Values of $\Delta G$, $-T\Delta S$, and $\Delta H$ for the model cavities as a function of the number of hydrogen bonds formed using the non-polarizable potential model (blue diamonds) and the polarizable model (red squares). The line shows a linear fit to the data for the polarizable potential.

studies that found TIP4P-FQ water is more likely to enter a lipid bilayer than non-polarizable water models [86]. The polarizable model has a lower $\Delta G$ by about 1.5 kcal/mol for the hydrophobic cavity. This difference appears to be purely enthalpic. The entropy changes between the polarizable and non-polarizable models are very close. The average dipole moment of the water molecule increases as hydrogen bond partners are added. For the purely hydrophobic cavity, the dipole moment is 1.88 Debye, close to the gas-phase value (1.85 Debye) [73], and increases up to 2.47 Debye for the four hydrogen cavity. (The dipole moment in the liquid is 2.62 Debye.) The differences between the polarizable

Table 3.5: Hydration thermodynamics for hydrophobic cavities with different sizes, calculated from the non-polarizable model.

| | change in hydrogen bonds | $R$ Å | $\langle r \rangle$ Å | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) | $\langle \delta \mathbf{r}_W^2 \rangle$ (Å$^2$) |
|---|---|---|---|---|---|---|---|---|
| 0F0M | 0.00 | 3.100 | 3.721(1) | 8.68(3) | -2.74(4) | 11.42(3) | 69.7(1) | 0.216(5) |
| 0F0M | 0.00 | 3.400 | 3.829(1) | 6.96(3) | -2.27(4) | 9.23(2) | 83.2(1) | 0.280(6) |
| 0F0M | 0.00 | 3.614 | 3.931(3) | 6.08(3) | -2.16(3) | 8.25(2) | 96.6(2) | 0.36(1) |
| 0F0M | 0.00 | 4.100 | 4.247(4) | 5.05(3) | -2.76(4) | 7.81(4) | 142.9(6) | 0.94(5) |
| 0F0M | 0.00 | 4.300 | 4.409(3) | 4.86(3) | -3.19(6) | 8.04(5) | 169.6(5) | 1.46(8) |
| 0F0M | 0.00 | 4.500 | 4.581(4) | 4.64(3) | -3.55(5) | 8.19(4) | 199.9(7) | 2.41(8) |

Table 3.6: Hydration thermodynamics for hydrophobic cavities with different sizes, calculated from the polarizable model.

| | change in hydrogen bonds | $R$ Å | $\langle r \rangle$ Å | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) | $\langle \delta \mathbf{r}_W^2 \rangle$ (Å$^2$) |
|---|---|---|---|---|---|---|---|---|
| 0F0M | 0.00 | 3.100 | 3.734(2) | 5.71(8) | -1.03(8) | 6.74(3) | 71.4(2) | 0.158(3) |
| 0F0M | 0.00 | 3.400 | 3.844(1) | 4.98(8) | -1.51(8) | 6.49(2) | 85.4(1) | 0.234(8) |
| 0F0M | 0.00 | 3.614 | 3.943(5) | 4.57(8) | -1.85(8) | 6.42(2) | 98.7(7) | 0.319(8) |
| 0F0M | 0.00 | 4.100 | 4.261(5) | 4.08(8) | -2.76(8) | 6.83(2) | 146.0(9) | 0.76(3) |
| 0F0M | 0.00 | 4.300 | 4.414(2) | 4.02(8) | -3.22(8) | 7.23(3) | 171.2(5) | 1.14(6) |
| 0F0M | 0.00 | 4.500 | 4.586(3) | 3.93(8) | -3.65(9) | 7.59(4) | 201.9(7) | 1.87(4) |

and non-polarizable models should be most noticeable when the dipole moment is most different from the bulk value, as it is for the hydrophobic cavities.

The thermodynamic properties as a function of cavity size are given in Tables 3.5-3.8. Tables 3.5 and 3.6 give the results for various sized non-polar cavities, with results from both the non-polarizable and polarizable potential.The results for cavities with four potential hydrogen bonding partners are given in Tables 3.7 and 3.8. The size of the cavities can be quantified in a number of ways. The average of the carbon positions, $\langle r \rangle$, gives one measure of the cavity size. Cavity size can also be determined by the solvent accessible volume. Another measure of the size of the cavity is the mean-square fluctu-

Table 3.7: Hydration thermodynamics for hydrophilic cavities with different sizes, calculated from the non-polarizable model.

| | change in hydrogen bonds | $R$ Å | $\langle r \rangle$ Å | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) | $\langle \delta \mathbf{r}_W^2 \rangle$ (Å$^2$) |
|---|---|---|---|---|---|---|---|---|
| 0F4M | 3.57(4) | 3.346 | 3.897(4) | 0.62(5) | 5.8(1) | -5.2(1) | 58.1(6) | 0.22(3) |
| 0F4M | 3.74(1) | 3.413 | 3.921(2) | -1.00(4) | 6.0(1) | -7.0(1) | 61.5(4) | 0.225(6) |
| 0F4M | 3.84(1) | 3.480 | 3.944(2) | -2.09(4) | 6.0(1) | -8.1(1) | 64.8(3) | 0.22(2) |
| 0F4M | 3.83(2) | 3.547 | 3.969(4) | -2.75(4) | 6.0(1) | -8.7(1) | 68.4(6) | 0.240(3) |
| 0F4M | 3.78(1) | 3.614 | 3.998(2) | -3.21(4) | 5.51(8) | -8.72(8) | 72.8(3) | 0.256(5) |
| 0F4M | 3.65(1) | 3.682 | 4.026(3) | -3.28(4) | 5.15(8) | -8.43(8) | 77.1(3) | 0.272(8) |
| 0F4M | 3.44(1) | 3.749 | 4.063(2) | -3.20(4) | 4.64(8) | -7.84(8) | 82.8(3) | 0.307(6) |
| 0F4M | 3.18(2) | 3.817 | 4.096(4) | -2.98(4) | 4.18(7) | -7.17(7) | 88.2(6) | 0.35(1) |
| 0F4M | 2.89(2) | 3.885 | 4.135(3) | -2.69(3) | 3.63(8) | -6.31(7) | 94.5(4) | 0.42(1) |
| 0F4M | 1.76(2) | 4.158 | 4.315(3) | -1.18(3) | 1.51(5) | -2.69(5) | 124.8(5) | 0.86(3) |
| 0F4M | 1.35(1) | 4.337 | 4.453(3) | -0.34(3) | 0.41(5) | -0.74(4) | 149.2(5) | 1.38(4) |

Table 3.8: Hydration thermodynamics for hydrophilic cavity with different sizes, calculated from polarizable model.

| | change in hydrogen bonds | $R$ Å | $\langle r \rangle$ Å | $\Delta G_{hyd}$ (kcal/mol) | $-T\Delta S_{hyd}$ (kcal/mol) | $\Delta H_{hyd}$ (kcal/mol) | Volume (Å$^3$) | $\langle \delta \mathbf{r}_W^2 \rangle$ (Å$^2$) |
|---|---|---|---|---|---|---|---|---|
| 0F4M | 3.74(3) | 3.346 | 3.883(2) | 0.4(1) | 6.9(2) | -6.6(2) | 57.2(3) | 0.19(1) |
| 0F4M | 3.85(1) | 3.413 | 3.909(3) | -1.00(9) | 7.0(2) | -8.0(2) | 60.2(3) | 0.184(6) |
| 0F4M | 3.91(1) | 3.480 | 3.932(3) | -1.96(9) | 6.9(2) | -8.9(1) | 63.5(5) | 0.195(5) |
| 0F4M | 3.90(1) | 3.547 | 3.961(3) | -2.48(8) | 6.4(1) | -8.9(1) | 68.0(3) | 0.219(7) |
| 0F4M | 3.82(1) | 3.614 | 3.988(1) | -2.66(8) | 6.2(1) | -8.81(9) | 71.8(1) | 0.226(8) |
| 0F4M | 3.70(3) | 3.682 | 4.020(3) | -2.69(8) | 5.5(1) | -8.15(8) | 76.5(5) | 0.245(8) |
| 0F4M | 3.28(2) | 3.749 | 4.055(3) | -2.50(8) | 4.9(1) | -7.4(1) | 82.0(4) | 0.289(7) |
| 0F4M | 3.18(3) | 3.817 | 4.089(3) | -2.21(8) | 4.0(1) | -6.26(9) | 87.5(5) | 0.325(7) |
| 0F4M | 2.83(1) | 3.885 | 4.129(2) | -1.91(8) | 3.3(1) | -5.24(7) | 94.0(3) | 0.391(6) |
| 0F4M | 1.67(1) | 4.158 | 4.316(3) | -0.42(8) | 0.8(1) | -1.26(5) | 125.1(5) | 0.84(2) |
| 0F4M | 1.28(2) | 4.337 | 4.451(3) | 0.35(8) | -0.2(1) | 0.55(4) | 149.0(4) | 1.272(8) |

ations of the position of the oxygen atom of the water molecule. These values are shown in Tables 3.5-3.8. The dependence of $\Delta G$, $-T\Delta S$, and $\Delta H$ on the cavity volume is shown in Fig. 3.4 for the non-polarizable model and Fig. 3.5 for the polarizable model. The dependence on volume for both a purely hydrophobic (all methane) cavity and a hydrophilic (with four possible hydrogen bonding molecules) is shown. For both potential models, the volume dependence is different for the hydrophobic (diamonds) and hydrophilic (filled circles) cavities. The free energy for the hydrophobic cavities decreases with volume, while the free energy for the hydrophilic cavities has a minimum around 75 $Å^3$ and increases at larger volumes. The entropies and enthalpies are much more strongly dependent on volume for the hydrophilic cavities.

For the hydrophilic cavities, the change in the number of hydrogen bonds is dependent on the volume (see Tables 3.7 and 3.8). For small volumes, adding a water molecule breaks up cage-cage hydrogen bonds, so the gain in hydrogen bonds is less than the number of hydrogen bonds the water makes with the cage. For example, a water in the smallest cavity (with $R$=3.346 Å) makes 3.92 hydrogen bonds with the surrounding methanol molecules, but the addition of the water disrupts the methanol-methanol hydrogen bonds so only 3.57 hydrogen bonds are gained. For the cavity with $R$=3.682 Å, which has the most negative $\Delta G$, only 3.65 water-methanol hydrogen bonds are made, so this cavity is less optimal than the smaller cavity for making water-cage hydrogen bonds. In this case, no methanol-methanol hydrogen bonds are disrupted by the water molecule, so the net number of hydrogen bonds gained by the addition of the water molecule is 3.65, slightly greater than that for the smaller cavity.

For larger volumes, the number of water-methanol hydrogen bonds continues to decrease. The number of hydrogen bonds falls from 3.65 to 1.35 in the largest case (Tables 3.7). If the contributions to $\Delta G$, $-T\Delta S$, and $\Delta H$ from the hydrogen bonds are given by the linear fits of eq 3.9 or 3.12, then the effects of volume can be separated from the effects of hydrogen bond number. For example, from the $\Delta G$ for the non-polarizable model a factor

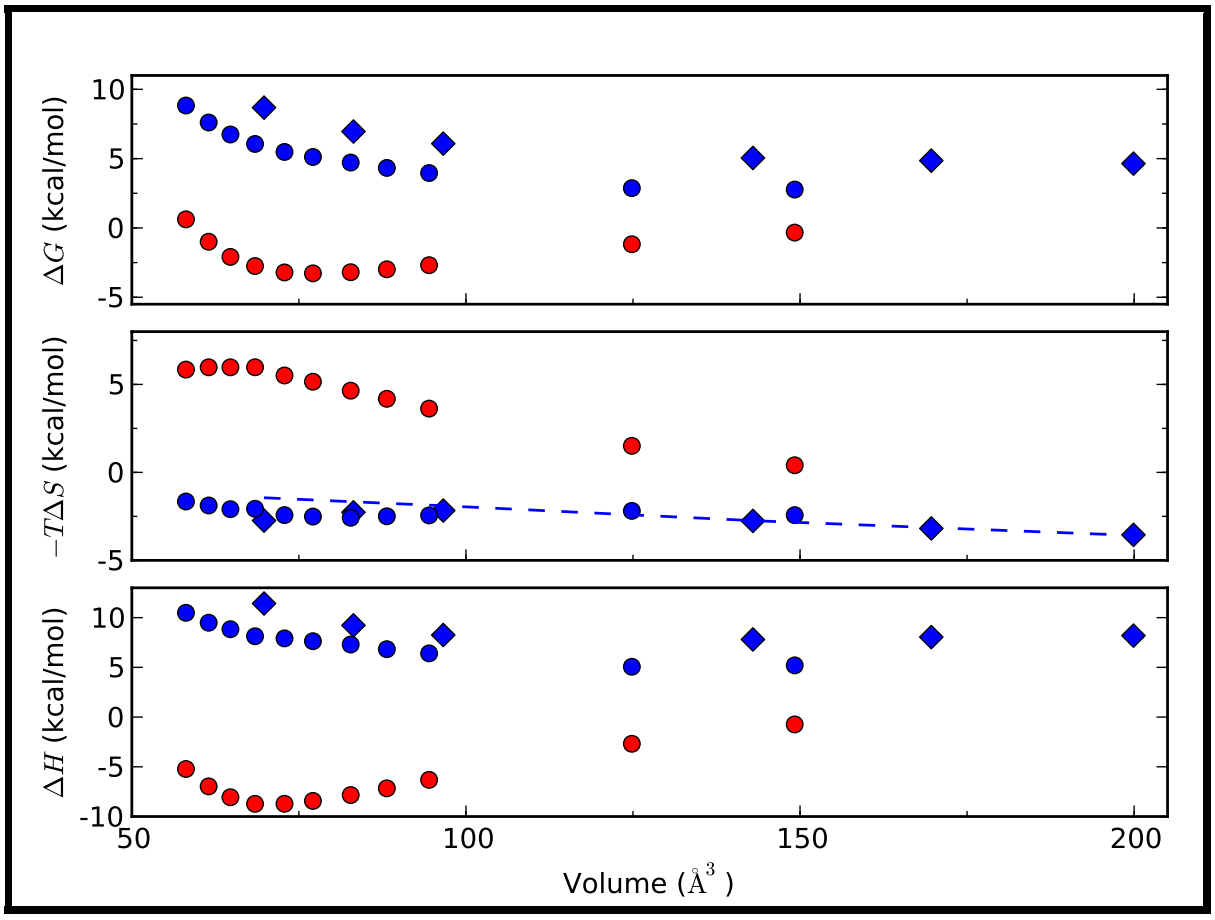


Figure 3.4: Values of $\Delta G$, $-T\Delta S$, and $\Delta H$ as a function of cavity volume for the purely hydrophobic cavities (blue diamonds), the hydrophilic cavities (red circles), and the hydrophilic cavities after subtraction of the hydrogen bond contribution (blue circles) for the non-polarizable model. The dashed line in the middle figure shows the harmonic estimate of the $-T\Delta S$ for the hydrophobic cavities.

of 2.3 kcal/mol times the number of hydrogen bonds is subtracted. Once the hydrogen bond contribution is eliminated, the results look more consistent between the hydrophilic and hydrophobic cavities (the blue circles in Figures 3.4 and 3.5). This free energy decreases as the volume increases, just like the hydrophobic cavities. The hydrogen-bond-corrected entropies are identical for the hydrophilic and hydrophobic cavities and do not have a strong volume dependence, which means that the strong volume dependence of the entropy for water in hydrophilic cavities is almost completely due the changes in hydrogen bonds.

