Thermodynamics of Water at Biomolecular Interfaces

by

Zheng Li

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy

The City University of New York

2006

UMI Number: 3231976

UMI Microform 3231976

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the
Graduate Faculty in Chemistry in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Themis Lazaridis

Date — Chair of Examining Committee

Michael Green

Date — Executive Officer

Marilyn Gunner

Ruth E. Stark

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Thermodynamics of Water at Biomolecular Interfaces

by

Zheng Li

Advisor: Professor Themis Lazaridis

Water molecules are often found at the binding interface of biomolecular complexes mediating the interaction between polar groups or simply filling space. Recent experimental or theoretical studies have demonstrated the importance of taking such water molecules into account in considering the thermodynamic parameter changes in complex formation. In my doctoral research studies, we calculated the different thermodynamic contributions of such water molecules in several complexes and discussed whether the displacement of water molecule would be favorable in ligand design. We calculated the contribution of the strongly bound water molecules at the interface of three complexes (HIV-1 protease-KNI 272, Concanavalin A-trimannoside 1 and Cyclophilin A-Cyclosporin A and analog) to the thermodynamic properties using statistical mechanical formulas for the energy and entropy. The requisite correlation functions were obtained by molecular dynamics simulations. We find that the entropic penalty of ordering is usually large but is outweighed by the favorable water-protein interactions. We also find large negative contributions from some of these water molecules to the heat capacity. We also considered the thermodynamic consequences of

displacement of the bound water molecule by ligand modification, including ligand desolvation, ligand conformational entropy loss and protein-ligand interactions. Our approach could be useful in rational drug design by estimating which bound water molecules would be most favorable to displace, which might be also helpful in the improvement of scoring function in predicting the protein-ligand binding affinity by incorporating the contributions from bound water.

ACKNOWLEDGMENTS

First, and foremost, I would like to send special and heartfelt thanks to my mentor, Professor Themis Lazaridis, for being the special person he is. Thank you for your guidance, encouragement and personal attention.

Thanks also to Professors Marylin Gunner and Ruth E. Stark for their guidance and helpful discussions and to my colleagues in the lab for their technical assistance.

As always, my special thanks go to my parents and my wife, Lina Nie for being wonderful supporters and, more importantly, for the never-ending encouragement expressed throughout the years I worked toward Ph. D..

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

IFST	Inhomogeneous Fluid Solvation Theory
MD	Molecular Dynamics
Con A	Concanavalin A
Cyp A	Cyclophilin A
Cys A	Cyclosporin A
ABNR	Adopted Basis Newton-Raphson

Part I. Introduction

Crystal structures of biomolecules typically reveal many water molecules on the biomolecular surface, in crevices, and sometimes buried within the biomolecular interior or at binding interfaces. The water molecules at binding interfaces can modify the shape and flexibility of the protein binding site, improving the steric complementarity between protein and ligand, or mediate the binding between the biomolecules with a hydrogen bond network.[1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13]. Analysis of 109 protein-DNA complexes showed that 6% of the crystallographic water mediates the binding.[14] For instance, in the trp-repressor/operator complex, in addition to the direct interactions between the phosphates of DNA and the protein, six water-mediated polar contacts to the bases were observed,[1] which are the determinants of specificity in this system.[6] An analysis of 75 protein-protein complexes showed on average one water molecule per 100Å^2 of interface area.[15] A more recent analysis found that water-mediated polar interactions are as abundant as direct protein-protein hydrogen bonds.[16]

Several reviews are relevant to this topic. Levitt and Park surveyed the available information on the location and dynamics of water molecules in and around protein structures.[17] Ladbury focused on the role of water in protein-ligand binding and its potential applications in drug design.[18] Janin also covered water-mediated interactions in protein-protein and protein-DNA complexes and their effect on affinity and specificity.[19] Cozzini et al. described different computational methods used to predict protein-ligand binding free energy with some emphasis on the contributions of water molecules.[20] Jayaram and Jain reviewed the role of water in protein-DNA recognition.[21]

Despite much work, many issues regarding interfacial water molecules remain unclear: Are water-mediated interactions equally strong as direct interactions? Do they provide more or less specificity? How does the presence of interfacial water molecules affect the enthalpy, entropy, and heat capacity of binding? What is the best way to incorporate water-mediated interactions in docking and drug design? In this section, we will review recent work that addresses some of these questions.

1.1 Experimental studies of the thermodynamics of interfacial water

The standard approach for studying the effect of interfacial water experimentally is to introduce modifications to the ligand or the protein that will introduce or displace it and measure the effect on binding thermodynamics. In some cases, ligands designed to displace the water molecules exhibit higher binding affinity. The best known example can be found among HIV-1 protease-inhibitors. Several water molecules were observed crystallographically at the binding interface between HIV 1-protease and KNI-272 or its analogs[22; 23; 24; 25; 26]. Cyclic urea inhibitors designed to displace and mimic the interactions of one of the bound water molecules were found to bind more strongly to the protein.[27; 28; 29] However, the cyclic urea inhibitors are entirely different from KNI-272, and therefore it is not certain that the gain in binding affinity really comes from water displacement. More clear-cut is the example of scytalone dehydratase. The crystal structure of this protein complexed with a salicylamide inhibitor revealed two water molecules at the binding interface. Analogs of this inhibitor with an additional nitrile group were found to displace one of the crystallographic water molecules and to have higher inhibitory potency.[30]

Crystal structures of OppA-dipeptide complexes revealed several ordered water molecules mediating the interactions between ligand and protein.[7; 10] These water molecules are displaced when the ligand changes from dipeptide to tri- and tetrapeptides. Isothermal titration calorimetry measurements of the binding of the peptides with different lengths to OppA indicate that the dipeptide is bound with about 60-fold lower affinity than related tri- and tetrapeptides.[10] The higher affinity for tri or tetrapeptides versus dipeptides is due to a more favorable binding entropy, which is only partly offset by a more unfavorable enthalpy. The favorable entropy was suggested to result from the displacement of the ordered water.[10]

The streptavidin-HABA (2-[(4'-dydroxyphenyl)-azo]-benzoate) crystal structure shows an immobilized water molecule bridging two hydroxyl groups. Analogs of HABA with additional aliphatic groups displaced this water molecule and exhibited higher binding affinity, due to a more favorable binding entropy. The enhanced affinity was attributed to water displacement and ligand flexibility.[31]

Higher binding affinity by displacement of interfacial water molecules can also be achieved by protein mutations. Crystallographic studies revealed that two water molecules bound to the tyrosine-82 hydroxyl group in unliganded wild type FKBP-12 are displaced upon formation of a complex with FK506.[32] In the Y82F mutant of the protein no ordered water molecules are observed. Thermodynamic measurements showed that the enthalpy of binding is 4.2 kcal/mol more negative for the mutant, but partial compensation by entropy led to a slightly more negative binding free energy (0.6 kcal/mol) for the mutant.[33] A simulation study of the E60A mutant found a reorientation

of the water and led to an intuitive prediction that the mutant will exhibit reduced binding affinity.[34]

In the examples given above, water displacement was found to be beneficial for binding. In other cases, however, the displacement of interfacial water weakens protein-ligand binding. The structure of the complex of concanavalin A with a trimannoside shows a conserved water molecule bridging the protein and the ligand with several hydrogen bonds.[35] To study the role of this water molecule in binding, an analog of the trimannoside was designed to replace the hydroxyl at C-2 of the central mannose with a hydroxyethyl group, which was expected to displace and mimic the interactions of this water molecule with the protein.[36] Calorimetric measurements showed that this displacement leads to a more favorable binding entropy, but also a more unfavorable and larger enthalpy term and thus weaker binding.[36] A crystallographic study of cyclosporin A bound to cyclophilin A revealed a cluster of water molecules mediating the protein-ligand binding.[3] One of the bound water molecules was found displaced and others reorganized in going from cyclosporin A to (5-hydroxynorvaline)-2-cyclosporin.[37] The binding affinity decreased by 1.3 kcal/mol.

Sharrow et al. studied the thermodynamic contribution of the hydrogen bond network formed between the interfacial waters and the mouse urinary protein and its ligands employing calorimetry and mutational studies.[38] Mutation of a protein Y to F disrupts a hydrogen bond between one bound water molecule and the protein and leads to a loss of enthalpy, which is only partially compensated by a favorable entropy change.

Several experimental thermodynamic studies investigated the role of interfacial water in antibody-antigen binding using site-directed mutagenesis. In mutational studies

of the antilysozyme-antibody D1.3, the interfacial waters in two of the mutant complexes (Y50S and W92D) appear to reorganize to partially alleviate the loss of antibody-antigen interactions.[39] Still, the binding is weakened. Yokota et al. constructed mutants of an antilysozyme antibody changing Y to F and S to A.[40] Crystallographic studies showed that the interfacial water molecules mediating the binding were not affected by the mutations. Instead, the mutations create an unfavorable gap between the waters and the antibody, which induces the loss of binding affinity due to loss of binding enthalpy. An Arg to Lys mutant of a different antilysozyme antibody introduces an additional water molecule at the binding interface and leads to a 1000-fold loss of binding affinity.[41]

In other studies, the introduction of water molecules at the interface by mutation was found to be energetically neutral. Watson et al. found similar binding affinities for different inhibitors of glycogen phosphorylase regardless of the presence of water-mediated interactions.[42] Dall' Acqua et al. found that an Asp18->Ala mutation introduced 3 water molecules at the interface but had negligible impact on the binding affinity.[43]

Mutagenesis has also been used to explore the contribution of bound water molecules to protein stability. One finding was that incorporation of water into an internal protein cavity is energy neutral.[44] Another study found that a water-mediated interaction is equivalent in the folded and unfolded states.[45] Another group concluded that protein-water hydrogen bonds contribute less to stability than protein-protein hydrogen bonds.[46; 47]

Denisov et al. performed an entropy analysis on experimental bound waters in BPTI. The amplitude and anisotropy of three water molecules in the protein cavities were assessed based on the nuclear magnetic relaxation dispersion (NMRD) data. The results

indicated that the binding of some of these water molecules are entropically driven even when they formed three or four hydrogen bonds with the protein.[48]

Several experimental thermodynamic studies focused on water mediating protein-DNA interactions. Osmotic stress has shown the importance of water molecules for specific DNA recognition by EcoRI.[49] The same approach applied to the complexes of the tryptophan and lac-repressor with their operators, showed a number of water molecules at the protein-DNA interface in good agreement with the number determined crystallographically.[50; 51]

Isothermal titration calorimetry has been used to measure the thermodynamics of a TATA-box binding protein with TATA-box and found that the bridging water molecules at highly polar biomolecular surfaces, as well as those in the binding interface, contribute substantially to the negative heat capacity change upon protein-DNA association.[52] This complements earlier results on the contribution of interfacial water molecules to the heat capacity of binding.[53; 54]

1.2 Prediction of interfacial water in biological systems

The identification and localization of water molecules in the crystal structures of biomolecules can be problematic due to crystallographic uncertainties. Some of the interfacial water in biological systems can be artifacts, particularly when they do not hydrogen-bond to the biomolecules.[55] Thus, in many instances, one needs to determine the location of bound water molecules theoretically. Several methods have been developed for this purpose. AQUARIUS[56] is a knowledge-based approach specially aimed at identifying water sites in protein from the electron density maps generated by

protein crystallography. GRID was developed to identify binding sites for water and other functional groups.[57] It involves calculating the interaction energy between the probe and the protein using an empirical energy function based on Lennard-Jones, electrostatic and hydrogen-bond terms. GRID has been reported to perform well in predicting protein-ligand interfacial water.[58; 59; 60] MCSS determines the location of functional groups via molecular dynamics simulations of multiple copies of such groups;[61] SuperStar is an empirical method based on the thermodynamics of small molecules and their crystal structures which treats protein binding site residues as small molecules or small structure fragments.[62] CS-Map predicts the most favorable binding positions of probes on a protein surface by calculating an interaction potential that accounts for van der Waals, electrostatic, and solvation contributions. The last two were calculated by a an implicit solvation model.[63] A new version of the Fold-X force field allows the prediction of the position of bound water molecules and can pick up 76% of water molecules interacting with at least two polar atoms of biomolecules in crystal structures with an average uncertainty of 0.8 Å.[64]

Garcia-Sosa et al. performed a multivariate regression analysis to establish a statistical correlation between the structural properties of water molecules in the binding site of a free protein crystal structure and the probability of observing the water molecules in the same location in the crystal structure of the complex.[65] The B-factor, the solvent contact surface area, and total hydrogen bond energy and the number of protein water contacts were found to correlate with the conservation of an ordered water molecule upon complex formation.