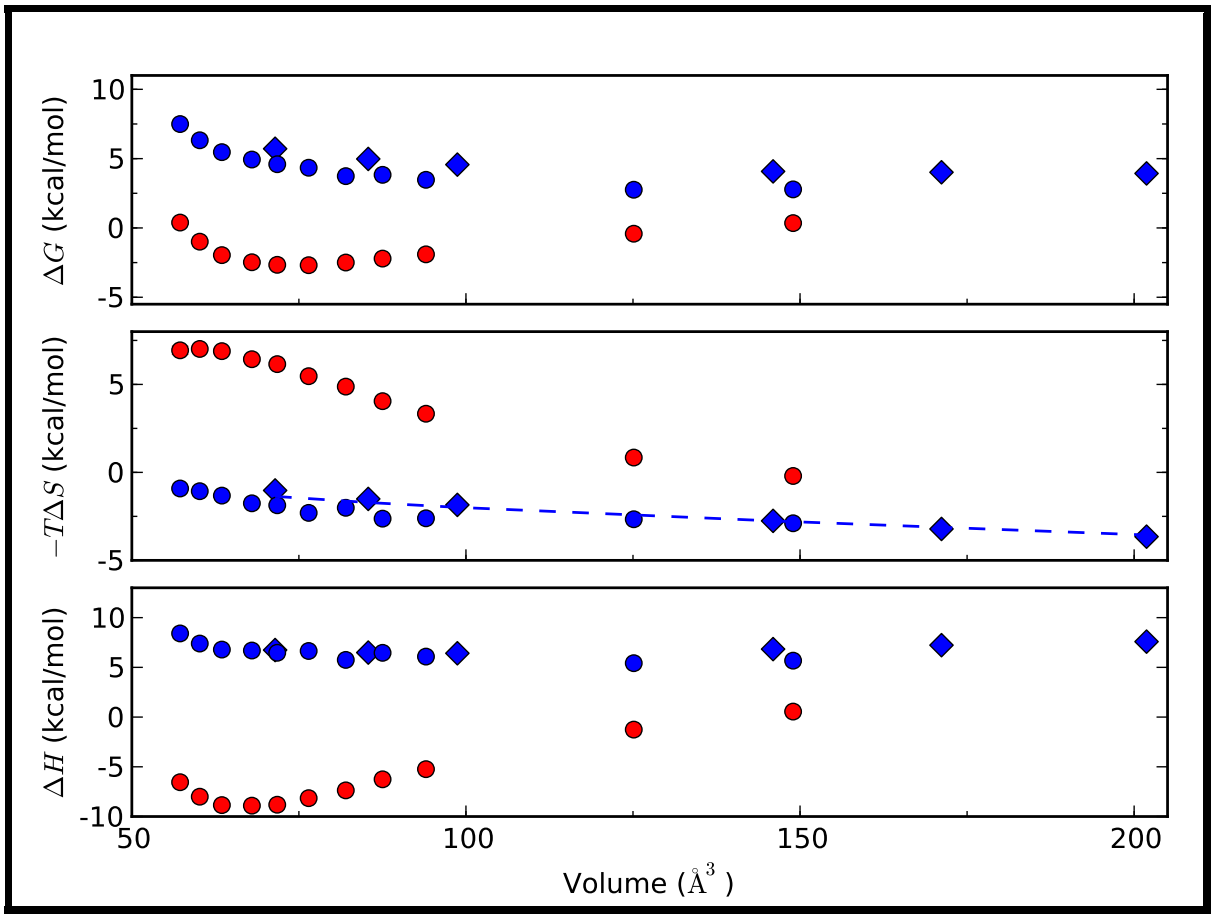


Figure 3.5: Values of $\Delta G$, $-T\Delta S$, and $\Delta H$ as a function of cavity volume for the purely hydrophobic cavities (blue diamonds), the hydrophilic cavities (red circles), and the hydrophilic cavities after subtraction of the hydrogen bond contribution (blue circles) for the polarizable model. The dashed line in the middle figure shows the harmonic estimate of the $-T\Delta S$ for the hydrophobic cavities.

The changes in volume increase the entropy (decreases $-T\Delta S$) of the hydrophobic cavities and the hydrophilic cavities (after the hydrogen bond contribution is subtracted) but only by one to two kcal/mol as the volume is increased from about 60 Å$^3$ to 200 Å$^3$. This weak dependence on volume can be understood with a harmonic model. The entropy of a harmonic oscillator is

$$S = 3k[1 - \ln(\hbar\omega/kT)] \tag{3.13}$$

where $\hbar$ is Planck's constant divided by $2\pi$. The frequency $\omega$ can be found from the fluctuations in the position of the oxygen atom in the water molecule, $\langle \delta \mathbf{r}_w^2 \rangle$ using

$$\omega^2 = 3kT/(m\langle \delta \mathbf{r}_w^2 \rangle) \tag{3.14}$$

where $m$ is the mass of the water molecule. This harmonic estimate of the entropy is shown in Fig. 3.4 and Fig. 3.5, using the $\langle \delta \mathbf{r}_w^2 \rangle$ from Tables 3.5 and 3.6. To find $-T\Delta S$, a constant (6.0 kcal/mol for the non-polarizable model and 6.2 for the polarizable model) has been added to the entropy from Eq. 3.13. This was chosen to that the harmonic estimate of the entropy agreed with entropies from the simulations in the large volume limit. Since the harmonic entropy includes only the vibrational entropy of the water molecule in the cavity, this constant includes the contributions to the entropy from everything else, including contributions from the liquid phase entropy, the translational entropy of water in the cages, and the changes in the entropy of the atoms making up the cavity. This analysis captures the volume dependence of the entropy, especially for larger volumes.

## 3.5 Conclusion

Our model cavities models are representative of cavities inside proteins, or at protein-ligand interfaces, that are large enough for a single water molecule. The thermodynamics of adding a water molecule to these model cavities agree with the results from simulations of entire proteins, and allow us to examine a large variety of local environments. The cavities are made to vary in two ways. The first is to vary the hydrophilicity by adding an carbonyl or alcohol moiety to increase the number of hydrogen bonds the water can make. These are chosen to represent neutral hydrogen bond forming atoms, like those on the protein backbone or on alcohol containing side chains. The second is to change the size of the cavity.

As the number of hydrogen bonds increases, $\Delta G$ decreases, with the enthalpic gain outweighing the entropic loss (Figures 3.2 and 3.3 and Tables 3.2 and 3.3).The free energy

decreases with the number of hydrogen bonds by about 2 kcal/mol, with the enthalpic contribution decreasing by 4 kcal/mol and the entropy contribution, $-T\Delta S$, increasing by 2 kcal/mol (see eqs 3.7-3.12). This dependence of the entropy on the number of hydrogen bonds is close to value found previously for water molecules at the protein/drug interface of 1.7 kcal/mol per hydrogen bond [30]. Also close is the theoretical estimate using the entropy of melting of ice which predicts $-T\Delta S$ should increase by 1.6 kcal/mol for each hydrogen bond [33]. In consideration of the effects of hydrogen bonds on the stability of the water, it is important to consider not only the hydrogen bonds that are made by the water molecule but the hydrogen bonds that are lost by the addition of the water. Water in cavities in which the water molecule forms four hydrogen bonds can be less stable than water those in cavities which form less hydrogen bonds, if the addition of those water disrupts less hydrogen bonds than in the other cavity. A similar conclusion was reached in our earlier study of water molecules at drug/protein interfaces [30].

The results of the calculations with different volumes for hydrophilic cavities show that the main dependence on the free energy, entropy, and enthalpy are mainly from changes in the hydrogen bonds that form. After the contributions from the hydrogen bonds are taken into account, the remaining changes are small in magnitude and similar to what is seen for the hydrophobic cavities. The small decrease in entropy as the volume increases is consistent with entropy estimated from a simple harmonic model.

# Bibliography


[1] Mark A. Williams, Julia M. Goodfellow, and Janet M. Thornton. Buried waters and internal cavities in monomeric proteins. *Protein Sci.*, 3:1224–1235, 1994.

[2] Sheldon Park and Jeffery G. Saven. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins: Struct. Funct. Bioinf.*, 60:450–463, 2005.

[3] Francis Rodier, Ranjit Prasad Bahadur, Pinak Chakrabarti, and Joël Janin. Hydration of protein-protein interfaces. *Proteins: Struct. Funct. Bioinf.*, 60:36–45, 2005.

[4] Yipin Lu, Renxiao Wang, Chao-Yie Yang, and Shaomeng Wang. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.*, 47:668–675, 2007.

[5] Robert Renthal. Buried water molecules in helical transmembrane proteins. *Protein Sci.*, 17:293–298, 2008.

[6] Szilvia Szep, Sheldon Park, Eric T. Boder, Gregory D. Van Duyne, and Jeffery G. Saven. Structural coupling between FKBP 12 and buried water. *Proteins: Struct. Funct. Bioinf.*, 74:603–611, 2009.

[7] Patrick Y. S. Lam, Prabhakar K. Jadhav, Charles J. Eyermann, C. Nicholas Hodge, Yu Ru, Lee T. Bacheler, James L. Meek, Michael J. Otto, Marlene M. Rayner, Y. Nancy Wong, Chong-Hwan Chang, Patricia C. Weber, David A. Jackson, Thomas R. Sharpe, and Susan Erickson-Viitanen. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, 263:380–384, 1994.

[8] Patrick R. Connelly, Robert A. Aldape, Frank J. Bruzzese, Stephen P. Chambers, Matthew J. Fitzgibbon, Mark A. Fleming, Susumu Itoh, David J. Livingston, Manuel A. Navia, John A. Thomson, and Keith P. Wilson. Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc. Natl. Acad. Sci. USA*, 91: 1964–1968, 1994.

[9] P. C. Weber, M. W. Pantoliano, D. M. Simons, and F. R. Salemme. Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.*, 116: 2717–2724, 1994.

[10] J. V. N. Vara Prasad, Elizabeth A. Lunney, Donna Ferguson, Peter J. Tummino, J. Ronald Rubin, Eric. L. Reyner, Barbra H. Stewart, Robert J. Guttendorf, John M. Domagala, Leonid I. Suvorov, Sergei V. Gulnik, Igor A. Topol, T. N. Bhat, and John W. Erickson. HIV protease inhibitors possessing a novel, high-affinity, and achiral $p_1$'/$p_2$'

ligand with a unique pattern of *i*n vitro resistance. importance of a conformationally-restricted template in the design of enzyme inhibitors. *J. Am. Chem. Soc.*, 117:11070–11074, 1995.

[11] William E. Royer, Jr., Animesh Pardanani, Quentin H. Gibson, Eric S. Peterson, and Joel M. Friedman. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. USA*, 93:14526–14531, 1996.

[12] Geoffrey A. Holdgate, Alan Tunnicliffe, Walter H. J. Ward, Simon A. Weston, Gina Rosenbrock, Peter T. Barth, Ian W. F. Taylor, Richard A. Pauptit, and David Timms. The entropic penalty of ordered water acccounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: A thermodynamic and crystallographic study. *Biochemistry*, 36:9663–9673, 1997.

[13] Christian Heiss, Maris Laivenieks, J. Gregory Zeikus, and Robert S. Phillips. The stereospecificity of secondary alcohol dehydrogenase from *t*hermoanaerobacter ethanolicus is partially determined by active site water. *J. Am. Chem. Soc.*, 123: 345–346, 2001.

[14] Christopher Clarke, Robert J. Woods, John Gluska, Alan Cooper, Margaret A. Nutley, and Geert-Jan Boons. Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.*, 123:12238–12247, 2001.

[15] Mark J. Snider and Richards Wolfenden. Site-bound water and the shortcomings of a less than perfect transition state analogue. *Biochemistry*, 40:11364–11371, 2001.

[16] Karen M. Peterson, K. V. Gopalan, Andreas Nandy, and D. K. Srivastava. Influcence of Glu-376 → Gln mutation on enthalpy and heat capacity changes for the binding of slightly altered ligands to medium chain acyl-CoA dehydrogenase. *Protein Sci.*, 10: 1822–1834, 2001.

[17] David L. Zechel, Alisdair B. Boraston, Tracey Gloster, Catherine M. Boraston, James M. Macdonald, D. Matthew G. Tilbrook, Robert V. Stick, and Gideon J. Davies. Iminosugar glycosidase inhibitors:â‰ structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to Î²-glucosidases. *J. Am. Chem. Soc.*, 125:14313–14323, 2003.

[18] Tracey M. Gloster, Spencer J. Williams, Shirley Roberts, Chris A. Tarling, Jacqueline Wicki, Stephen G. Withers, and Gideon J. Davies. Atomic resolution analyses of the binding of xylobiose-derived deoxynojirimycin and isofagomine to xylanase Xyn10A. *Chem. Commun.*, pages 1794–1795, 2004.

[19] Daniel Lafitte, Valérie Lamour, Philippe O. Tsvetkov, Alexander A. Makarov, Michel Klich, Pierre Deprez, Dino Moras, Claudette Briand, and Robert Gilli. DNA gyrase interaction with coumarin-based inhibitors: The role of the hydroxybenzoate isopentenyl moiety and the 5'-methyl group of the noviose. *Biochemistry*, 41:7217–7223, 2002.

[20] Antonio Hernandez Daranas, Hiroki Shimizu, and Steve W. Homans. Thermodynamics of binding of d-galactose and deoxy derivatives thereof to the l-arabinose-binding protein. *J. Am. Chem. Soc.*, 126:11870–11876, 2004.

[21] Rebecca C. Wade, Michael H. Mazor, J. Andrew McCammon, and Florante A. Quiocho. Hydration of cavities in proteins: A molecular dynamics approach. *J. Am. Chem. Soc.*, 112:7057–7059, 1990.

[22] Rebecca C. Wade, Michael H. Mazor, J. Andrew McCammon, and Florante A. Quiocho. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers*, 31:919–931, 1991.

[23] Li Zhang and Jan Hermans. Hydrophilicity of cavities in proteins. *Proteins: Struct. Funct. Bioinf.*, 24:433–438, 1996.

[24] Benoît Roux, Mafalda Nina, Régis Pomès, and Jeremy C. Smith. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.*, 71:670–681, 1996.

[25] L. Renee Olano and Steven W. Rick. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.*, 126:7991–8000, 2004.

[26] Volkhard Helms and Rebecca C. Wade. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: A molecular dynamics study. *Biophys. J.*, 69:810–824, 1995.

[27] Donald Hamelberg and J. Andrew McCammon. Standard free energy of releasing a localized water molecule from the binding of pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc.*, 126:7683–7689, 2004.

[28] Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.*, 128: 11830–11839, 2006.

[29] Caterina Barillari, Justine Taylor, Russell Viner, and Jonathan W. Essex. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.*, 129:2577–2587, 2007.

[30] Hongtao Yu and Steven W. Rick. Free energies and entropies of water molecules at the inhibitor–protein interface of DNA gyrase. *J. Am. Chem. Soc.*, 131:6608–6613, 2009.

[31] Julien Michel, Julian Tirado-Raves, and William L. Jorgensen. Energetics of displacing water molecules from protein binding sites: Consequences for ligand optimization. *J. Am. Chem. Soc.*, 131:15403–15411, 2009.

[32] Jack D. Dunitz. The entropic cost of bound water in crystals and biomolecules. *Science*, 264:670, 1994.

[33] Alan Cooper. Heat capacity effects in protein folding and ligand binding: A re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.*, 115:89–97, 2005.

[34] Zheng Li and Themis Lazaridis. Thermodynamics contribution of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.*, 125:6636–6637, 2003.

[35] Zheng Li and Themis Lazaridis. The effect of water displacement on binding thermodynamics: Concanavalin A. *J. Phys. Chem. B*, 109:662–670, 2005.

[36] Zheng Li and Themis Lazaridis. Thermodynamics of buried water clusters at a protein-ligand binding interface. *J. Phys. Chem. B*, 110:1464–1475, 2006.

[37] Subramanian Vaitheeswaran, Hao Yin, Jayendran C. Rasaiah, and Gerhard Hummer. Water clusters in nonpolar cavities. *Proc. Natl. Acad. Sci. USA*, 101:17002–17005, 2004.

[38] Brian W. Matthews and Lijun Liu. A review about nothing: Are apolar cavities in protein really empty? *Protein Sci.*, 18:494–502, 2009.

[39] Elizabeth Barratt, Richard J. Bingham, Daniel J. Warner, Charles A. Laughton, Simon E. V. Phillips, and Steve W. Homans. Van der waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.*, 127:11827–11834, 2005.

[40] Tom Young, Robert Abel, Byungchan Kim, Bruce J. Berne, and Richard A. Friesner. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. USA*, 104:808–813, 2007.