1.3 Interfacial water in scoring, docking, and ligand design

Interfacial water molecules have begun to be taken into consideration in *de novo* drug design, protein-ligand docking programs and the development of scoring functions for prediction of binding affinity.[60; 65; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75]

A computational approach based on the HINT empirical energy function was developed to study the binding and energetic roles of water molecules located in the binding interface of 23 HIV 1-protease ligand complexes.[60] The correlation coefficients between the calculated HINT scores and the experimental binding protein-ligand affinity of these complexes were significantly improved by taking the contributions of water into account. HINT was also used to estimate the energetics of water molecules in a wider variety of binding sites.[73] The water molecules bridging protein and ligand were found to have the largest binding energies (–1.3 kcal/mol). The HINT score together with the number and quality of hydrogen bonds of water molecules at protein active sites were suggested as criteria for predicting whether they are favorable or not to be displaced by ligand.

Inclusion of the interfacial water molecules in deriving both binding site structure- and ligand- based pharmacophores shows a profound effect on ligand design.[70] Wang et al. used the number of bridging water molecules in a QSAR-type predictor of the binding affinity of tyrosine phosphatase inhibitors.[68] Mancera found that inclusion of explicit water greatly facilitated *de novo* ligand design for bacterial neuraminidase.[75]

A few docking programs have developed ways to include discrete water molecules. Rarey et al. developed an approach for taking water into account in the docking program

FlexX. The water molecule is represented by a particle, which can form at most four hydrogen bonds. Possible positions for a water molecule in the protein binding site are precomputed is normalized for reasonable scoring and geometry parameters.[66] The geometry of the hydrogen bonds and the steric constraints resulting from the water is used to optimize the orientation of the ligand during the docking process. Osterberg treated water as a variable in an ensemble of protein structures which are combined to produce a single target representation for use with Autodock.[67] Schnecke and Kuhn predicted possible water molecules from the crystal structure of the ligand-free protein and considered their effects on molecular docking allowing them to interact with the ligands or incorporating a desolvation penalty into the scoring function when they are displaced by ligand atoms.[76] Verdonk et al. extended the program GOLD to include the bound water molecules in protein-ligand docking. In this approach, the water molecules in all-atom model are allowed to rotate and switch between the bound and displaced status during the docking protocol. The loss of rigid-body entropy of water is added to the scoring function as a penalty when water is present, thus, rewarding water displacement.[72] The program GLIDE docks explicit water molecules into the binding site and employs empirical scoring terms to account for their impact.[77]

Significant improvements in docking performance by considering the contributions from interfacial water have been reported.[66; 67; 71; 72; 76; 78; 79] In general, when crystallographic or even predicted water molecules are retained in protein structures, docking performance is improved.[71; 72; 76; 78; 79] These water molecules can usually accept two and donate two hydrogen bonds at most in the complex. However, other studies found that including water molecules had little effect on the quality of docking.[80]

A "solvated rotamer" approach was introduced by Jiang et al. to account for water-mediated hydrogen bonds in designing protein-protein interfaces.[81] They expanded a rotamer library by building fixed water molecules onto polar-side chain atoms, as well as backbone atoms. These rotamers were used to calculate protein-water interactions that include a Lennard-Jones repulsive potential and hydrogen bonding potential. It appears that the simple water-mediated hydrogen-bonding model can significantly improve prediction of amino acid. This model can also be applied to calculate protein-protein interaction energies. However, the predicted water molecules are not as helpful as crystallographic water molecules in predicting interaction energies. One problem with the approach is the increased combinatorial complexity of the new rotamer library.

Wolynes and coworkers derived direct and water mediated contact potentials for protein and found that the combination of the two performs best in discriminating native protein-protein interfaces.[82] In more recent work, water-mediated contacts along with a new one-dimensional burial term improved the quality of protein structure prediction, especially for larger proteins.[83] Water-mediated potentials also improved the quality of β-sheets in α/β protein structure prediction.[84]

1.4 Theoretical studies of the thermodynamics of interfacial water

The simplest estimate of the entropy cost of tying up a water molecule into a protein cavity was given by Dunitz, based on the entropies of inorganic salts.[85] He gave an upper bound of the entropy cost as 7 cal/mol K (about 2 kcal/mol at 298 K). Similarly, an estimated upper bound of the enthalpy gain of water ordering was given as -3.8 kcal/mol.[18] Combining these two estimates one arrives at -1.8 kcal/mol for the free

energy of inclusion of a single water molecule in a protein cavity or binding interface. Nevertheless, these are simple, generic estimates which are of limited utility in practical drug design situations.

Covell and Wallqvist used a knowledge-based energy function, including crystal waters, to compute the effect of mutations at protein-protein interfaces on binding affinity. 25% of the total binding affinity was calculated to arise from the crystallographic water molecules at the interface.[86]

Petrone and Garcia calculated the contributions of bound water molecules in major histocompatibility complex (MHC) -peptide binding. This is a very interesting system because it is characterized by low specificity (promiscuity). They used an approximate method for calculating the chemical potential of water based on a Gaussian model for the water interaction energy distribution. They found that water that simply fills space can bind at the interface with a net gain in entropy relative to the bulk.[87]

An entirely different methodology, normal mode analysis, was used by Fischer et al. to measure the effect of an isolated and well-ordered water molecule on the vibrational entropy of bovine pancreatic trypsin inhibitor (BPTI).[88] The increase in entropy upon addition of the water molecule was decomposed into the librational contributions of the water (9.4 cal/mol K) and an increase in protein flexibility (4 cal/mol K). This work shows that the low frequency vibrational modes of the water are important for the protein-ligand binding processes. Later work found that the water-protein electrostatic interactions are key to this phenomenon.[89]

On a totally different theory level, a quantum-chemical *ab initio* method was used to study the geometry and energetics of six water-mediated base pairs in RNA including

their geometries and interaction energies.[90] In most cases, the geometry of the water mediated base pairs was similar to that observed in crystal structures. The energies of the water-mediated and direct base pairs were compared. It was found that the absolute energy per hydrogen bond was higher for water-mediated interactions in all cases except for the UA pair. Of course, these studies only give interaction energies and do not include entropic effects, unlike the other methods mentioned above. Another quantum mechanical study aimed to quantify the differential binding energies of water versus carbohydrates to concanavalin A, where the carbohydrates were modeled by methanol and the protein by a limited number of amino acid side chains.[91] The binding energy of water and methanol to a cluster including three key amino acid side chains of concanavalin A and the key water, was found similar to each other.

1.5 Free energy perturbation and thermodynamic integration studies of interfacial water

Free energy perturbation (FEP) and thermodynamic integration (TI) are rigorous methods for computing the relative free energy between two thermodynamic states. Over the past two decades they have been applied to a wide variety of biological problems, including molecular recognition and protein stability. Studies on the contributions of ordered water molecules to free energy employing FEP or TI usually involve a hypothetical process by which a bound water molecule is transferred from the binding site or cavity of the protein to the bulk.[92; 93; 94; 95; 96] Applying these techniques, Wade et al. found that the free energy of transferring water molecules from bulk solvent into protein cavities is positive for empty cavities and negative for cavities containing water molecules.[92] Roux et al. used free energy perturbation to study the stability of water molecules in bacteriorhodopsin.[93]

Tashiro and Stuchebrukhov calculated the thermodynamic properties of water molecules in a hydrophobic cavity around the catalytic center of cytochrome c oxidase employing thermodynamic integration.[97] Zhang and Hermans studied the energetics of buried water molecules in the subtilisin Carlsberg-eglin C complex by calculating the free energies of transferring water molecules into the protein cavities employing molecular dynamics simulations and a harmonic restraint potential for each water.[94] Their results show that such free energies for the hydrated cavities (with a negative value up to –8.3 kcal/mol) are dramatically different from those for the empty cavities (with a positive value up to +14.6 kcal/mol). Using FEP, Olano and Rick calculated the free energies of transferring water molecules from pure liquid into two different protein cavities, i.e., BPTI (hydrophilic) and the I76A mutant of barnase (hydrophobic), which were obtained as –4.7 and + 4.7 kcal/mol at 298 K.[96] They also calculated the entropy of transfer and found it to be unfavorable for the polar cavity and favorable for the nonpolar one.

Hamelberg and McCammon studied interfacial water molecules in two biological systems using the double-decoupling approach.[95] In this approach, the binding thermodynamics of two molecules (A and B) is calculated from the free energy of transferring B from solution to the gas phase and transferring B from the complex AB in solution to the gas phase.[98] The free energy change of the first process is calculated by simply removing B from the solution applying free energy perturbation, while the other one is evaluated by thermodynamic integration applying constraints on the coordinates of B. They applied this method to the crucial water molecule bridging HIV 1-protease and the ligand KNI-272 and a similar water molecule in the complex of trypsin with

benzyldiamine. The free energy of transfer of water into the protein-ligand complex was calculated to be -3.1 kcal/mol in the first system and -1.9 in the second.[95]

Archontis et al. used free energy simulations to study the binding affinity between glycogen phosphorylase (GP) and its inhibitors: glucopyranose spirohydantoin (hydan) and its analogues (m-hydan and n-hydan).[99] Both this study and crystallography show that in the binding interface of GP-hydan complex, one water molecule prefers to stay at a well defined position, while in the GP-m-hydan (or n-hydan) complexes this water prefers another position. The free energy profile for the translocation of the water between the two positions showed that the calculated binding free energies are sensitive to the preference of this interfacial water to occupy the two positions. A free energy decomposition analysis employing the thermodynamic integration method along with a restraining potential on the water showed the free energy of moving a water molecule from one position to another in the GP-hydan complex to be 3.0 kcal/mol.[99]

A method based on free energy simulations has been proposed to compute the number of water molecules in the interfaces of protein-protein complexes.[100] This approach determines the optimal number of water molecules from the condition that the chemical potential of water in the cavity be equal to that in the bulk. It has been applied to determine the water content in 211 interfacial cavities in 26 antigen-antibody complexes.

1.6 Overview of the present work

In our work, we calculated the contributions of interfacial water molecules in HIV-1 protease-KNI-272, Concanavalin A-carbohydrate and Cyclophilin A-ligand complexes to the thermodynamic properties using statistical mechanical formulas for the energy and entropy and tried to connect the thermodynamic contributions of interfacial water to the protein-ligand binding thermodynamics. We found that the entropic penalty of ordering is usually large but outweighed by the favorable water-protein interactions.

In the HIV 1 protease-KNI 272 complex, a large negative contribution from the bound water molecule to the heat capacity was obtained. This work showed that the theory can yield the thermodynamic contributions of ordered water molecules and qualitatively estimate which bound water molecules would be most favorable to displace. However, no connection to *binding* thermodynamics could be made in that work because the inhibitor that displaced the ordered water molecule (DMP450) was entirely different from KNI-272. This complicates the attribution of the binding free energy difference to one or a few factors.

In our next work, we studied the Con A-trimannoside complexes to make this connection. As introduced above**,** the structures of the ligands (trimannoside **1** and **2**) and their binding to the protein are very similar (Figs. 4 and 5). Therefore, only a few contributions to binding affinity will be different for the two ligands, namely, the contribution of the ordered water molecule in the Con A-trimannoside **1** complex, the direct interactions $\Delta E_{L\text{-}P}$ of **1** and **2** with the protein, the desolvation entropy and enthalpy of the ligands and the conformational entropy of the additional hydroxyethyl side chain in trimannoside **2**. All these contributions are calculated to obtain an estimate of the difference in binding enthalpy and entropy. We found that the above inventory of

contributions is consistent with experimental data and provides insights into the sensitive balance of factors that determine whether water displacement is favorable or not.

In another work, we extended the analysis to water molecules that are not fully buried or form a cluster. Such a water cluster is observed in complexes of cyclophilin A (CypA) with cyclosporin A (CsA, **1**) and analogues. CsA is an immunosuppressive drug, which can prevent graft rejection after transplant surgery by forming a complex with its soluble intracellular receptor protein CypA, which then interacts with calcineurin and inhibits its phosphatase activity. As introduced above, in the complex of CypA-CsA (**1**), several bound water molecules were found at the binding interface mediating the interactions between CypA and the Abu residue (L-a-aminobutyric acid) of CsA (**1**). Three of them (WTR5, 6 and 7) are well ordered and one is less tightly bound (WTR133). (5-hydroxynorvaline)-2-cyclosporin (**2**) (figure 8, compound 2) is a derivative of **1**, which was designed to form additional direct interactions with CypA. Displacement of two of the bound water molecules in the Abu pocket was observed going from **1** to **2**. X-ray crystal structures of these two complexes show that the two ligands have identical backbone conformation and the protein binding sites have similar geometry (figure 9). We calculated the contributions of each bound water molecule to the thermodynamic properties separately and compared them to those of other water molecules, revealing the different roles they play at the binding interface. The change in binding affinity by the ligand modification is accounted for by considering the contributions from the water clusters, ligand-protein interactions, ligand desolvation and ligand conformational entropy loss.