[41] Johan Qvist, Monika Davidovic, Donald Hamelberg, and Bertil Halle. A dry ligand-binding cavity in a solvated protein. *Proc. Natl. Acad. Sci. USA*, 105:6296–6301, 2008.

[42] Jeffrey Ewell, Bruce C. Gibb, and Steven W. Rick. Water inside a hydrophobic cavitand molecule. *J. Phys. Chem. B*, 112:10272–10279, 2008.

[43] Stefan Fischer and Chandra S. Verma. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. USA*, 96:9613–9615, 1999.

[44] Yi Mao, Mark A Ratner, and Martin F. Jarrold. One water molecule stiffens a protein. *J. Am. Chem. Soc.*, 122:2950–2951, 2000.

[45] Stefan Fischer, Jeremy C. Smith, and Chandra S. Verma. Dissecting the vibrational entropy change on protein/ligand binding: Burial of a water molecule in bovine pancreatic trypsin inhibitor. *J. Phys. Chem. B*, 105:8050–8055, 2001.

[46] Neil A. Farrow, Ranjith Muhandiram, Alex U. Singer, Pascal Steven M., Cyril M. Kay, Gerry Gish, Steven E. Shoelson, Tony Pawson, Julie D. Forman-Kay, and E. Kay, Lewis. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by $^{15}$N NMR relaxation. *Biochemistry*, 33:5984–6003, 1994.

[47] James T. Stivers, Chitrananda Abeygunawardana, Albert S. Mildvan, and Christian P. Whitman. $^{15}$N NMR relaxation studies of free and inhibitor-bound 4-oxalocrotonate tautomerase: Backbone dynamics and entropy changes of an enzyme upon inhibitor binding. *Biochemistry*, 35:16036–16047, 1996.

[48] Liping Yu, Chang-Xi Zhu, Yuk-Ching Tse-Dinh, and Stephen W. Fesik. Backbone dynamics of the C-terminal domain of *escherichia coli* topoisomerase I in the absence and presence of single-stranded DNA. *Biochemistry*, 35:9661–9666, 1996.

[49] Donald K. Phelps, Peter J. Rossky, and Carol B. Post. Influence of an antiviral compound on the temperature dependence of viral protein flexibility and packing: a molecular dynamics study. *J. Mol. Biol.*, 276:331–337, 1998.

[50] Peng Yuan, Vincent P. Marshall, Gary L. Petzold, Roger A. Poorman, and Brian J. Stockman. Dynamics of stromelysin/inhibitor interactions studied by 15N NMR relaxations measurements: Comparison of ligand binding to the S1-S3 and S1'-S3' subsites. *J. Biomol. NMR*, 15:55–64, 1999.

[51] Martin J. Stone. NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc. Chem. Res.*, 34:379–388, 2001.

[52] Sunggoo Yun, Do Soo Jang, Do-Hyung Kim, Kwan Yong Choi, and Hee Cheon Lee. $^{15}$N NMR relaxation studies of backbone dynamics in free and steroid-bound $\Delta^3$-3-ketosteroid isomerase from *pseudomonas testosteroni*. *Biochemistry*, 40:3967–3973, 2001.

[53] Richard J. Bingham, John B. C. Findlay, Shih-Yang Hsieh, Arnout P. Kalverda, Alexandra Kjellberg, Chiara Perazzolo, Simon E. V. Phillips, Kothandaraman Seshadri, Chi H. Trinh, W. Bruce Turnbull, Geoffrey Bodenhausen, and Steve W. Homans. Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J. Am. Chem. Soc.*, 126: 1675–1681, 2004.

[54] Christopher A. MacRaild, Antonio Hernández Daranas, Agnieszka Bronowska, and Steve W. Homans. Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein. *J. Mol. Biol.*, 368:822–832, 2007.

[55] Chia-en A. Chang, Wei Chen, and Michael K. Gilson. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. USA*, 104:1534–1539, 2007.

[56] Benjamin J. Killian, Joslyn Yudenfreund Kravitz, Sandeep Somani, Paramita Dasgupta, Yuan-Ping Pang, and Michael K. Gilson. Configurational entropy in protein-peptide binding: Computational study of Tsg101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide. *J. Mol. Biol.*, 389:315–335, 2009.

[57] Jerry Tsai, Robin Taylor, Cyrus Chothia, and Mark Gerstein. The packing density in proteins: Standard radii and volumes. *J. Mol. Biol.*, 290:253–266, 1999.

[58] Michael K. Gilson and Barry H. Honig. The dielectric constant of a folded protein. *Biopolymers*, 25:2097–2119, 1986.

[59] Paul E. Smith, Roger M. Brunne, Alan E. Mark, and Wilfred F. van Gunsteren. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. *J. Phys. Chem.*, 97:2009–2014, 1993.

[60] Thomas Simonson and Charles L. Brooks III. Charge screening and the dielectric constant of proteins: Insights from molecular dynamics. *J. Am. Chem. Soc.*, 118: 8452–8458, 1996.

[61] Claudia N. Schutz and Arieh Warshel. What are the dielectric "constants" of proteins and how to validate electrostatic models. *Proteins: Struct. Funct. Bioinf.*, 44:400–417, 2001.

[62] Jed W. Pitera, Michael Falta, and Wilfred F. van Gunsteren. Dielectric properties of proteins from simulation: The effect of solvent, ligands, pH, and temperature. *Biophys. J.*, 80:2546–2555, 2001.

[63] Carolyn A. Fitch, Daniel A. Karp, Kelly K. Lee, Wesley E. Stites, Eaton E. Lattman, and García-Moreno E. Bertrand. Experimental $pK_a$ values of buried residues: Analysis with continumm methods and role of water penetration. *Biophys. J.*, 82:3289–3304, 2002.

[64] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, Oxford, 1987.

[65] Michael R. Shirts, David L. Mobley, John D. Chodera, and Vijay S. Pande. Accurate and effecient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B*, 111:13052–13063, 2007.

[66] Voet Donald, Judith G. Voet, and Pratt Charlotte W. *Fundamentals of Biochemistry*. John Wiley & Sons, New York, 2002.

[67] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[68] Hannes Fischer, Igor Polikarpov, and Aldo F. Craievich. Average protein density is a molecular-weight-dependent function. *Protein Sci.*, 13:2825–2828, 2004.

[69] Michael K. Gilson, James A. Given, Bruce L. Bush, and J. Andrew McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, 72:1047–1069, 1997.

[70] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[71] Sandeep Patel and Charles L. Brooks III. A nonadditive methanol force field: Bulk liquid and liquid-vapor interfacial properties via molecular dynamics simulations using a fluctuating charge model. *J. Chem. Phys.*, 122:024508, 2005.

[72] Steven W. Rick and B. J. Berne. Free energies of the hydrophobic interaction from molecular dynamics simulations: The effects of solute and solvent polarizability. *J. Phys. Chem. B*, 101:10488–10493, 1997.

[73] Steven W. Rick, Steven J. Stuart, and B. J. Berne. Dynamical fluctuating charge force field: Application to liquid water. *J. Chem. Phys.*, 101:6141–6156, 1994.

[74] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[75] Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.

[76] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[77] Jean-Paul Ryckaert, Glovanni Ciccotti, and Herman J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[78] M. Zacharias, T. P. Straatsma, and J. A. McCammon. Separation-shifted scaling, a new scaling method for lennard-jones interaction in thermodynamic intergration. *J. Chem. Phys.*, 100:9025–9031, 1994.

[79] Steven W. Rick and B. J. Berne. The aqueous solvation of water: A comparison of continuum methods with molecular dynamics. *J. Am. Chem. Soc.*, 116:3949–3954, 1994.

[80] A. Ben-Naim and Y. Marcus. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.*, 81:2016–2027, 1984.

[81] William L. Jorgensen, James F. Blake, and J. Kathleen Buckner. Free energy of TIP4P water and the free energies of hydration of $CH_4$ and $Cl^-$ from statistical perturbation theory. *Chem. Phys.*, 129:193–200, 1989.

[82] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nosé-hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97:2635–2643, 1992.

[83] Ailan Cheng and Kenneth M. Merz, Jr. Application of the nosé-hoover chain algorithm to the study of protein dynamics. *J. Phys. Chem.*, 100:1927–1937, 1996.

[84] Gerard J. Kleywegt and T. Alwyn Jones. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst. D*, 50:178–185, 1994.

[85] R. L. Mancera and A. D. Buckingham. Temperature effects on the hydrophobic hydration of ethane. *J. Phys. Chem.*, 99:14632–14640, 1995.

[86] Joseph E. Davis, Obaidur Rahaman, and Petal Sandeep. Molecular dynamics simulations of a DMPC bilayer using nonadditive interaction models. *Biophys. J.*, 96: 385–402, 2009.

# Chapter 4

# Bound Water and Protein Structure and Flexibility

## 4.1 Introduction

The intimate relationship between protein function and its structure and flexibility has long been recognized [1, 2, 3]. Bound water molecules are frequently found in protein crystal structures and sometimes conserved in proteins belonging to the same homologous family [4, 5]. These water molecules are usually involved in the mediating of protein-protein and protein-ligand interactions and may contribute to their structural stability and conformational flexibility. In this study, we preformed Molecular Dynamics simulations for proteins with bound water included in and excluded from the water-binding sites, and the influences of the bound water molecules on the protein structure and flexibility are investigated.

## 4.2 Methods

### 4.2.1 Simulated Systems and Structure Preparation

Four proteins were simulated in our study, the wild-type bovine pancreatic trypsin inhibitor (BPTI, PDB entry 5PTI [6]), the wild-type hen egg white lysozyme (HEWL, PDB entry 4LZT [7]), and two variants of the wild-type Staphylococcal nuclease (SNase), PHS (PDB entry 1YE8 [8]) and PHS/V66E [9]. PHS is a hyperstable form of wild-type SNase with

three mutations (P117G, H124L, and S128A); PHS/V66E is the mutant of PHS with Val-66 replaced by a glutamic acid. Two separate simulations were performed for each protein, differing in the hydration states of protein cavities. In the first simulation, all bound water molecules are retained in the structure preparation step, and no perturbation was added during the course of simulation. In the other simulation, a repulsive potential applied to all solvent molecules at each water-binding site was used to prevent the solvent from entering into these regions. The repulsive potential was taken as the repulsive term of the Lennard-Jones interaction. Four bound water molecules are well identified in the 5PTI structure (WAT111, WAT112, WAT113 and WAT122) [6]. Among them, water WAT122 is contained in an isolated cavity, while the remaining three are located in a channel-like cavity forming a water cluster (red spheres in Figure 4.1(a)). In the 4LZT structure of HEWL [7], four water molecules (WAT1001, WAT1004, WAT1007 and WAT1015) are observed to be buried in a large cavity in the "hinge" region between two domains, and one water molecule (WAT1008) is located in the loop region of the smaller domain (Figure 4.1(b)). Beside these five completely buried water molecules, two other water molecules (WAT1003 and WAT1073) are also found to be partially buried at the surface (green spheres in Figure 4.1(b)). The structure of PHS has several bound water molecules, of which WAT202, WAT203, and WAT211 are deeply buried water molecules and are inaccessible to the solvent (red spheres in Figure 4.1(c)), while water WAT201, WAT214, WAT242 and WAT223 are partially buried water molecules (green spheres in Figure 4.1(c)). All these water molecules are conserved in the PHS/V66E and wild-type Snase (1STN [10]) structures except for WAT242, which is only observed in the PHS structure. The V66E mutation creates a cavity in the protein interior and introduces two more water molecules (WAT223 and WAT220 in Figure 4.1(d)). These two water molecules are next to the Glu-66 side chain and form portion of a chain of four water molecules reaching to the exterior of the protein. In the simulations of proteins without bound water, both the completely buried and partially buried water molecules are

repelled from the HEWL and PHS interiors, while only the deeply buried water molecules are excluded from the BPTI and PHS/V66E protein interiors (Figure 4.1).

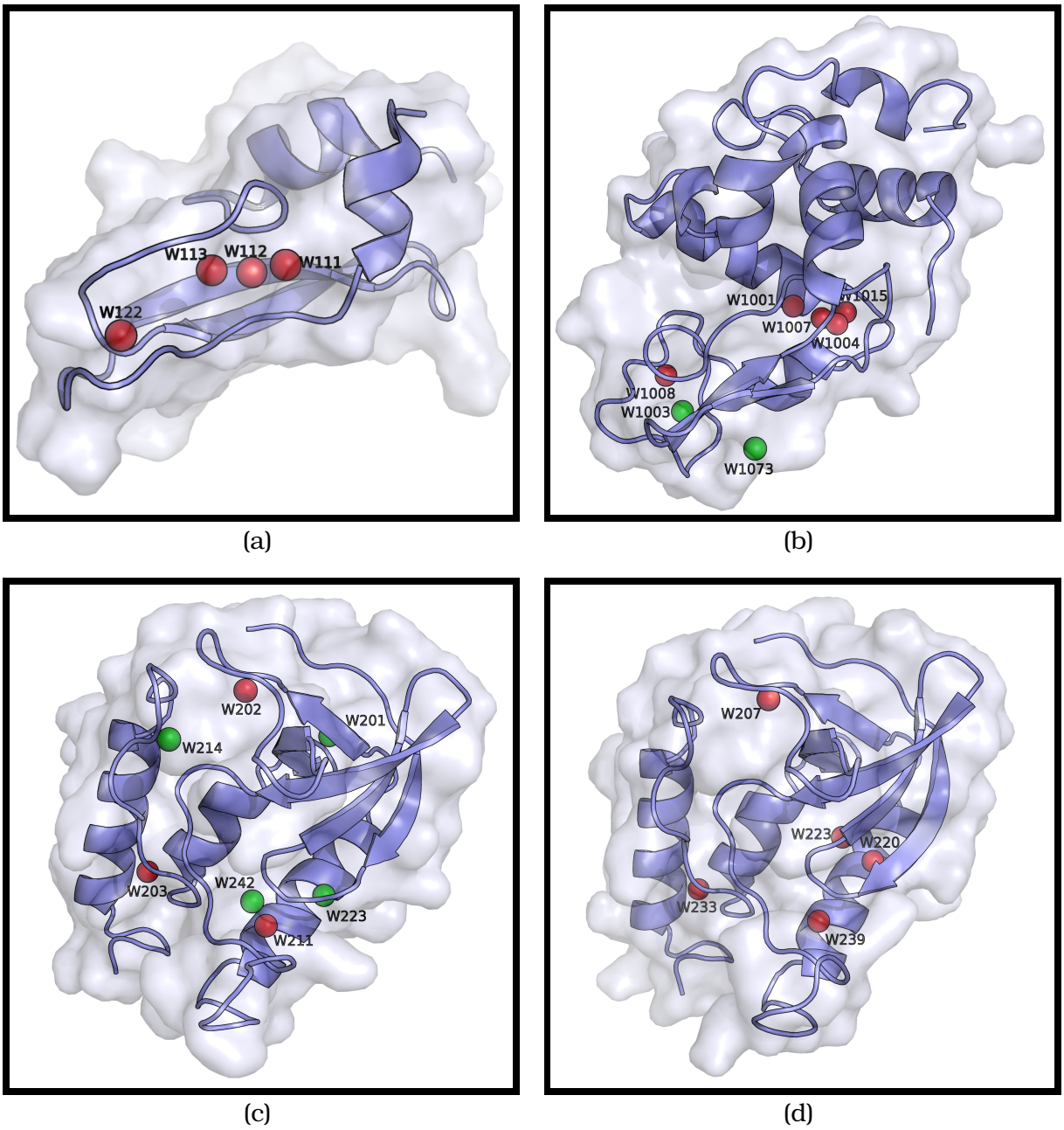


Figure 4.1: Crystal Structures of 4 proteins simulated in this study, the bound water molecules are shown as spheres. The red spheres represent deeply buried water molecules, and green ones are water molecules close to the surfaces. The solvent accessible surfaces(Connolly surface) for each protein is shown in light bllue. (a) BPTI, (b) HEWL, (c) PHS, (d) PHS/V66E

Since all simulations were carried out at physiological (neutral) pH, care was taken while assigning the charged states of histidine residues. In PHS and PHS/V66E, each protein contains three histidines: HIS-8, HIS-46, HIS-121. Their corresponding experimental $pK_a$ values are 6.53±0.02, 5.90±0.02 and 5.31±0.02 in PHS [11]. Although we found no experimental $pK_a$ data currently available for histidines in PHS/V66E, the $pK_a$ values for them in the wild-type and a series of mutants of SNase are in the range of 5.3 to 6.5 [11, 12]. Therefore, it is reasonable to presume that the $pK_a$s for these three histidines in PHS/V66E are also in this range. In our simulations, all histidines in both PHS and PHS/V66E proteins were assumed to be uncharged and were assigned the following protonation states: HID-8, HIE-46, HID-121, where HID is the N$\delta$ tautomer and HIE is the N$\epsilon$ tautomer. This choice is the same as that used in two other PHS/V66E simulations [13, 14], but different from that used in a $pK_a$ calculation with continuum method [11]. It should be noted that, Although the $pK_a$ values for all aspartic acids and glutamic acids measured experimentally in a number of PHS mutants are in the range of 2.1–4.0 and 2.5–4.6 (with one exception, the $pK_a$ for ASP-21 is about 6.5), respectivelly [15], The $pK_a$ of GLU-66 in PHS/V66E is estimated to be 8.8 [16]. In our simulations, all the ASP and GLU residues were assumed to be in their deprotonationed states except GLU-66 in PHS/V66E, which was in the neutral state and the hydrogen was added on the OE2 atom. There is one histidine residue (HIS-15) in the HEWL protein. The $pK_a$ value for this histidine is about 5.5 [17, 18]. This residue was assumed to be in the HID-15 protonation state based on the analysis of hydrogen bonding pattern in 4LZT crystal structure. The addition of the hydrogen on the N$\delta$ atom can make this histidine accepts one proton from the hydroxyl group in THR-89 and donates one proton to the carboxyl group in ALA-11.