Part II. Methods

2.1 Inhomogeneous fluid solvation theory (IFST)[101; 102]

IFST is a statistical mechanical theory that provides expressions for the energy and entropy as functionals of the molecular correlation functions based on an inhomogeneous view of infinitely dilute solutions. This view considers the solute as fixed at the origin and uses spatially-dependent correlation functions, rather than the standard homogeneous correlation functions that are averages over the entire body of the fluid. This allows one to focus on the vicinity of the solute where most of the contributions to the solvation thermodynamics originate. The expression for the entropy is an infinite series which cannot be calculated exactly. Usually, only two-particle contributions to the entropy are considered and the contribution of three and higher-particle correlations is neglected. This is equivalent to the Kirkwood superposition approximation (KSA), i.e., $g^{(3)}_{1,2,3} = g^{(2)}_{1,2}\, g^{(2)}_{1,3}\, g^{(2)}_{2,3}$. This approximation has given good results for the entropy in simple fluids.[102] However, for cases where triplet correlations are important, the results of this approach should be viewed with caution. Obviously, for isolated water molecules, only one-body distribution functions are involved.

This theory has been applied to several systems: the excess entropy in pure liquid water;[103] solvation thermodynamics in simple Lennard-Jones and hard sphere fluids;[102] and the solvent reorganization energy and entropy of hydration of methane.[104]

In this approach, ordered water molecules in the binding interface of biomolecular complexes are considered as part of the solvent. Their contributions to the solvation free energy are decomposed into four components: the solute-solvent energy (E_{sw}), the solute-solvent entropy (S_{sw}), the solvent reorganization energy (ΔE_{ww}) and solvent

reorganization entropy (ΔS_{ww}). The latter are due to the difference in solvent-solvent interactions and correlations in the bulk and in the complex.

$$\Delta G_{solv} = E_{sw} + \Delta E_{ww} - T\,(S_{sw} + \Delta S_{ww}) \quad (1)$$

All of these components can be expressed as integrals over the solute-solvent correlation function $g_{sw}(\mathbf{r}, \omega)$ and solvent-solvent correlation function $g_{ww}^{inh}(\mathbf{r}, \mathbf{r'}, \omega, \omega')$.[104]

$$S_{sw} = -k\,\rho/\Omega * \int g_{sw}(\mathbf{r}, \omega)\ln g_{sw}(\mathbf{r},\omega)\, d\mathbf{r}\, d\omega. \quad (2)$$

$$E_{sw} = \rho/\Omega \int g_{sw}(\mathbf{r}, \omega)\, u_{sw}(\mathbf{r}, \omega) d\mathbf{r} d\omega \quad (3)$$

$$\Delta S_{ww} = -\frac{1}{2} k \frac{\rho^2}{\Omega^2} \int g_{sw}(\mathbf{r},\omega)[g_{sw}(\mathbf{r'},\omega')\{g_{ww}^{inh}(\mathbf{r},\mathbf{r'},\omega,\omega')\ln g_{ww}^{inh}(\mathbf{r},\mathbf{r'},\omega,\omega') - g_{ww}^{inh}(\mathbf{r},\mathbf{r'},\omega,\omega') + 1\} - \{g_{ww}^{o}(R,\omega^{rel})\ln g_{ww}^{o}(R,\omega^{rel}) - g_{ww}^{o}(R,\omega^{rel}) + 1\}]\, d\mathbf{r} d\mathbf{r'} d\omega d\omega' \quad (4)$$

$$\Delta E_{ww} = -\frac{1}{2}\frac{\rho^2}{\Omega^2} \int g_{sw}(\mathbf{r},\omega)[g_{sw}(\mathbf{r'},\omega') g_{ww}^{inh}(\mathbf{r},\mathbf{r'},\omega,\omega') - g_{ww}^{o}(R,\omega^{rel})]u_{ww}(R,\omega^{rel}) d\mathbf{r} d\mathbf{r'} d\omega d\omega' \quad (5)$$

where k is Boltzmann's constant, ρ is the density of bulk solvent, **r** and **r**' denote the position of two water molecules; ω, and ω' denote the orientation of these two water molecules with respect to the solute, each of which is expressed as three Euler angles; Ω is the integral over ω ($\Omega = 8\pi^2$), R is the distance between two water molecules, ω^{rel} are the five angles which describe the relative orientation of two water molecules, and $\Omega^{rel} = \int d\omega^{rel} = 32\pi^3$; $g_{ww}^{o}(R, \omega^{rel})$ and $g_{ww}^{inh}(\mathbf{r},\mathbf{r'},\omega,\omega')$ are the solvent-solvent correlation function in the pure solvent and in the complex, respectively; $u_{sw}(\mathbf{r}, \omega)$ and $u_{ww}(R,\omega^{rel})$ are water-solute and water-water potentials, respectively.

The integrals in eqs. 2 and 4 are over all space V. $g_{sw}(\mathbf{r}, \omega)$ is zero over the regions occupied by the solute; the only contributions come from regions occupied by the solvent. Because any integral over V can be split into a sum of integrals over distinct subregions ($V = \sum_i \nu_i$, $V' = \sum_j \nu_j$), the contributions of specific water molecules can be determined. As a result, eqs. 2 and 4 can be written as:

$$
\begin{aligned}
S_{sw} &= -k\,\rho/\Omega * \int_{\nu} g_{sw}(\mathbf{r}, \omega)\, \ln g_{sw}(\mathbf{r},\omega)\, d\mathbf{r}\, d\omega \\
&= -k\,\rho/\Omega \int_{\sum_i \nu_i} g_{sw}(\mathbf{r}, \omega)\, \ln g_{sw}(\mathbf{r},\omega)\, d\mathbf{r}\, d\omega \\
&= -k\,\rho/\Omega \sum_i \int_{\nu_i} g_{sw}(\mathbf{r}, \omega)\, \ln g_{sw}(\mathbf{r},\omega)\, d\mathbf{r}\, d\omega \\
&\equiv \sum_i S_{sw(i)}
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
\Delta S_{ww} &= -\frac{1}{2} k\rho^2/\Omega^2 \int_V d\mathbf{r} \int_{V'} d\mathbf{r}' g_{sw}(\mathbf{r},\omega)[g_{sw}(\mathbf{r}',\omega')\{g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega')\ln g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega') \\
&\quad - g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega') + 1\} - \{g_{ww}^{o}(R,\omega^{rel})\ln g_{ww}^{o}(R,\omega^{rel}) - g_{ww}^{o}(R,\omega^{rel}) + 1\}]\, d\omega d\omega' \\
&= -\frac{1}{2} k\rho^2/\Omega^2 \sum_i \int_{\nu_i} d\mathbf{r} \int_{\sum_j \nu_j} d\mathbf{r}' g_{sw}(\mathbf{r},\omega)[g_{sw}(\mathbf{r}',\omega')\{g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega')\ln g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega') \\
&\quad - g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega') + 1\} - \{g_{ww}^{o}(R,\omega^{rel})\ln g_{ww}^{o}(R,\omega^{rel}) - g_{ww}^{o}(R,\omega^{rel}) + 1\}]\, d\omega d\omega' \\
&= -\frac{1}{2} k\rho^2/\Omega^2 \sum_i \sum_j \int_{\nu_i} d\mathbf{r} \int_{\nu_j} d\mathbf{r}' g_{sw}(\mathbf{r},\omega)[g_{sw}(\mathbf{r}',\omega')\{g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega')\ln g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega') \\
&\quad - g_{ww}^{inh}(\mathbf{r},\mathbf{r}',\omega,\omega') + 1\} - \{g_{ww}^{o}(R,\omega^{rel})\ln g_{ww}^{o}(R,\omega^{rel}) - g_{ww}^{o}(R,\omega^{rel}) + 1\}]\, d\omega d\omega' \\
&\equiv \sum_i \sum_j \Delta S_{w(i)w(j)}
\end{aligned}
\tag{7}
$$

With the identities,

$$g_{sw}(\mathbf{r},\omega) = g_{sw}^{tr}(\mathbf{r})\, g_{sw}^{or}(\omega|\mathbf{r}) \tag{8}$$

$$g_{ww}(\mathbf{r},\mathbf{r}',\omega,\omega) = g_{ww}^{r}(\mathbf{r},\mathbf{r}')g_{ww}^{or}(\omega,\omega'|\,\mathbf{r},\mathbf{r}') \tag{9}$$

the integrals in each subregion (ν_i,ν_j) can be decomposed into a translational and an orientational contribution:

$$S_{sw(i)} = -k\rho \int_{\nu_i} g_{sw}^{tr}(\mathbf{r})\ln g_{sw}^{tr}(\mathbf{r})d\mathbf{r} - k\rho/\Omega \int_{\nu_i} g_{sw}^{tr}(\mathbf{r})d\mathbf{r} \int_{\nu_i} g_{sw}^{or}(\omega|\mathbf{r})\ln\, g_{sw}^{or}(\omega|\mathbf{r})d\omega \tag{10}$$

and,

$$\begin{aligned}\Delta S_{w(i)w(j)} = &-\frac{1}{2}k\rho^2 \int_{\nu_i} d\mathbf{r} g_{sw}^{r}(\mathbf{r}) \int_{\nu_j} d\mathbf{r}'[g_{sw}^{r}(\mathbf{r}')\{g_{ww}^{r,inh}(\mathbf{r},\mathbf{r}')\ln g_{ww}^{r,inh}(\mathbf{r},\mathbf{r}') - g_{ww}^{r,inh}(\mathbf{r},\mathbf{r}') + 1\} - \{g_{ww}^{r,o}(R)\\ &\ln g_{ww}^{r,o}(R) - g_{ww}^{r,o}(R) + 1\}] - \frac{1}{2}k\rho^2 \int_{\nu_i} d\mathbf{r} g_{sw}^{r}(\mathbf{r}) \int_{\nu_j} d\mathbf{r}'[g_{sw}^{r}(\mathbf{r}')g_{ww}^{r,inh}(\mathbf{r},\mathbf{r}')S_{ww}^{m}(\mathbf{r},\mathbf{r}') - g_{ww}^{r,o}(R)S^{o}(R)]\end{aligned} \tag{11}$$

with $S_{ww}^{m}(\mathbf{r},\mathbf{r}')$ and S^o defined as,

$$S_{ww}^{m}(\mathbf{r},\mathbf{r}') = \frac{1}{\Omega^2}\int g_{sw}^{or}(\omega\,|\,\mathbf{r})g_{sw}^{or}(\omega'|\,\mathbf{r}')g_{ww}^{or,inh}(\omega,\omega'|\,\mathbf{r},\mathbf{r}')\ln g_{ww}^{or,inh}(\omega,\omega'|\,\mathbf{r},\mathbf{r}')d\omega d\omega' \tag{12}$$

$$S^{o}(R) = \frac{1}{\Omega^{rel}}\int g_{ww}^{r,o}(\omega^{rel}\,|\,R)\ln g_{ww}^{r,o}(\omega^{rel}\,|\,R)d\omega^{rel} \tag{13}$$

If we further assume that $g_{sw}^{or}(\omega\,|\,\mathbf{r})$, $g_{sw}^{or}(\omega'|\,\mathbf{r}')$ are independent of **r** or **r**' within the small regions ν_i,ν_j, i.e., $g_{sw}^{or}(\omega\,|\,\mathbf{r}) \approx g_{sw}^{or}(\omega)$, $g_{sw}^{or}(\omega'|\,\mathbf{r}') \approx g_{sw}^{or}(\omega')$, eqs. 10 and 12 become,

$$S_{sw(i)}=-k\rho \int_{\nu_i} g_{sw}^{tr}(\mathbf{r})\ln g_{sw}^{tr}(\mathbf{r})d\mathbf{r} - k N_{wat(i)}^{\nu_i}/\Omega \int_{\nu_i} g_{sw}^{or}(\omega)\ln g_{sw}^{or}(\omega)d\omega \quad (14)$$

$$S_{ww}^{m}(\mathbf{r},\mathbf{r'}) = \frac{1}{\Omega^2}\int g_{sw}^{or}(\omega)g_{sw}^{or}(\omega')g_{ww}^{or,inh}(\omega,\omega'|\mathbf{r},\mathbf{r'})\ln g_{ww}^{or,inh}(\omega,\omega'|\mathbf{r},\mathbf{r'})d\omega d\omega' \quad (15)$$

where, $N_{wat(i)}^{\nu_i}$ is the number of water molecules in the region of space ν_i, ($N_{wat(i)}^{\nu_i}=\rho \int_{\nu_i} g_{sw}^{tr}(\mathbf{r})\, d\mathbf{r}$).

The translational correlation function and orientational correlation function were calculated as products of one-dimensional functions, e.g., $g_{sw}^{tr}(\mathbf{r}) = g_{sw}^{tr}(r)\, g_{sw}^{tr}(\theta')\, g_{sw}^{tr}(\phi')$ and $g_{sw}^{or}(\theta,\phi,\psi) = g_{sw}^{or}(\theta)\, g_{sw}^{or}(\phi)\, g_{sw}^{or}(\psi)$, where r, θ' and ϕ' are spherical coordinates of the water oxygen with respect to its average position; θ, ϕ and ψ are Euler angles. In this work, the bin sizes for the numerical evaluation of integrals are set to dr = 0.1 Å, dθ' = π/10, dϕ' = π/10 and dθ = π/10, dϕ = 2π/10, dψ = 2π/10. In eqs. 11 and 15, the translational solvent pair correlation function $g_{ww}^{r,inh}(\mathbf{r}, \mathbf{r'})$ and the orientational solvent pair correlation function $g_{ww}^{or,inh}(\omega, \omega'|\mathbf{r}, \mathbf{r'})$ are six-dimensional functions, which are very difficult to obtain directly from simulation. The Kirkwood Superposition Approximation (KSA) assumes that the inhomogeneous pair correlation function is equal to the bulk solvent pair correlation function and depends only on the distance and the relative orientation between two solvent molecules.

$$g_{ww}^{inh}(\omega,\omega',\mathbf{r},\mathbf{r'}) = g_{ww}^{o}(\omega^{rel},R) \quad (16)$$

It has been successfully employed to calculate the solvent reorganization enthalpy and entropy for methane in water.[104]

The KSA can be applied separately to the translational and orientational parts

$$g_{ww}^{r,inh}(\mathbf{r},\mathbf{r}') = g_{ww}^{r,o}(R) \tag{17}$$

$$g_{ww}^{or,inh}(\omega,\omega'|\mathbf{r},\mathbf{r}') = g_{ww}^{or,o}(\omega^{rel}|R) \tag{18}$$

Thus, eq.15 becomes,

$$S_{ww}^{m}(\mathbf{r},\mathbf{r}') = \frac{1}{\Omega^2}\int g_{sw}^{or}(\omega)g_{sw}^{or}(\omega')g_{ww}^{or,o}(\omega^{rel}|R)\ln g_{ww}^{or,o}(\omega^{rel}|R)d\omega d\omega' \tag{19}$$

In our calculations, R is set to the distance of the average positions of the two water molecules. For each ω, ω' and R we calculate the five relative angles of the two water molecules. As a result $S_{ww}^{m}(\mathbf{r},\mathbf{r}')$ has a unique value for each water pair. To simplify the calculation of the five-dimensional orientational function in bulk water $g_{ww}^{or,o}(\omega^{rel}|R)$, a factorization is employed.

$$g_{ww}^{or,o}(\omega^{rel}|R) = g(\theta_1,\theta_2|R)g(\theta_1,\chi_2|R)g(\theta_2,\chi_1|R)g(\chi_1,\chi_2|R)g(\phi|R)/[g(\theta_1|R)g(\theta_2|R)g(\chi_1|R)g(\chi_2|R)] \tag{20}$$

This factorization was found to underestimate the magnitude of the entropy in pure water (about 5%).[103] The error this brings to the final result for the solvent-solvent entropy in

the complex S_{ww} for any bound water molecule is no more than 0.3 cal/mol K in this study.

Eq. 7 is essentially the difference in entropy between a bound water and a water in the bulk. In our work, it has been proven to be more convenient to calculate the bound water terms and then subtract the entropy of bulk water. The entropy of bulk water calculated theoretically with similar formulas is –15.2 cal/mol K. The solvent reorganization entropy of one bound water molecule is,

$$\Delta S_{w(i)w} = \sum_j S^{trans}_{w(i)w(j)} + \sum_j S^{or}_{w(i)w(j)} + 15.2 \text{ cal/mol K} \tag{21}$$

where, the solvent-solvent translational part $S^{trans}_{w(i)w(j)}$ and the solvent-solvent orientational part $S^{or}_{w(i)w(j)}$ of the entropy of w(i) can be calculated by excluding the terms of bulk water in equation 11, i.e.,

$$S^{trans}_{w(i)w(j)} = -\frac{1}{2}k\rho^2 \int_{\nu_i \nu_j} g^r_{sw(i)}(\mathbf{r}) g^r_{sw(j)}(\mathbf{r'}) \{ g^{r,inh}_{w(i)w(j)}(\mathbf{r},\mathbf{r'}) \ln g^{r,inh}_{w(i)w(j)}(\mathbf{r},\mathbf{r'}) - g^{r,inh}_{w(i)w(j)}(\mathbf{r},\mathbf{r'}) + 1 \} d\mathbf{r} d\mathbf{r'} \tag{22}$$

$$S^{or}_{w(i)w(j)} = -\frac{1}{2}k\rho^2 \int_{\nu_i \nu_j} g^r_{sw(i)}(\mathbf{r}) g^r_{sw(j)}(\mathbf{r'}) g^{r,inh}_{w(i)w(j)}(\mathbf{r},\mathbf{r'}) S^m_{w(i)w(j)}(\mathbf{r},\mathbf{r'}) d\mathbf{r} d\mathbf{r'} \tag{23}$$

The solvation entropy ΔS_{solv} is the sum of the solute-solvent entropy and solvent reorganization entropy of all bound water molecules.

$$\Delta S_{solv} = \sum_i (\Delta S_{w(i)w} + S_{sw(i)}) \quad (24)$$

Thus, the contribution of specific water molecules (or regions of space) to the solvation entropy can be determined.