All the simulated systems were prepared with the *tleap* and *sander* modules in the Amber 9 suite of programs [19]. In the first step of structure preparation, residues with alternative conformations were identified and only "A" conformation was retained. All water molecules found in the PDB files were also retained except for BPTI, in which only

the bound water molecules were retained. The well prepared protein structure was then loaded into *tleap* to add hydrogen atoms and counter ions to neutralize the protein charge. For PHS and PHS/V66E, only atomic coordinates from residue 7 to 141 are available in the crystal structures. Acetyl (ACE) beginning groups and N-methylamine ending groups were added to these two proteins to represent the terminal amino acids. After the addition of ending groups, short energy minimizations of 200 steps were performed to remove bad contacts. In the last step, the protein was immersed into a rectangular water box with a minimum distance of 9 Å between the protein and the edge of the box. For the simulation of protein devoid of bound water molecules, all identified water molecules were removed after solvating protein into the water box. The details of the systems simulated in this study are summarized in Table 4.1.

Table 4.1: Details of simulated systems

| Protein | Excluded Water | residues | Protonation state | $Cl^-$ | $H_2O$ | Time (ns) |
|---|---|---|---|---|---|---|
| BPTI(W) | | 1–58 | | 6 | 3628 | 31 |
| BPTI(NW) | 122,111,112,113 | 1–58 | | 6 | 3624 | 13 |
| HEWL(W) | | 1–129 | HID-15 | 8 | 6135 | 20 |
| HEWL(NW) | 1001,1003,1004,1007, 1008,1015,1073 | 1–129 | HID-15 | 8 | 6128 | 17 |
| PHS(W) | | 7–141 | HID-8,HIE-46,HID-121 | 9 | 5982 | 23 |
| PHS(NW) | 201,202,203,211, 214,223,242 | 7–141 | HID-8,HIE-46,HID-121 | 9 | 5975 | 20 |
| PHSV66E(W) | | 7–141 | HID-8,HIE-46,HID-121, GLH-66 | 9 | 5822 | 20 |
| PHSV66E(NW) | 207,220,223,233 239 | 7-141 | HID-8,HIE-46,HID-121 GLH-66 | 9 | 5817 | 20 |

### 4.2.2 Molecular Dynamics Simulation

All protein Molecular Dynamics (MD) simulations were performed with Amber 9 suite of programs [19] in conjunction with the AMBER99SB force field [20] and TIP3P

water [21]. Before any MD simulations, each solvated system was subjected to a 100 cycles of steepest-descent energy minimization followed by a 100 cycles of conjugate gradient minimization. The solvated and minimized structure was then heated gradually in the canonical ensemble (constant N, V, T) in steps of 50 K with 100 ps per step. After reaching the simulation temperature of 298 K, the simulation was switched to a 200 ps equilibration process under constant pressure of 1 bar and temperature of 298 K (NPT) condition. Totally 800 ps MD simulation was conducted for each system in the heatup and equilibration stages. Production runs were performed subsequently after the equilibration and MD trajectory data were collected at an interval of 1 ps. In our simulation, a time step of 2 fs was used and all bonds were constrained (except for heatup simulation, where only bonds involving hydrogen atoms were constrained) with SHAKE algorithm [22]. The temperature was regulated through the Langevin dynamics with a collision frequency of 2.0 $ps^{-1}$ and pressure was maintained utilizing Berendsen weak coupling algorithm [23] with isotropic position scaling and a relaxation time of 2.0 ps. Short-range nonbonded interactions were cut off at 9.0 Å and long-range electrostatics were treated with particle mesh Ewald method [24].

To simulate protein devoid of bound water, a repulsive potential of form $u = 4\epsilon \sum_o [\sigma/(|\mathbf{r}_o - \mathbf{r}_x| - r_{eq})]^{12}$ was applied to each binding site, where $\mathbf{r}_o$ is the position of a solvent water oxygen and $\mathbf{r}_x$ is the position of water-binding site, the summation is over all solvent molecules in the simulated system. The parameter $\epsilon$ is the repulsion strength and $r_{eq}$ is the repulsion radius. The repulsive energy decreases rapidly with distance beyond the repulsion radius. An uniform value of $\epsilon$ was used for all the simulations, but different values of $r_{eq}$ were chosen for different binding sites so that the water molecules can be repelled without disturbing too much of the solvent structure. The values of $r_{eq}$ are shown in Table 4.2. The protein may undergo rotation and translation during the simulation process, the water-binding site must move along with the protein. This was implemented by fitting the crystal structure onto the simulation structure in every MD

step to minimize the root mean square deviation between $C_a$ atoms. The position of the water-binding site, $\mathbf{r}_x$, was taken to be the location of the oxygen atom of the crystal bound water.

Table 4.2: Repulsion radii for water-binding sites

| Protein | Repulsion radius $r_{eq}$ (Å) | | | | | | |
|---|---|---|---|---|---|---|---|
| BPTI(NW) | 111:2.6 | 112:3.2 | 123:3.2 | 122:3.2 | | | |
| PHS(NW) | 201:2.6 | 202:4.0 | 203:4.0 | 211:3.6 | 214:2.6 | 223:2.6 | 242:2.6 |
| HEWL(NW) | 1001:4.0 | 1003:2.6 | 1004:4.0 | 1007:3.0 | 1008:3.2 | 1015:4.0 | 1073:2.0 |
| PHSV66E(NW) | 207:3.6 | 220:3.2 | 223:3.6 | 233:4.0 | 239:2.6 | | |

## 4.3 Results

### 4.3.1 Monitoring Bound Water Molecules

Although at least 20 ns MD simulation was performed for each protein (both with and without bound water), not all the trajectories generated in the simulation of protein without bound water were used for data analyses because of the water penetration. The protein may undergo conformational changes due to the loss of bound water. Some conformational changes can create new empty sites in the protein interior, the solvent water molecules will penetrate into and occupy these new sites. Since in this study, we want to investigate how the protein flexibility is influenced by the bound water molecules, the trajectories with water molecules penetrated into the protein interior must be excluded from our analyses.

To monitor if the water has been repelled successfully by the repulsive perturbation. the minimum distances of solvent molecules to each binding sites are calculated. These data are shown in Figures 4.2–4.5. In Figure 4.2, there is a sharp drop in the minimum distances for W122 site in BPTI at about 14 ns. A careful check on the simulation structure indicates that this drop is caused by a water molecule penetrating into the pro-

tein interior. Although repulsive potentials have been applied to each of the binding sites, the conformational change of BPTI due to the void of bound water creates another empty site between W122 and W113. A water molecule is trapped in this new site and remains there during the rest of the simulation. The typical fluctuation of the minimum distance is about 2 Å, while this fluctuation drops to 1 Å after 14 ns at both W122 and W113 sites in BPTI, both decreases are due to the presence of this water molecule (WAT427 in Figure 4.6). Similar decreases in the minimum distances and their fluctuations at W112 and W111 sites are also seen in Figure 4.2. The examination of the simulation structure finds no water penetration, but some relatively stable solvent molecules bind to the protein surface close to the entrance of the channel-like cavity (WAT3609 in Figure 4.6). Due to the penetration of water, all following analyses for BPTI without bound water are based on the first 13 ns simulation. In the simulation of HEWL, a water molecule is also found to enter into the protein interior at the end of 18 ns, as indicated by the sharp decreases in the minimum solvent distance and its fluctuation for W1001 site in Figure 4.3. For the HEWL analysis, we will use on the first 17 ns simulation trajectory.

Although the minimum distances drop slowly for the W211 and W242 sites in PHS simulation (see Figure 4.4), no water penetration is observed at these sites. These two sites locate next to each other in the region close to the mobile Ω-loop of SNase, the bound water molecules at these sites play important roles in bridging the loop and the rest part of the protein. The loss of bound water at these sites breaks the connection and exposes the binding sites to the solvent. For the analysis of PHS and PHS/V66E, the whole trajectories generated in the simulations are used. The simulation times used for each protein analysis in this study are listed in Table 4.1. It should be noted that, the minimum distances for some binding sites are far beyond the repulsion radii, for example, W203 in PHS and W223 in PHS/V66E, this is because these sites are deeply buried in the protein interiors and the entries of solvent molecules are deterred by the steric interactions with the protein atoms.

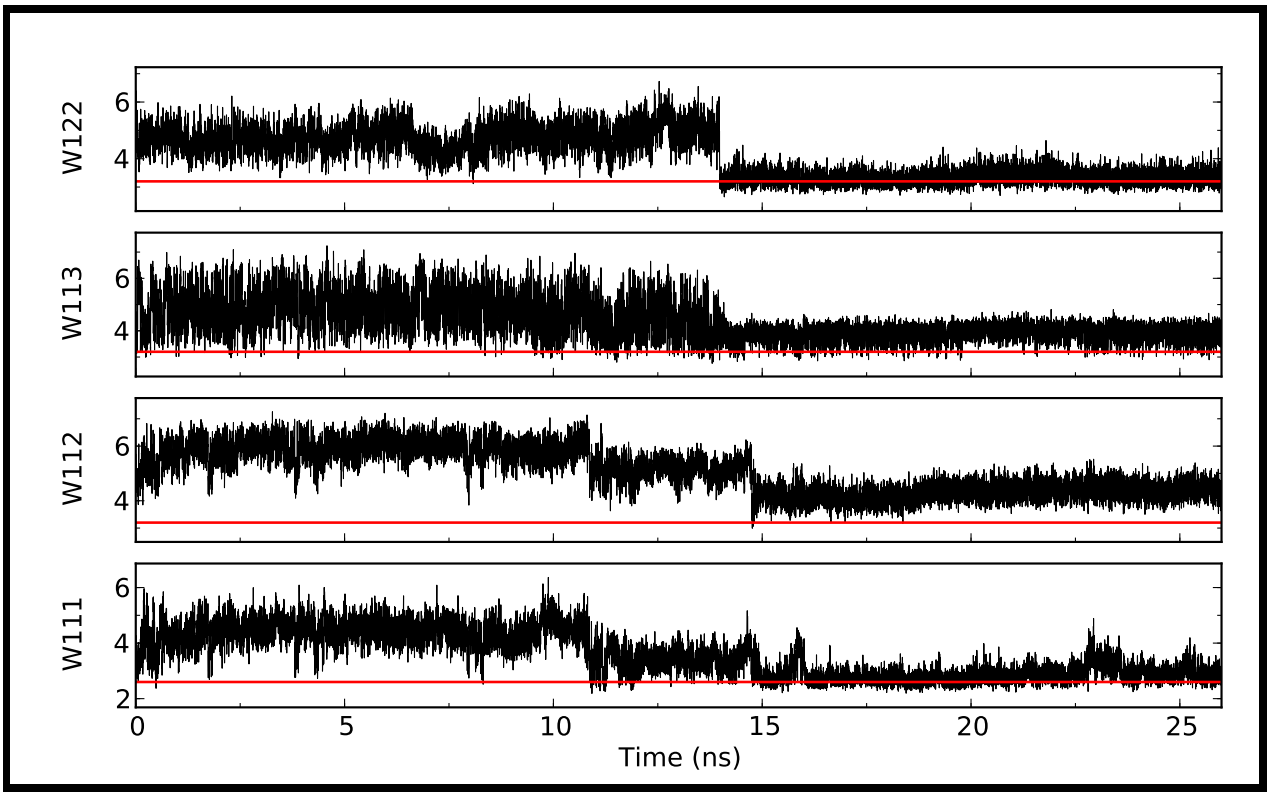


Figure 4.2: Minimum solvent distance to each water-binding site in the simulation of BPTI without bound water. The x-axis is the simulation time and y-axis is the minimum distance (unit Å). The red line indicates the repulsion radius for each binding site.

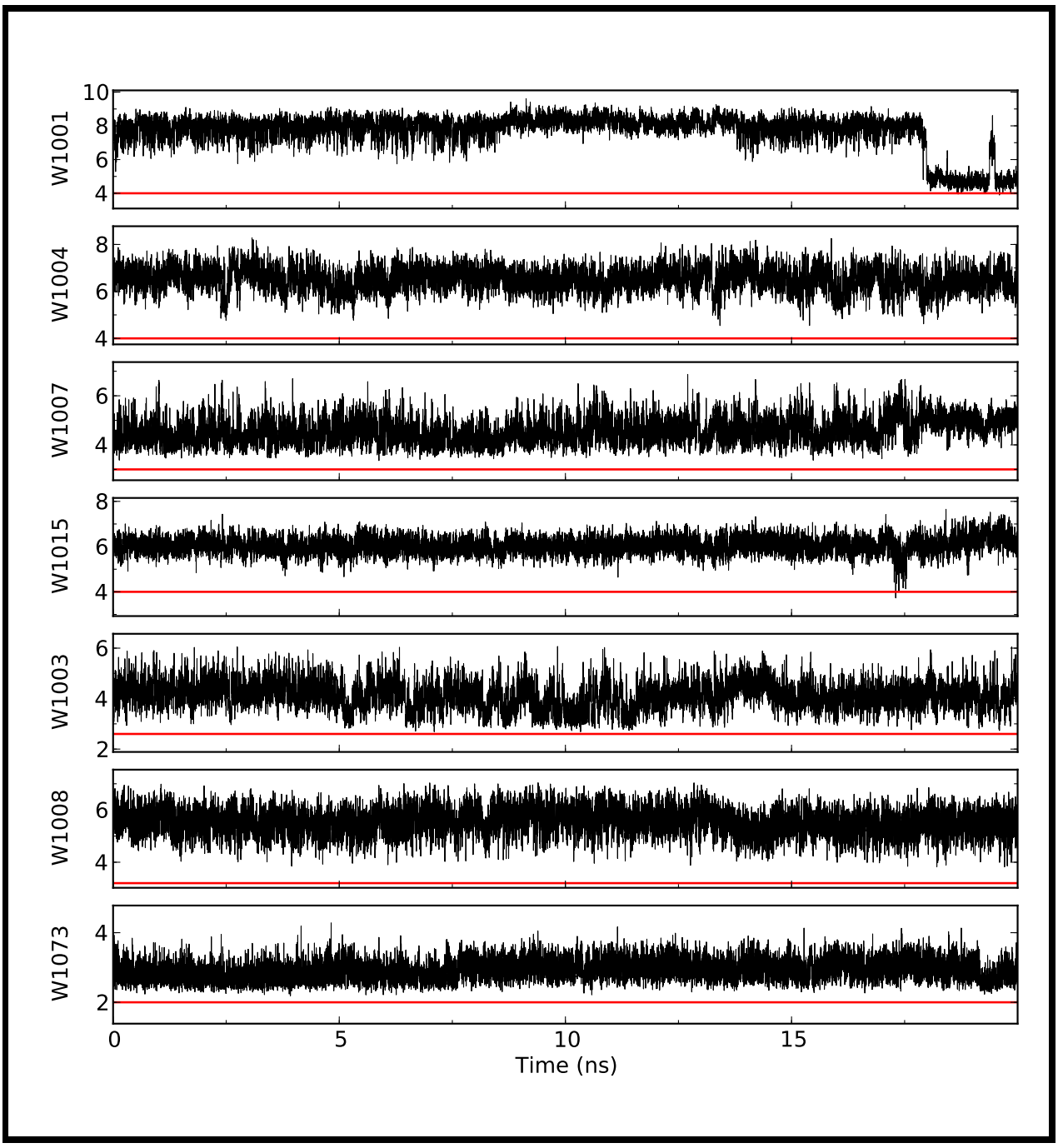


Figure 4.3: Minimum solvent distance to each water-binding site in the simulation of HEWL without bound water. The x-axis is the simulation time and y-axis is the minimum distance (unit Å. The red line indicates the repulsion radius for each binding site.

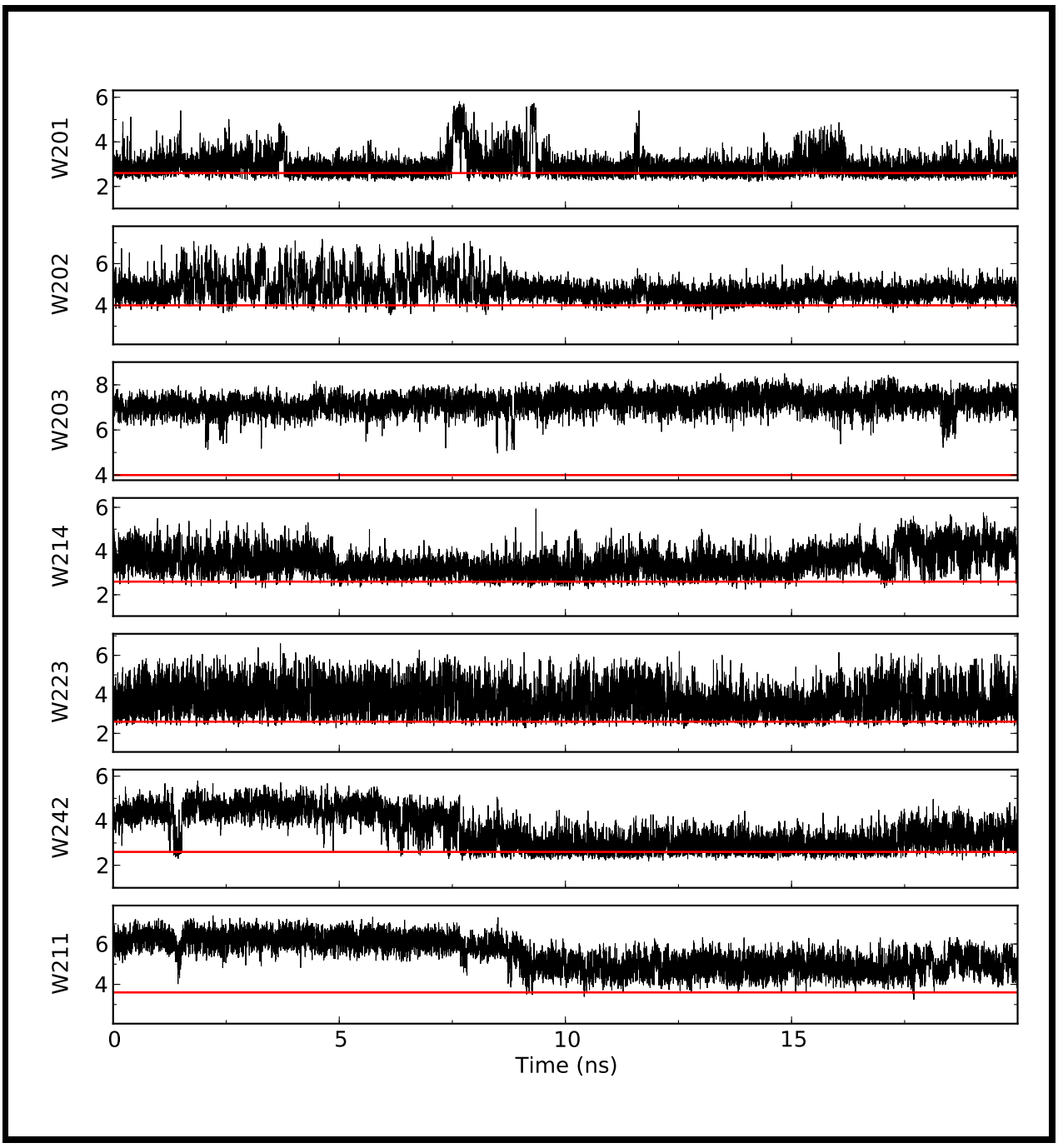


Figure 4.4: Minimum solvent distance to each water-binding site in the simulation of PHS without bound water. The x-axis is the simulation time and y-axis is the minimum distance (unit Å). The red line indicates the repulsion radius for each binding site.

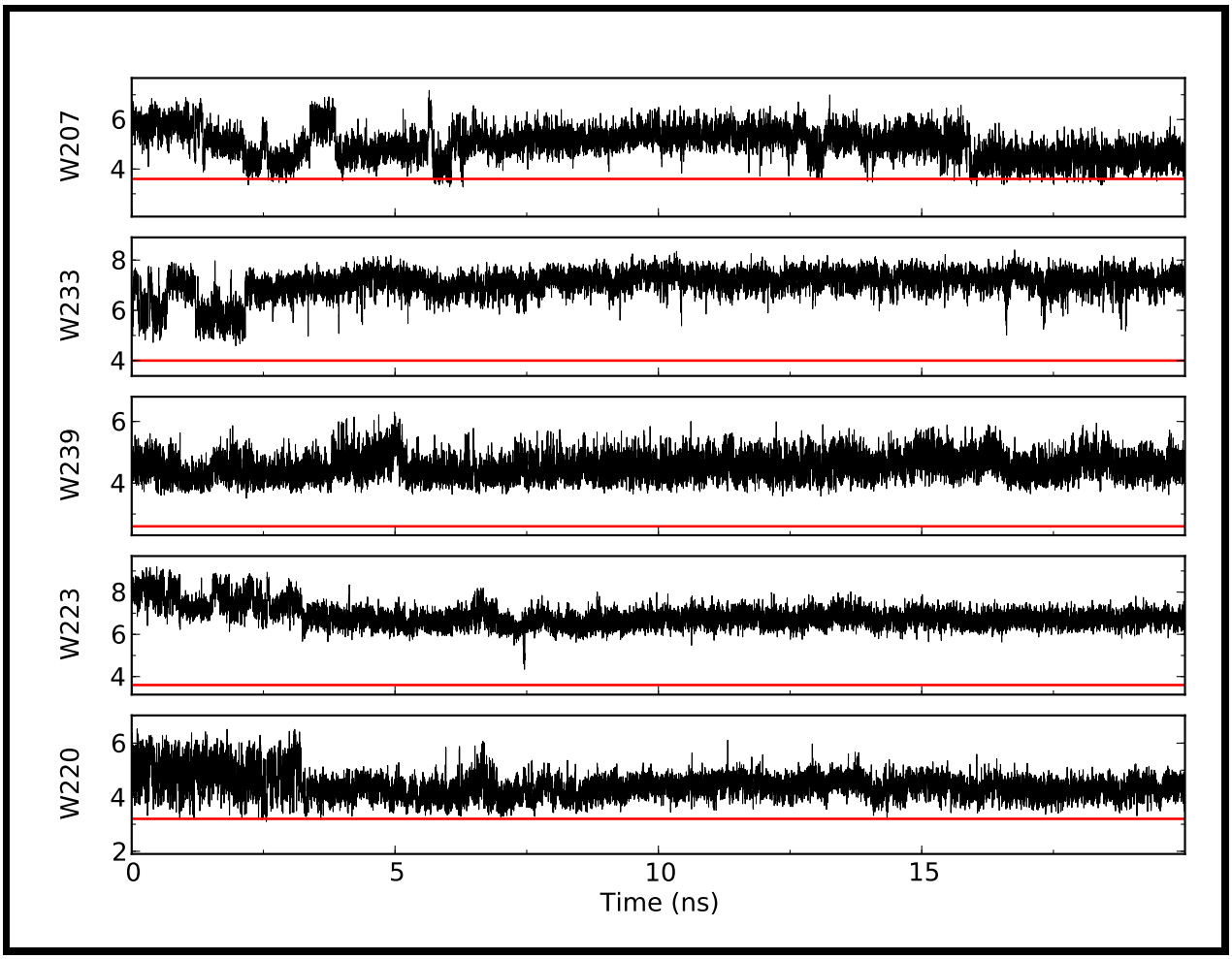


Figure 4.5: Minimum solvent distance to each water-binding site in the simulation of PHSV66E without bound water. The x-axis is the simulation time and y-axis is the minimum distance (unit Å). The red line indicates the repulsion radius for each binding site.

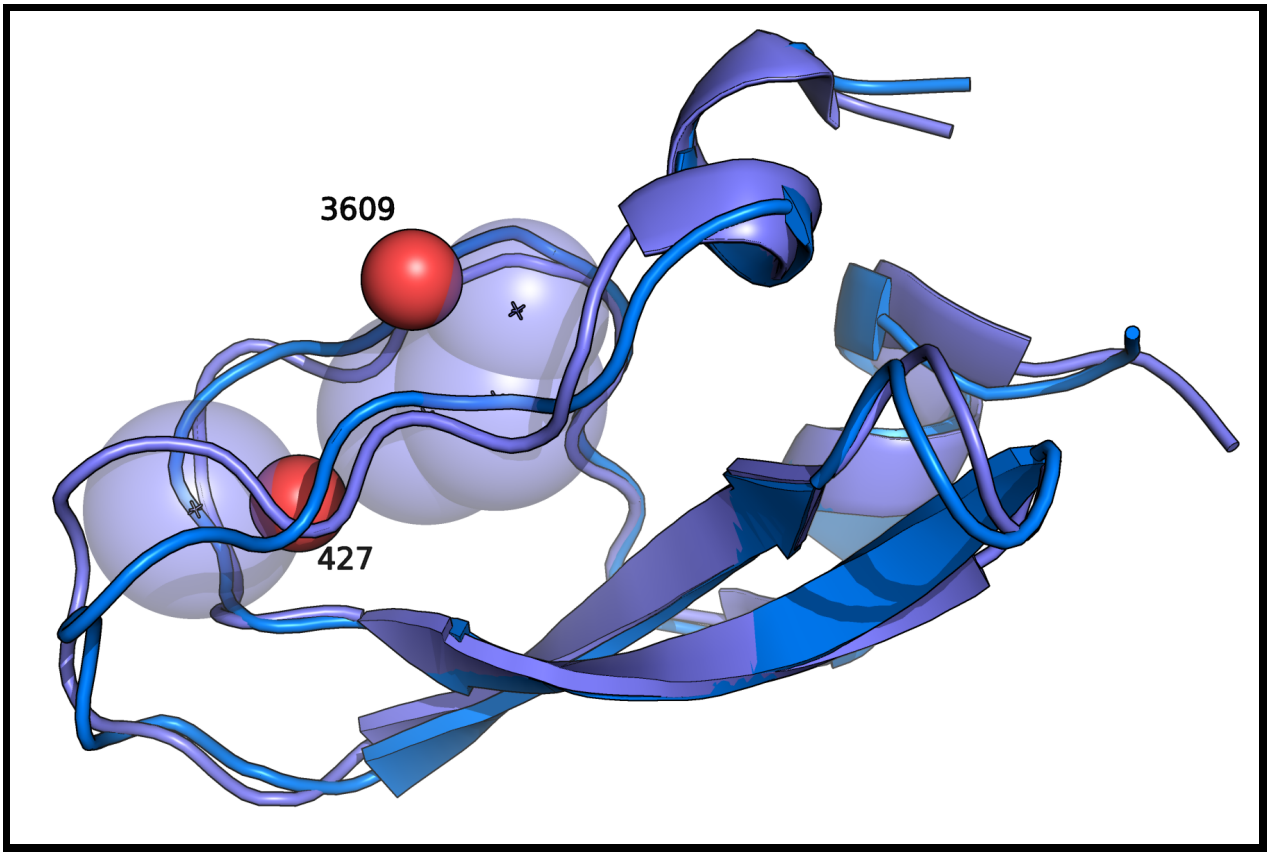


Figure 4.6: The water molecules caused drops in minimum solvent distances and their fluctuations in BPTI. The crystal and simulation structures of the protein are shown in purple and blue, respectively. The water binding sites are shown as black crosses and their repulsion radii are illustrated by purple spheres. The conformational change due to the void of bound water molecules creates a new empty site in the protein interior, a water molecule, W427, penetrates into this empty site at about 14 ns and is trapped stably during the rest of simulation. Both decreases in the minimum distances and their fluctuations for W122 and W113 sites are due to this water. Water WAT3609 bind to the protein surface at the entrance of a channel-like cavity with relative stability and contributes to the drops at W111 and W112 binding sites.

### 4.3.2 Bond water and Protein Structural Change

To investigate how the bound water influence the protein structure, the root mean square deviation (RMSD) could be evaluated. The RMSD of the protein with respect to some reference structure at simulation time $t$ is defined by

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(\mathbf{r}_i(t) - \mathbf{r}_i^{ref}\right)^2} \tag{4.1}$$

where $N$ is the total number of atoms, $\mathbf{r}_i(t)$ and $\mathbf{r}_i^{ref}$ are the Cartesian coordinates of atom $i$ in the simulation structure of time $t$ and in the reference structure, respectively. For comparison, both RMSDs for protein with and without bound water are determined by using the crystal structure as the reference structure. The RMSDs for the 4 proteins as a function of simulation time are illustrated in Figure 4.7

For BPTI with bound water, the RMSD values are between 0.75 and 2.46 Å, with the maximum value is found at 13.355 ns. The RMSD exhibits a quasi-periodic behavior in the first 15 ns, its value then decreases monotonically to about 1.25 Å and maintains low structural changes after 25 ns. The average value of RMSD during the 31 ns simulation is 1.43±0.09 Å (All error bars in this Chapter were calculated based on 1 ns window). For BPTI without bound water, the RMSD is relatively stable during the 13 ns simulation, no apparent decrease or increase is observed. The maximum RMSD is 2.62 Å and minimum value is 1.18 Å, with a average RMSD of 1.74±0.06 Å. In the simulations of HEWL, both proteins with and without bound water show very similar structure with respect to the protein crystal. The RMSD values are very low and essentially do not change in the whole simulation processes. The simulation of HEWL with bound water has a average RMSD value of 0.84±0.03 Å, whereas the simulation without bound water has a slightly higher average RMSD of 0.98±0.03 Å. For the simulations of PHS, The RMSD of protein with bound water increase rapidly from ~0.8 Å to ~ 1.2 Å in the first nanosecond. the protein structure then maintains fairly stable structure, no dramatic change in RMSD is seen

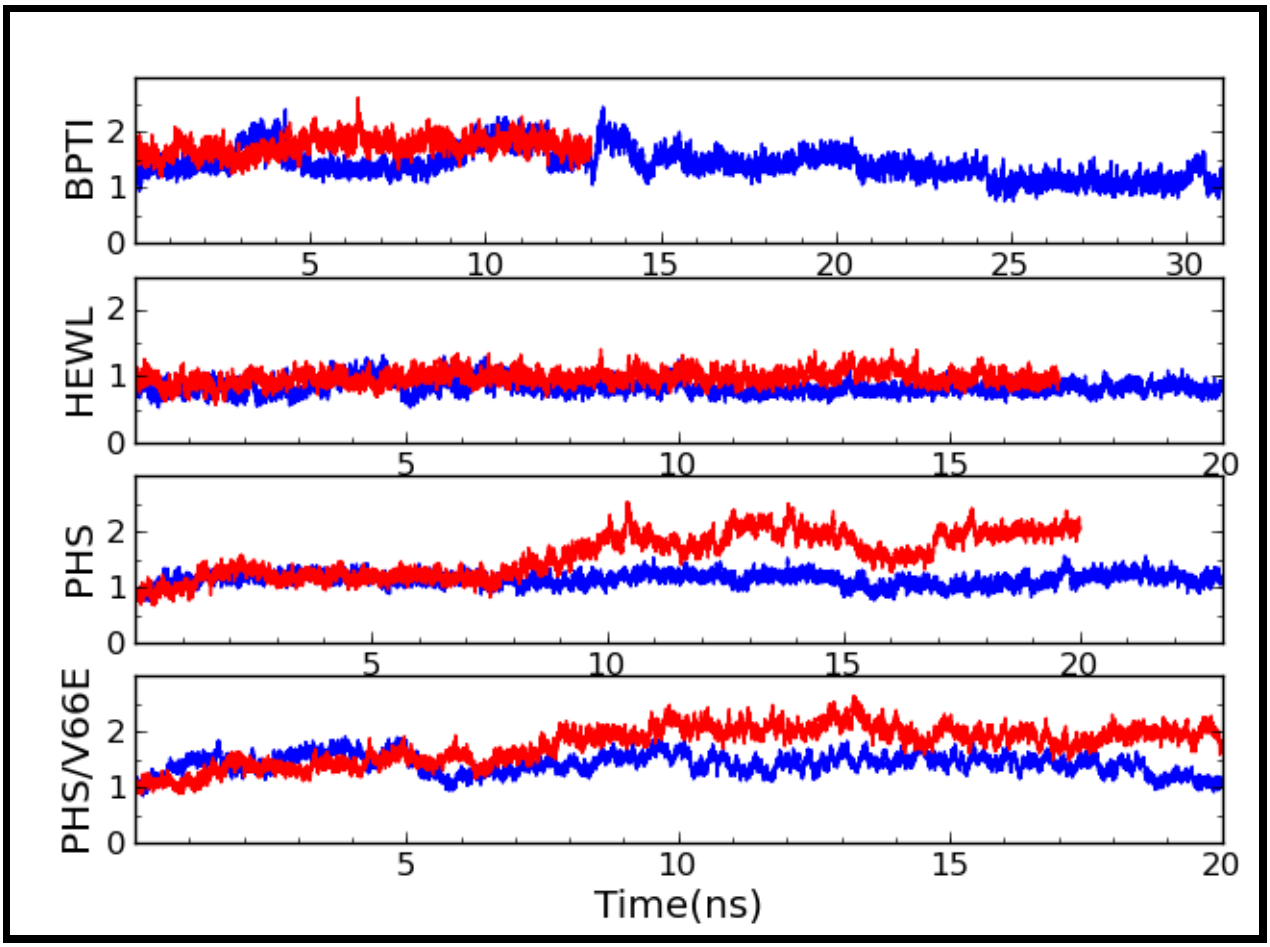


Figure 4.7: RMSDs of $C_a$ atoms relative to the crystal structures for 4 proteins as a function of simulation time. The blue line is for the simulation of protein with bound water while the red line is for the simulation without bound water.

in the rest of simulation. The average RMSD for PHS with bound water is 1.14 Å. The simulation of PHS without bound water shows a different behavior. In the first 2 ns, the protein RMSD increases gradually from below 1.0 Å to ~1.2 Å, it then maintains this value for about 6 ns. After 8 ns, the RMSD value increases quickly to ~2.0 Å within 2 ns and undergo larger fluctuations within 1.5–2.5 Å thereafter. The average RMSD value for PHS without bound water is 1.15 Å in the first 8 ns and is 1.90 Å in the last 10 ns. The RMSD for PHS/V66E with bound water increase rapidly from its initial value of 1.23 Å to 1.81 Å within 1.5 ns. Its value decreases back to ~1.0 Å before increasing again to ~1.6 Å. The protein structure is stable until 18 ns where a gradually decrease to ~ 1.1 Å is observed. The RMSD for PHS/V66E without bound water increases gradually from ~1.0 Å to ~2.0

Å within the first 10 ns and then undergo quasi-periodic small fluctuations. The average RMSDs for PHS/V66E with and without bound water are 1.42±0.06 Å and 1.8±0.2 Å, respectively. The RMSD values averaged over the whole simulation time for the 4 proteins with and without bound water molecules are reported in Figure 4.8. These indicate the long-time scale fluctuations of the structures.

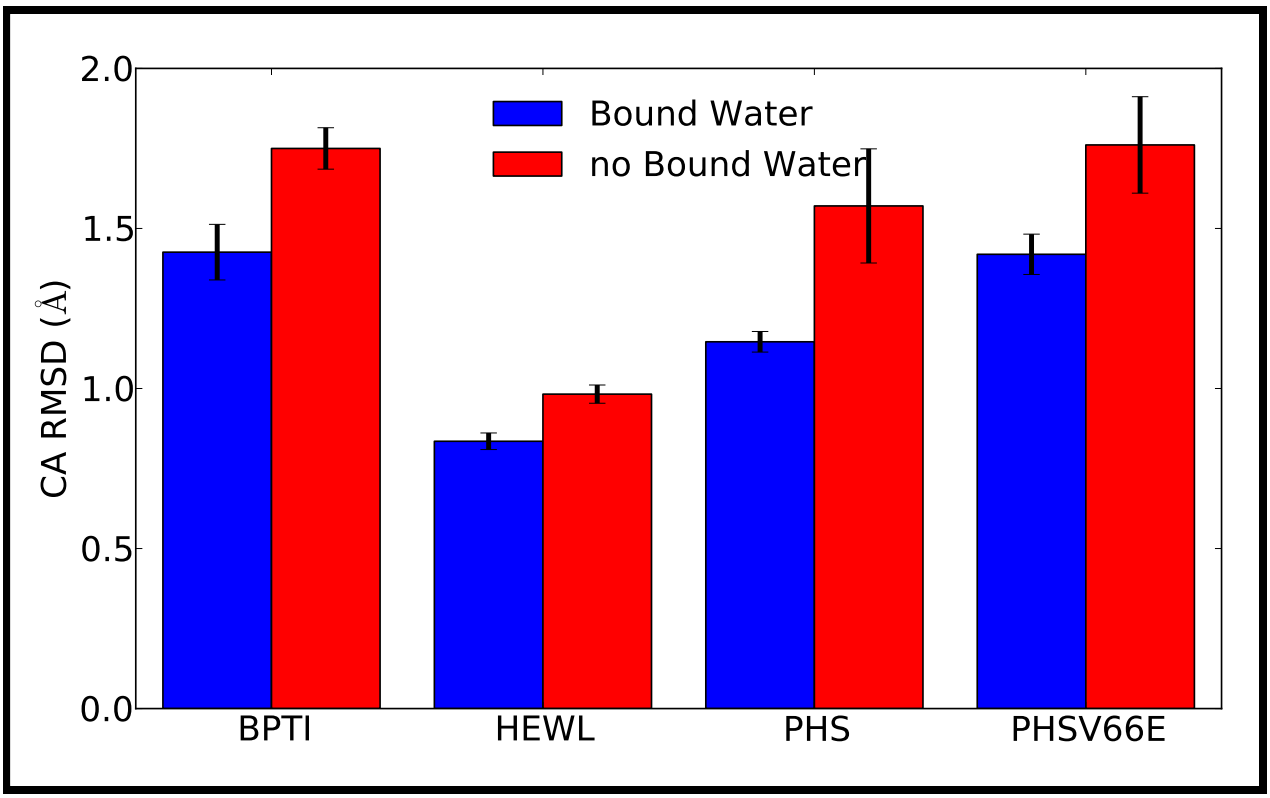


Figure 4.8: RMSD values of proteins calculated based on $C_a$ atoms

The apparent differences in RMSD values for the 4 proteins with and without bound water, especially in the last stages of the PHS and PHS/V66E simulations, imply that the proteins may undergo large structural changes due to the absence of bound water. Furthermore, although no significant RMSD differences are found in the simulation of HEWL, a small RMSD change is observed by averaging over the simulation time. In addition, an invariance in overall protein RMSD does not necessarily mean that the protein do not change its structure during the simulation processes. Small structural changes may occur in local regions and their visibilities in RMSD are depressed by averaging over

all protein atoms. It would be of interest to measure RMSD per residue to access which residues suffer large structural changes. In our study, the RMSD value for each residue is estimated based on the $C_a$ atom using the equation

$$\mathrm{RMSD}(i) = \sqrt{\frac{1}{\tau}\sum_{t}^{\tau} |\mathbf{r}_i(t) - \mathbf{r}_i^{ref}|} \tag{4.2}$$

These RMSD values (with respect to crystal structure) are plotted in Figure 4.9.

For the simulation of BPTI, significant structural changes are observed in two regions, which comprise residues 8–17 and 36–41. The overage RMSD for these two regions is 0.77 Å for protein with bound water, while this value increases to 1.94 Å for protein without bound water. In the protein crystal, these two regions have loop structures and form the cavities that hold the bound water molecules. The bound water molecules make several hydrogen bonds with the loop residues and bridges their interaction. Due to the loss of bound water molecules, the loops experience large structural changes. It should be noted that large RMSD values are seen for the C-terminal residues, ALA-58, in both simulations of BPTI with and without bound water. ALA-58 is a small hydrophobic residue with a methyl side chain. Crystal structure examination reveals that this residue exposes isolatedly to the solvent environment, making no direct interactions with the other part of the protein. The unfavorable interactions with the solvent molecules cause this residue undergo large structural change.

For HEWL, both simulations show small overall RMSD values (below 1.0 Å). The examination of RMSD value for each $C_a$ reveals that a majority of residues (including some terminal residues) have low RMSD values indicating that the proteins do not suffer significant structural changes. However, three regions with relative large conformational changes are still identified, as indicated by their large RMSD values. The three regions are residues 45–49, 67–71, and 100-103. Furthermore, large RMSD fluctuations are also observed in these regions (data not shown), reflecting that they consist of highly flexibly residues. Comparing the simulations of proteins with and without bound water, structural

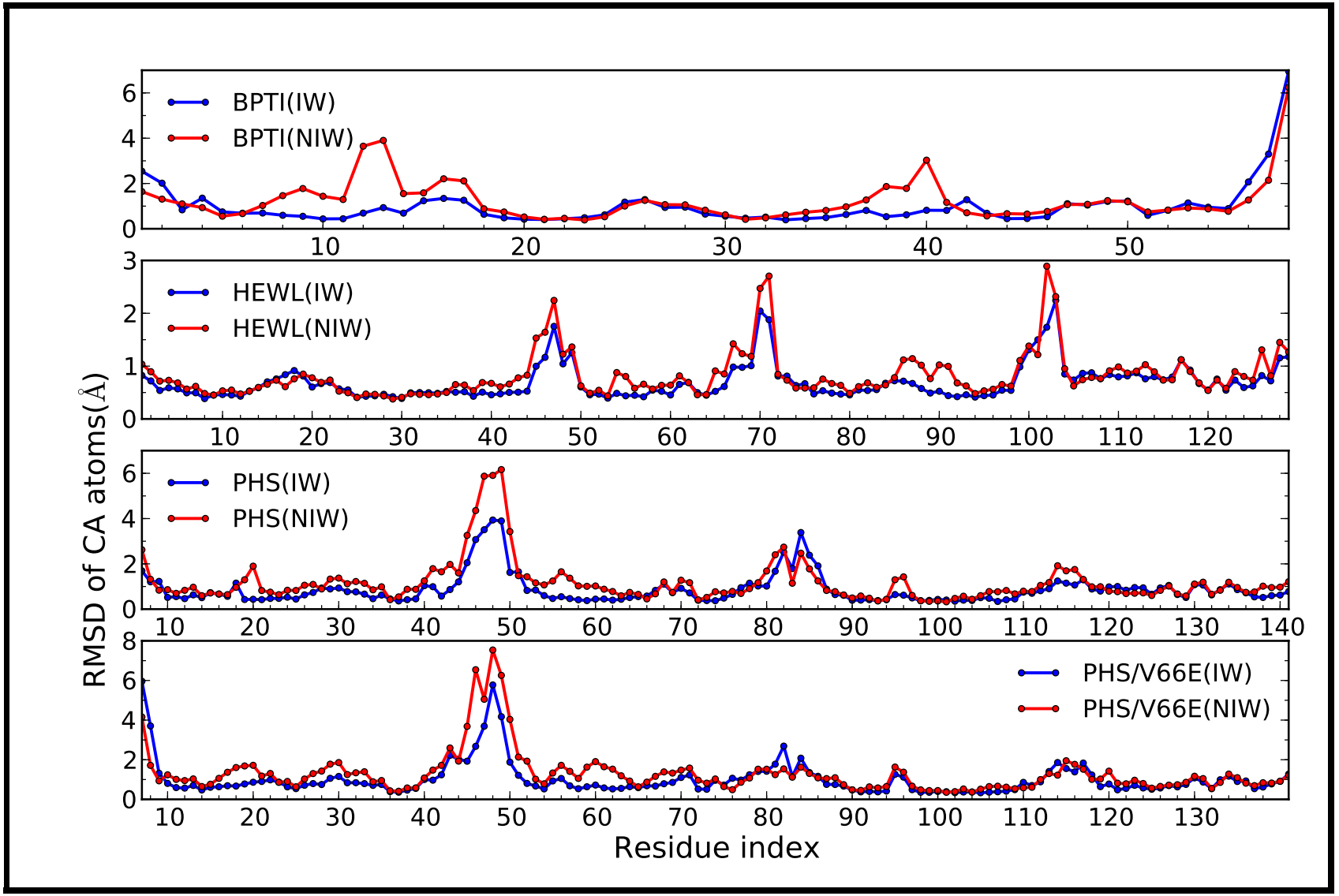


Figure 4.9: Average RMSDs for each of $C_a$ atoms in the 4 proteins relative to the crystal structures. The blue lines and dots are for the simulations of proteins with bound water and the red lines and dots are for simulations without bound water.

changes are observed in residues 86–90. The average RMSD value of these 5 residues in protein with bound water is 0.60 Å, while this value increases to 1.01 Å in protein without bound water, revealing that these residues undergo small structural changes after losing bound water molecules.

The simulations of PHS and PHS/V66E are less stable than HEWL, as demonstrated by their larger protein RMSD values and fluctuations. The analysis of RMSD for each $C_a$ in PHS and PHS/V66E reveals that, except for the N-terminal residue, two regions are also undergo significant structural changes. They are residues 40–51 and 78–86. The average RMSD values for these residues are 1.96 Å and 2.05 Å in PHS and PHS/V66E with bound water, respectively. Although no significant RMSD differences are found for PHS/V66E with and without bound water, a large RMSD increase is observed in residues 46-50 for PHS without bound water compared with PHS with bound water. The rapidly change in PHS simulation without bound water occur at 8 ns may attribute to these 5 residues. In addition, several small RMSD increases upon removing of bound water are seen in both PHS and PHS/V66E simulations, indicating that these regions at least suffer small structural changes, Maybe they are too rigid to be distorted significantly.

### 4.3.3 Bound Water and Protein Flexibility Change

To study how the protein changes its flexibility upon removing of bound water, we calculated the mean square positional fluctuations (MSF) of atoms in the protein. The MSF for atom $i$ in the simulation is given by

$$\mathrm{MSF}(i) = \frac{1}{\tau}\sum_{t}^{\tau}(\mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle)^2 \tag{4.3}$$

where $\tau$ is the simulation time, $\langle \mathbf{r}_i \rangle$ is the average position of atom $i$ during the simulation time $\tau$. The protein's MSF is evaluated by a mass-weighted averaging over all atoms

$$\mathrm{MSF} = \frac{\sum_i^N m_i \times \mathrm{MSF}(i)}{\sum_i^N m_i} \tag{4.4}$$

where $m_i$ is the mass of the $i$th atom. To eliminate the contributions from the translational and rotational motions, the simulation structure has been superposed onto the crystal structure before MSF calculation. The atomic MSF($i$) can be converted to the crystallographic atomic B-factor, $B_i$, using the relation [25, 26, 27]:

$$\mathrm{MSF}(i) = \frac{3B_i}{8\pi^2} \tag{4.5}$$

The MSF values for the 4 proteins calculated based on main chain $C_a$ atoms are ploted in Figure 4.11. All these 4 proteins, except PHS/V66E, become more flexible after losing bound water. The MSF value for HEWL with bound water is 0.28±0.01 Å$^2$, this value increases to 0.34±0.02 Å$^2$ after removing bound water molecules. Although PHS with bound water is slightly more rigid than HEWL, with a MSF value of 0.23±0.02 Å$^2$, its flexibility increases up to 0.30±0.02 Å$^2$ after losing bound water, comparable with that of HEWL without bound water. Both protein have a flexibility change of about 0.65 Å$^2$. Comparing with HEWL and PHS, BPTI suffers less flexibility change, The MSF fluctuation for BPTI protein with and without bound water are 0.27±0.01 Å$^2$ and 0.31±0.02 Å$^2$, respectively. Only a flexibility change of 0.03±0.02 Å$^2$ is observed. For PHS/V66E protein, no apparent change is seen in its flexibility, even considering the MSF of all non-hydrogen atoms (see Table 4.3).

The MSF fluctuation for each $a$ carbon atoms are illustrated in Figure 4.11. In BPTI protein, two regions are observed obviously to change their flexibilities after the loss of bound water molecules. These regions include residues 8–12 and 39–42 which are residues close to the bound water cluster of WAT111–WAT113. Although it seems that residues 3–4 and 25-28 also change their flexibility, a definitive conclusion is prevented by the large error bars. The WAT122 forms totally 4 hydrogen bonds with THR-11, CYS-14 and CYS-38, however, their flexibilities have almost no change. HEWL has several regions in which flexibilities are changed: 3 segments, residues 39–42, 54–57 and 81–93, locate nearby the 4 water cluster, while 1 region including 102–105 is far away from any bound

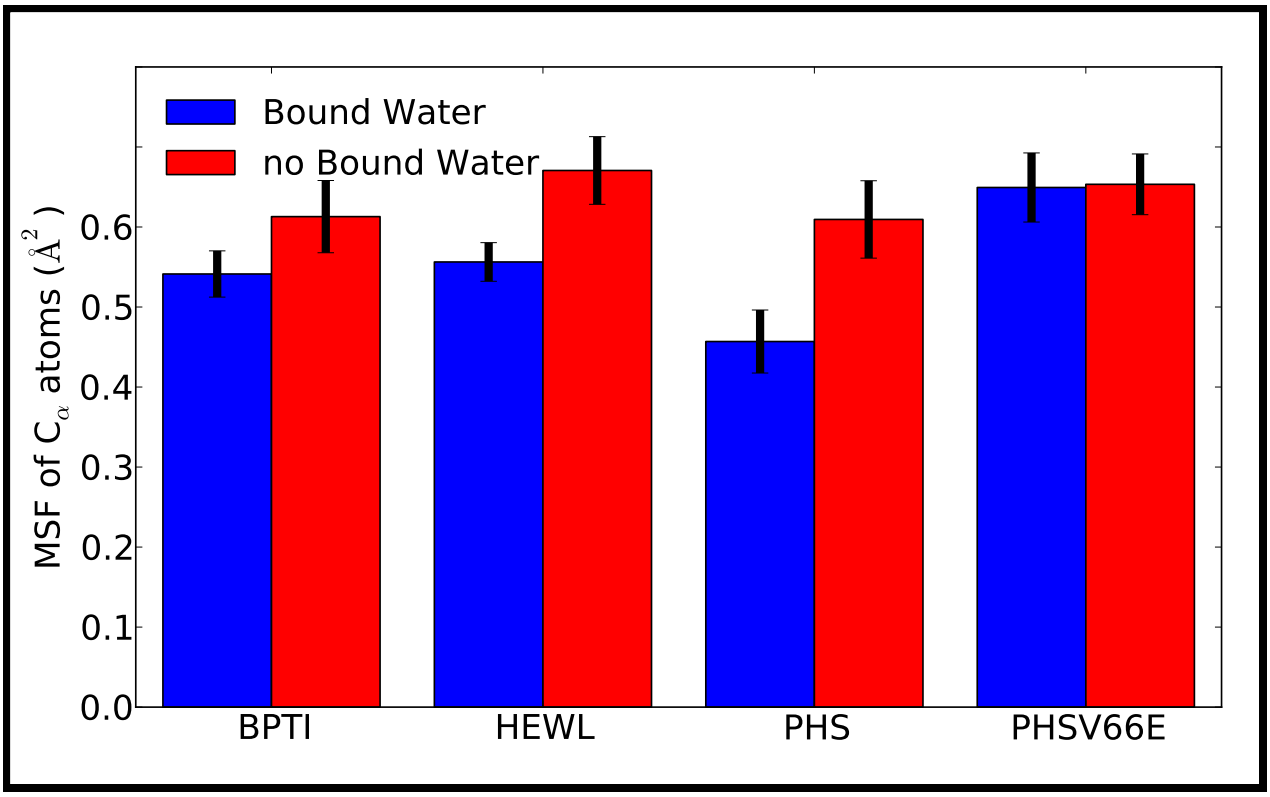


Figure 4.10: MSF values of proteins calculated based on $C_a$ atoms

water molecules. Although two water molecules are buried by residues 60–74, apparently no flexibility changes are found for them. PHS is the protein observed to have largest flexibility change among the 4 proteins studied. The significant flexibility change seems is due to residues 42–49, which form the Ω loop and locate close to the two bound water WAT211 and WAT242. The rest neighbors of these two water molecules, residues 18–21, are also observed to have a small flexibility change. Besides these two regions, flexibility change that is so large and can not be missed is also seen in residues THR-93 and ALA-94. These 2 residues reside close the surface bound water WAT201. No apparent flexibility change is observed in any residues of PHS/V66E.

## 4.4 Summary

To summarize, we have performed Molecular Dynamics simulations for four proteins with bound water included in and excluded from the protein interiors. In our study,

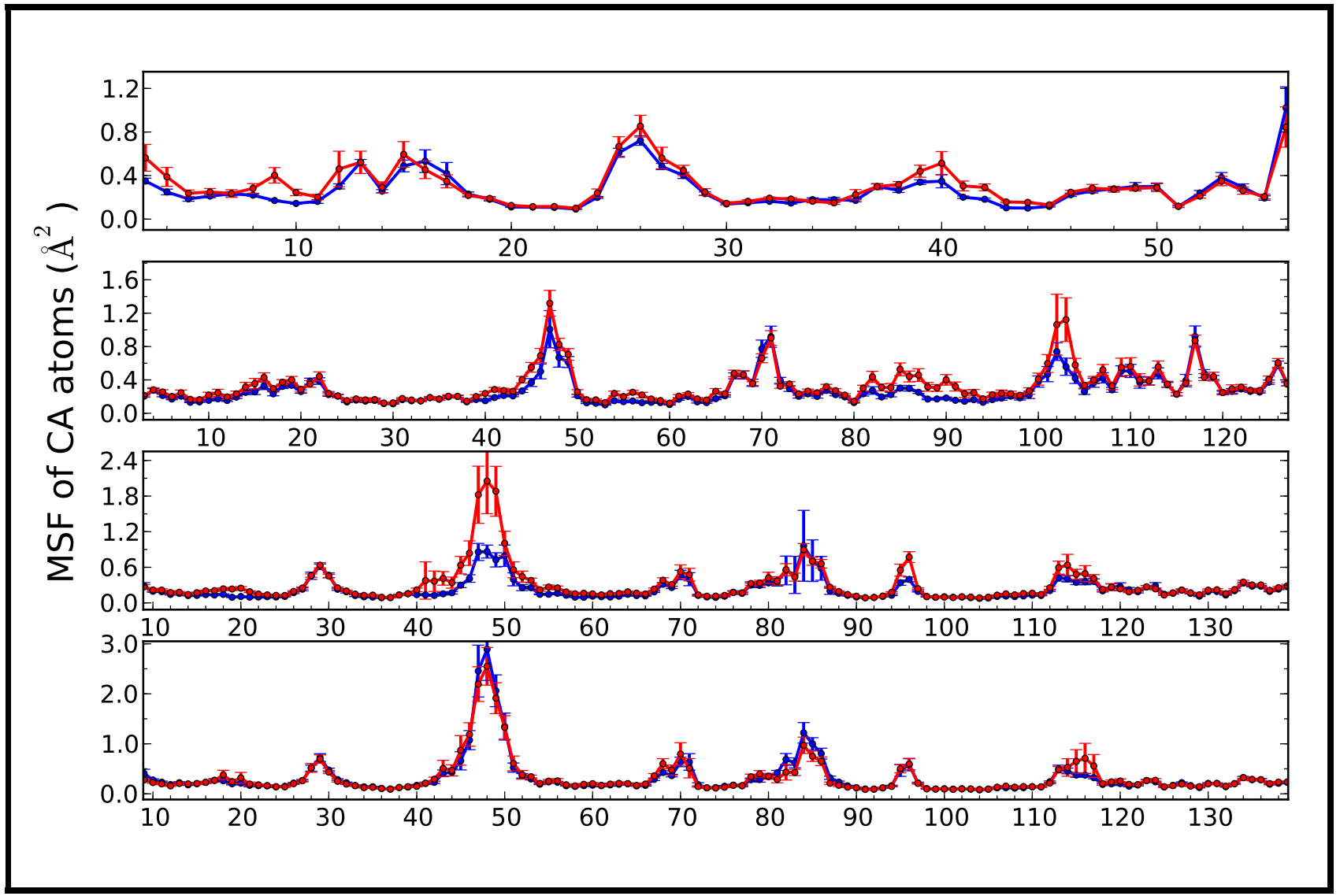


Figure 4.11: Average RMSDs for each of $C_a$ atoms in the 4 proteins relative to the crystal structures. The blue lines and dots are for the simulations of proteins with bound water and the red lines and dots are for simulations without bound water.

Table 4.3: Protein MSF values computed based on different components.

| Protein | MSF (Å$^2$) | | |
|---|---|---|---|
| | $C_a$ atoms | Backbone atoms | non-hydrogen atoms |
| BPTI(W) | 0.27±0.01 | 0.29±0.02 | 0.52±0.03 |
| BPTI(NW) | 0.31±0.02 | 0.33±0.02 | 0.55±0.04 |
| HEWL(W) | 0.28±0.01 | 0.30±0.01 | 0.44±0.02 |
| HEWL(NW) | 0.34±0.02 | 0.35±0.02 | 0.52±0.03 |
| PHS(W) | 0.23±0.02 | 0.25±0.02 | 0.44±0.03 |
| PHS(NW) | 0.30±0.02 | 0.32±0.02 | 0.56±0.03 |
| PHSV66E(W) | 0.32±0.02 | 0.34±0.02 | 0.58±0.03 |
| PHSV66E(NW) | 0.32±0.02 | 0.34±0.02 | 0.60±0.03 |

we focus on the protein structure and flexibility changes upon removing of bound water molecules. All the four proteins simulated in our study, BPTI, HEWL, PHS and PHS/V66E, change their structures after losing of bound water. The structure change does not happen only in the local regions close to the bound water, residues far away from the water binding site also suffer small structure changes. Three of the four proteins are observed to have increased flexiblities after the bound water is removed, while no obvious flexibility change is found in PHS/V66E protein. Our finding is in agreement with the Mao et al. conclusion [28], but different form the studies of Fischer, Verma and co-works [29, 30, 31] and Olano and Rick [32]. A detailed analysis on pattern changes of hydrogen bonds may be required for understanding the complex influences of water on protein flexibility.

# Bibliography


[1] Robert Huber and William S. Bennett, Jr. Functional significance of flexibility in proteins. *Biopolymers*, 22:261–279, 1983.

[2] Gregory A. Petsko and Dagmar Ringe. *Protein Structure and Function*. New Science Press, London, 2003.

[3] David Whitford. *Proteins: Structure and Function*. John Wiley & Sons, 2005.

[4] J. T. Edsall and H. A. McKenzie. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. *Adv. Biophys.*, 16:53–183, 1983.

[5] Uma Sreenivasan and Paul H. Axelsen. Buried water in homologous serine proteases. *Biochemistry*, 31:12785–12791, 1992.

[6] Alexander Wlodawer, Joseph Nachman, Gary L. Gilliland, Warren Gallagher, and Clare Woodward. Strcuture of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.*, 198:469–480, 1987.

[7] Martin A. Walsh, Thomas R. Schneider, Larry C. Sieker, Zbigniew Dauter, Victor S. Lamzin, and Keith S. Wilson. Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta Cryst. D*, 54:522–546, 1998.

[8] Junmei Chen, Zhiqiang Lu, Joshua Sakon, and Wesley E. Stites. Increasing the thermostability of staphylococcal nuclease: Implications for the origin of protein thermostability. *J. Mol. Biol.*, 303:125–130, 2000.

[9] Coordinates for PHS/V66E nuclease at cryogenic temperature kindly provided by Bertrand García-Moreno E.

[10] Thomas R. Hynes and Robert O. Fox. The crystal structure of staphylococcal nucelase refined at 1.7 å resolution. *Proteins: Struct. Funct. Bioinf.*, 10:92–105, 1991.

[11] Carolyn A. Fitch, Daniel A. Karp, Kelly K. Lee, Wesley E. Stites, Eaton E. Lattman, and García-Moreno E. Bertrand. Experimental $pK_a$ values of buried residues: Analysis with continumm methods and role of water penetration. *Biophys. J.*, 82:3289–3304, 2002.

[12] Kelly K. Lee, Carolyn A. Fitch, and Bertrand García-Moreno E. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci.*, 11: 1004–1016, 2002.

[13] Ana Damjanović, Bertrand García-Moreno E., Eaton E. Lattman, and Angel E. García. Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins: Struct. Funct. Bioinf.*, 60:433–449, 2005.

[14] Ana Damjanović, Jamie L. Schlessman, Carolyn A. Fitch, Angel E. García, and Bertrand. Role of flexibility and polarity as determinats of the hydration of internal cavities and pockets in proteins. *Biophys. J.*, 93:2791–2804, 2007.

[15] Michael J. Harms, Carlos A. Castañeda, Jamie L. Schlessman, Gloria R. Sue, Daniel G. Isom, Brian R. Cannon, and Bertrand García-Moreno E. The $pk_a$ values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol.*, 389:34–47, 2009.

[16] John J. Dwyer, Apostolos G. Gitts, Daniel A. Karp, Eaton E. Lattman, Daniel S. Spencer, Wesley E. Stites, and Bertrand E. García-Moreno. High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.*, 79:1610–1620, 2000.

[17] Takuya Takahashi, Haruki Nakamura, and Akiyoshi Wada. Electrostatic forces in two lysozymes: Calculations and measurements of histidine $pk_a$ values. *Biopolymers*, 32: 897–909, 1992.

[18] Kristin Bartik, Redfield Christina, and Christopher M. Dobson. Measurement of the individual $pk_a$ values of acidic residuesof hen and turkey lysozymes by two-dimensional $^1$H NMR. *Biophys. J.*, 66:1180–1184, 1994.

[19] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, and P. A. Kollman. Amber 9, 2006. University of Carlifornia, San Francisco.

[20] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinf.*, 65:712–725, 2006.

[21] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[22] Jean-Paul Ryckaert, Glovanni Ciccotti, and Herman J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[23] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81: 3684–3690, 1984.

[24] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $n\log(n)$ method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[25] Kenneth N. Trueblood. Diffraction studies of molecular motions in crystals. In Aldo Demenicano and Istvan Hargittai, editors, *Accurate Molecular Structures: Their Determination and Importance*, pages 199–219. Oxford University Press, Oxford, 1992.

[26] Akinori Kidera and Nobuhiro Gō. Normal mode refinement: Crystallographic refinement of protein dynamic structure: I. theory and test by simulated diffraction data. *J. Mol. Biol.*, 225:457–475, 1992.

[27] Akinori Kidera, Koji Inaka, Masaaki Matsushima, and Nobuhiro Gō. Normal mode refinement: Crystallographic refinement of protein dynamic structure : II. application to human lysozyme. *J. Mol. Biol.*, 225:477–486, 1992.

[28] Yi Mao, Mark A Ratner, and Martin F. Jarrold. One water molecule stiffens a protein. *J. Am. Chem. Soc.*, 122:2950–2951, 2000.

[29] Stefan Fischer and Chandra S. Verma. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. USA*, 96:9613–9615, 1999.

[30] Stefan Fischer, Jeremy C. Smith, and Chandra S. Verma. Dissecting the vibrational entropy change on protein/ligand binding: Burial of a water molecule in bovine pancreatic trypsin inhibitor. *J. Phys. Chem. B*, 105:8050–8055, 2001.

[31] Jeremy C. Smith, Franci Merzel, Chandra S. Verma, and Stefan Fischer. Protein hydration water: Structure thermodynamics. *J. Mol. Liq.*, 101:37–33, 2002.

[32] L. Renee Olano and Steven W. Rick. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.*, 126:7991–8000, 2004.

# Chapter 5

# Appendix

## 5.1 Supplementary Information for Chapter 3

Table 5.1: Non-local corrections for hydration free energies of cavities with different hydrogen bond forming molecules.

| | | Non-polarizable model | | Polarizable model | |
|---|---|---|---|---|---|
| | $R$(Å) | $\Delta G_{KO}$(kcal/mol) | $\Delta G_{LJ}$(kcal/mol) | $\Delta G_{KO}$(kcal/mol) | $\Delta G_{LJ}$(kcal/mol) |
| 0F0M | 3.614 | -0.1615(3) | -0.4332(7) | -0.1199(3) | -0.431(2) |
| 1F0M | 3.614 | -0.1617(3) | -0.4337(9) | - | - |
| 0F1M | 3.614 | -0.1603(4) | -0.430(2) | -0.1418(4) | -0.4280(6) |
| 2F0M | 3.614 | -0.1612(3) | -0.4323(9) | - | - |
| 1F1M | 3.614 | -0.1603(5) | -0.430(2) | - | - |
| 0F2M | 3.614 | -0.1591(3) | -0.4269(9) | -0.1658(3) | -0.4254(4) |
| 2F1M | 3.614 | -0.1609(3) | -0.4315(9) | - | - |
| 1F2M | 3.614 | -0.1591(2) | -0.4269(6) | - | - |
| 0F3M | 3.614 | -0.1573(2) | -0.4220(6) | -0.1771(4) | 0.4228(4) |
| 2F2M | 3.614 | -0.1595(2) | -0.4278(4) | - | - |
| 1F3M | 3.614 | -0.1577(2) | -0.4231(6) | - | - |
| 0F4M | 3.614 | -0.1561(2) | -0.4188(4) | -0.2020(2) | 0.4209(2) |

Table 5.2: Non-local corrections for hydration free energies of hydrophobic cavities with different sizes.

| | | Non-polarizable model | | Polarizable model | |
|---|---|---|---|---|---|
| | $R$(Å) | $\Delta G_{KO}$(kcal/mol) | $\Delta G_{LJ}$(kcal/mol) | $\Delta G_{KO}$(kcal/mol) | $\Delta G_{LJ}$(kcal/mol) |
| 0F0M | 3.100 | -0.1804(1) | -0.4827(3) | -0.1351(2) | -0.4794(5) |
| 0F0M | 3.400 | -0.1704(1) | -0.4564(2) | -0.1268(1) | -0.4529(2) |
| 0F0M | 3.614 | -0.1615(3) | -0.4332(7) | -0.1199(3) | -0.431(2) |
| 0F0M | 4.100 | -0.1378(3) | -0.3705(7) | -0.1013(3) | -0.3680(9) |
| 0F0M | 4.300 | -0.1217(2) | -0.3430(5) | -0.0938(1) | -0.3422(3) |
| 0F0M | 4.500 | -0.1175(2) | -0.3166(6) | -0.0864(2) | -0.3158(4) |

Table 5.3: Non-local corrections for hydration free energies of hydrophilic cavities with different sizes.

| | | Non-polarizable model | | Polarizable model | |
|---|---|---|---|---|---|
| | $R$(Å) | $\Delta G_{KO}$(kcal/mol) | $\Delta G_{LJ}$(kcal/mol) | $\Delta G_{KO}$(kcal/mol) | $\Delta G_{LJ}$(kcal/mol) |
| 0F4M | 3.346 | -0.1644(3) | -0.4408(9) | -0.251(1) | -0.4439(5) |
| 0F4M | 3.413 | -0.1624(2) | -0.4354(4) | -0.238(1) | -0.4381(7) |
| 0F4M | 3.480 | -0.1605(2) | -0.4304(4) | -0.226(2) | -0.4330(7) |
| 0F4M | 3.547 | -0.1584(3) | -0.4250(9) | -0.216(2) | -0.4267(6) |
| 0F4M | 3.614 | -0.1561(2) | -0.4188(4) | -0.2020(2) | -0.4209(2) |
| 0F4M | 3.682 | -0.1539(2) | -0.4130(6) | -0.191(2) | -0.4142(6) |
| 0F4M | 3.749 | -0.1510(2) | -0.4055(4) | -0.1797(7) | -0.4071(6) |
| 0F4M | 3.817 | -0.1485(3) | -0.3989(8) | -0.1697(6) | -0.4003(6) |
| 0F4M | 3.885 | -0.1457(2) | -0.3913(6) | -0.1602(2) | -0.3925(4) |
| 0F4M | 4.158 | -0.1333(2) | -0.3586(5) | -0.1307(2) | -0.3584(5) |
| 0F4M | 4.337 | -0.1248(2) | -0.3359(5) | -0.1171(2) | -0.3363(5) |

Table 5.4: Hydrogen bonding information for cavities with different hydrogen bond forming molecules. Calculation is done with non-polarizable potential.[a]

| | $R$ | changes in | changes in | $\lambda$ = 1.0 | | | $\lambda$ = 0.0 |
|---|---|---|---|---|---|---|---|
| | (Å) | hydrogen bonds | hb3 | hb1 | hb2 | hb3 | hb3 |
| 0F0M | 3.614 | 0.00(0) | 0.0000(0) | 0.0000(0) | 0.0000(0) | 0.0000(0) | 0.0000(0) |
| 1F0M | 3.614 | 0.91(1) | 0.0000(0) | 0.0000(0) | 0.906(5) | 0.0000(0) | 0.0000(0) |
| 0F1M | 3.614 | 0.89(1) | 0.0000(0) | 0.066(6) | 0.82(2) | 0.0000(0) | 0.0000(0) |
| 2F0M | 3.614 | 1.86(2) | 0.0000(0) | 0.0000(0) | 1.86(2) | 0.0000(0) | 0.0000(0) |
| 1F1M | 3.614 | 1.84(4) | -0.010(2) | 0.37(3) | 1.48(4) | 0.0000(0) | 0.010(2) |
| 0F2M | 3.614 | 1.84(3) | -0.0008(4) | 0.51(2) | 1.33(3) | 0.0000(0) | 0.0008(4) |
| 2F1M | 3.614 | 2.91(1) | 0.0000(0) | 0.977(6) | 1.93(2) | 0.0000(0) | 0.0000(0) |
| 1F2M | 3.614 | 2.88(1) | -0.014(3) | 0.999(7) | 1.89(1) | 0.0002(4) | 0.014(3) |
| 0F3M | 3.614 | 2.79(2) | -0.0016(8) | 0.998(7) | 1.79(2) | 0.0000(0) | 0.0016(8) |
| 2F2M | 3.614 | 3.88(3) | -0.010(2) | 1.92(3) | 1.964(8) | 0.0000(0) | 0.010(2) |
| 1F3M | 3.614 | 3.84(1) | -0.028(2) | 1.89(1) | 1.971(4) | 0.0000(0) | 0.028(2) |
| 0F4M | 3.614 | 3.78(1) | -0.0044(5) | 1.838(5) | 1.943(3) | 0.0000(0) | 0.0044(5) |

[a] hb1 is the number of hydrogen bonds where the water molecule acts as a proton acceptor, hb2 is for that with water as a proton donor, and hb3 are hydrogen bonds formed between residue molecules.

Table 5.5: Hydrogen bonding information for cavities with different hydrogen bond forming molecules. Calculation is done with polarizable potential.

| | $R$ | changes in | changes in | $\lambda$ = 1.0 | | | $\lambda$ = 0.0 |
|---|---|---|---|---|---|---|---|
| | (Å) | hydrogen bonds | hb3 | hb1 | hb2 | hb3 | hb3 |
| 0F0M | 3.614 | 0.00(0) | 0.0000(0) | 0.00(0) | 0.00(0) | 0.00(0) | 0.0000(0) |
| 0F1M | 3.614 | 0.86(2) | 0.0000(0) | 0.10(2) | 0.76(2) | 0.00(0) | 0.0000(0) |
| 0F2M | 3.614 | 1.83(3) | -0.0004(5) | 0.80(2) | 1.04(2) | 0.00(0) | 0.0004(5) |
| 0F3M | 3.614 | 2.73(2) | 0.0000(0) | 1.12(2) | 1.61(1) | 0.00(0) | 0.0000(0) |
| 0F4M | 3.614 | 3.82(1) | -0.0028(7) | 1.885(7) | 1.940(6) | 0.00(0) | 0.0028(7) |

Table 5.6: Hydrogen bonding information for hydrophilic cavities with different sizes. Calculation is done with non-polarizable potential.

| | $R$ | changes in | changes in | $\lambda$ = 1.0 | | | $\lambda$ = 0.0 |
|---|---|---|---|---|---|---|---|
| | (Å) | hydrogen bonds | hb3 | hb1 | hb2 | hb3 | hb3 |
| 0F4M | 3.346 | 3.57(4) | -0.34(3) | 1.94(3) | 1.980(3) | 0.02(2) | 0.36(3) |
| 0F4M | 3.413 | 3.74(1) | -0.176(7) | 1.939(8) | 1.975(3) | 0.004(2) | 0.180(6) |
| 0F4M | 3.480 | 3.84(1) | -0.065(6) | 1.93(1) | 1.974(2) | 0.002(1) | 0.067(6) |
| 0F4M | 3.547 | 3.83(2) | -0.021(3) | 1.89(2) | 1.963(4) | 0.0006(5) | 0.022(3) |
| 0F4M | 3.614 | 3.78(1) | -0.0044(5) | 1.838(5) | 1.943(3) | 0.0000(0) | 0.0044(5) |
| 0F4M | 3.682 | 3.65(1) | -0.0022(4) | 1.75(1) | 1.904(7) | 0.0000(0) | 0.0022(4) |
| 0F4M | 3.749 | 3.44(1) | -0.0002(4) | 1.60(1) | 1.841(5) | 0.0000(0) | 0.0002(4) |
| 0F4M | 3.817 | 3.18(2) | 0.0000(0) | 1.42(2) | 1.758(7) | 0.0000(0) | 0.0000(0) |
| 0F4M | 3.885 | 2.89(2) | 0.0000(0) | 1.26(2) | 1.634(9) | 0.0000(0) | 0.0000(0) |
| 0F4M | 4.158 | 1.76(2) | 0.0000(0) | 0.73(2) | 1.03(2) | 0.0000(0) | 0.0000(0) |
| 0F4M | 4.337 | 1.35(1) | 0.0000(0) | 0.54(1) | 0.81(1) | 0.0000(0) | 0.0000(0) |

Table 5.7: Hydrogen bonding information for hydrophilic cavities with different sizes. Calculation is done with polarizable potential.

| | $R$ | changes in | changes in | $\lambda$ = 1.0 | | | $\lambda$ = 0.0 |
|---|---|---|---|---|---|---|---|
| | (Å) | hydrogen bonds | hb3 | hb1 | hb2 | hb3 | hb3 |
| 0F4M | 3.346 | 3.74(3) | -0.23(4) | 1.984(5) | 1.988(3) | 0.002(2) | 0.23(4) |
| 0F4M | 3.413 | 3.85(1) | -0.11(2) | 1.972(4) | 1.987(1) | 0.0004(5) | 0.11(2) |
| 0F4M | 3.480 | 3.91(1) | -0.032(6) | 1.96(1) | 1.980(3) | 0.0000(0) | 0.032(6) |
| 0F4M | 3.547 | 3.90(1) | -0.008(2) | 1.94(2) | 1.966(4) | 0.0002(4) | 0.008(2) |
| 0F4M | 3.614 | 3.82(1) | -0.0028(0) | 1.885(7) | 1.940(6) | 0.0000(0) | 0.0028(7) |
| 0F4M | 3.682 | 3.70(3) | -0.0008(4) | 1.81(3) | 1.89(2) | 0.0000(0) | 0.0008(4) |
| 0F4M | 3.749 | 3.28(2) | 0.0000(0) | 1.68(3) | 1.80(2) | 0.0000(0) | 0.0000(0) |
| 0F4M | 3.817 | 3.18(3) | 0.0000(0) | 1.51(3) | 1.67(2) | 0.0000(0) | 0.0000(0) |
| 0F4M | 3.885 | 2.83(1) | 0.0000(0) | 1.315(7) | 1.515(6) | 0.0000(0) | 0.0000(0) |
| 0F4M | 4.158 | 1.67(1) | 0.0000(0) | 0.74(2) | 0.923(8) | 0.0000(0) | 0.0000(0) |
| 0F4M | 4.337 | 1.28(2) | 0.0000(0) | 0.57(2) | 0.71(2) | 0.0000(0) | 0.0000(0) |

Table 5.8: Occurrence probabilities and the force filed parameters for 27 types of atoms, the unit for $\sigma$ is Å and for $\epsilon$ is kcal/mol.

| atom type $i$ | $w_i$ (%) | $\sigma_i$ | $\epsilon_i$ | atom type $i$ | $w_i$ (%) | $\sigma_i$ | $\epsilon_i$ |
|---|---|---|---|---|---|---|---|
| HS | 0.1190 | 1.0740 | 0.0157 | HP | 0.7392 | 1.9690 | 0.0157 |
| HO | 0.9961 | 0.0000 | 0.0000 | HC | 25.5842 | 2.6617 | 0.0157 |
| HA | 2.3742 | 2.6116 | 0.0150 | O2 | 1.4534 | 2.9735 | 0.2100 |
| C* | 0.0877 | 3.4153 | 0.0860 | NA | 0.3759 | 3.2650 | 0.1700 |
| CB | 0.0877 | 3.4153 | 0.0860 | C | 7.7304 | 3.4153 | 0.0860 |
| CN | 0.0877 | 3.4153 | 0.0860 | CC | 0.1441 | 3.4153 | 0.0860 |
| H | 10.0984 | 1.0740 | 0.0157 | CA | 3.1385 | 3.4153 | 0.0860 |
| O | 6.8032 | 2.9735 | 0.2100 | N | 6.8032 | 3.2650 | 0.1700 |
| S | 0.1378 | 3.5800 | 0.2500 | CR | 0.1441 | 3.4153 | 0.0860 |
| N2 | 0.9585 | 3.2650 | 0.1700 | N3 | 0.3696 | 3.2650 | 0.1700 |
| CW | 0.2318 | 3.4153 | 0.0860 | CT | 19.8897 | 3.4153 | 0.1094 |
| OH | 0.9961 | 3.0806 | 0.2104 | H1 | 10.1547 | 2.4827 | 0.0157 |
| H4 | 0.2318 | 2.5221 | 0.0150 | H5 | 0.1441 | 2.4326 | 0.0150 |
| SH | 0.1190 | 3.5800 | 0.2500 | | | | |

# Vita

Hongtao Yu was born in Tongxu, Henan, P. R. China. He received a Bachelor of Science degree in 2005 from the Department of Modern Physics of the University of Science and Technology of China. He entered the Department of Chemistry of the University of New Orleans in the fall of 2006 and joined Steven Rick's group.