For isolated and fully buried water molecules, e.g., the key water molecules in HIV 1 protease- KNI 272 complex and concanavalin A-trimannoside 1 complex (see below), the interaction and correlation between them and the other solvent molecules are negligible. Therefore, the ΔS_{ww} is simply the entropy of removing a water molecule from bulk water $\Delta S_{w(i)w}$ = + 15.2 cal/mol K. Otherwise, for those water molecules that are not fully buried or form a water cluster, we need to first split the region occupied by these water molecules into distinct spherical regions (*i*) based on the average positions of each bound water molecule obtained from the MD simulation (within a radius 1.2 Å in our work). This cutoff value was selected because in bulk water, the nearest neighbor's distance is about 2.8 Å for oxygen-oxygen pair, and half of this value is 1.4 Å. Also, as shown in figure 9, the density of each bound water molecule decreases to 0 before r reaches 1.2 Å. Next, we calculate the occupancy $O(i)$ of each region. We identify any water molecule located in a region (*i*) with the corresponding bound water molecule [w(*i*)], i.e., we allow for exchange between water molecules. The solvent-solvent energy and entropy terms ($\Delta E_{w(i)w(j)}$ and $\Delta S_{w(i)w(j)}$) were calculated separately for each water pair. w(*i*)w(*j*) denotes the water-water pair composed of the bound water molecule in region *i* and any other water molecule close to it and lying in different subregions (*j*). The subregions (*j*) were defined by scanning a spherical space of a bound water molecule [w(*i*)] within a radius of

5 Å (the magnitude of the interaction energy E_{ww} is no more than 0.2 kcal/mol beyond this distance) and looking for locations of high water density.

The solute-solvent energy (E_{sw}), and the solvent reorganization energy (ΔE_{ww}) solvation energy are evaluated directly from the simulations. This section is based on the publication (Li, Z. and Lazaridis, T. *J. Phys. Chem. B* **2006**, *110*, 1464-1475)

2.2 Construction of Initial Structures and force field

The initial structures of the systems we studied are all based on the X-ray structures of the complexes: HIV-1 protease-KNI-272 (1HPX[22]), HIV-1 protease-DMP450 (1DMP[29]), cyclophilin A-cyclosporin A (1CWA[3]) cyclophilin A-ligand **2** (1MIK[37]) and concanavalin A - trimannoside **1** (1ONA[35]). The interfacial water molecules were included according to the crystal structure. The crystal structure concanavalin A (Con A)-trimannoside **1** shows that three of the Con A subunits (I, II, and III) are almost identical, while subunit IV is a little different. Here, we select subunit I and the trimannoside bound to it from the X-ray crystal structure. The bound water molecule bridging trimannoside **1** and Con A, trimannoside **1,** and the ions (Ca^2+ and Mn^2+) were included according to the crystal structure. No bonds of any kind were established between the ions and protein atoms. The Con A-trimannoside **2** complex was constructed based on the first one by removing the bound water and replacing the hydroxyl group in trimannoside **1** with a hydroxyethyl group.

2.3 MD simulations.

The statistical analyses introduced above are based on a large amount of samples providing dynamic and atomic insights into the internal motions of the interfacial water molecules in the biological systems. Molecular dynamics (MD) and Monte Carlo (MC) simulations are usually performed for this purpose.[3] In our work, the CHARMM22 force field[105] was used for the proteins and ligands**.** Partial charges for the ligands were obtained with QUANTA using CHARMM template charges. MD simulations were carried out with the program CHARMM. A 15 Å sphere of TIP3P water molecules was added around the active site and subjected to spherical stochastic boundary conditions.[106] Solvent molecules overlapping with the protein, ligands or bound water molecules were deleted. This procedure did not result in the insertion of any additional water molecules at the binding interface. The SHAKE procedure was used to constrain the bonds involving hydrogen. The integration time step of the MD simulations was 2 fs. A cutoff distance of 30 Å was applied for the computation of non-bonded interactions. First, potential energy minimizations were performed using the ABNR method. Then, MD simulations at 300 K were performed, starting from the energy-minimized system, lasting for 8 ns with the protein and the ligands (except trimannoside **1** and **2**, because their conformations were found to affect the motion and interaction of the bound water with the protein and ligand) kept fixed.

2.4 Calculation of Euler angles

According to Euler's orientation theorem, any orientation may be described using three Euler angles. Any rotation *A* can be written as

$$A = BCD = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (25)$$

where θ, φ, ψ are the Euler angles.

If X and X' are the coordinates of any point in the original and final coordinate system before and after the rotation *A*, respectively, they are related by

$$X = A^{-1} X' \quad (26)$$

To simplify the calculations, the oxygen atoms of the ordered water molecule in each frame were first translated to their original position. Next, a body-fixed transformation was performed with the *y*-axis on the bisector of HOH, the *x*-axis perpendicular to *y*-axis on the plane of water molecule, and the *z*-axis perpendicular to this plane. 6 equations were obtained for the coordinates of the two hydrogen atoms (X_{H1}, Y_{H1}, Z_{H1}, X_{H2}, Y_{H2}, Z_{H2}) in each frame:

$$\begin{vmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & \sin\theta\sin\phi \\ \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & -\sin\theta\cos\phi \\ \sin\theta\sin\psi & \sin\theta\cos\psi & \cos\theta \end{vmatrix} \begin{vmatrix} X_{H_{01}} \\ Y_{H_{01}} \\ Z_{H_{01}} \end{vmatrix} = \begin{vmatrix} X_{H_1} \\ Y_{H_1} \\ Z_{H_1} \end{vmatrix}$$

$$\begin{vmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & \sin\theta\sin\phi \\ \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & -\sin\theta\cos\phi \\ \sin\theta\sin\psi & \sin\theta\cos\psi & \cos\theta \end{vmatrix} \begin{vmatrix} X_{H_{02}} \\ Y_{H_{02}} \\ Z_{H_{02}} \end{vmatrix} = \begin{vmatrix} X_{H_2} \\ Y_{H_2} \\ Z_{H_2} \end{vmatrix}$$

where X_{H01}, Y_{H01}, Z_{H01} and X_{H02}, Y_{H02}, Z_{H02} are the coordinates of the two hydrogen atoms at the reference frame. Solving these equations, we obtained the Euler angles (θ, φ, ψ).

Part III. Results

3.1 Thermodynamic contributions of the key water in the binding interface of HIV 1 protease-KNI 272 complex

This section is based on the publication (Li, Z. and Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 6636-6637).

The crystal structures of the HIV-1 protease with a number of inhibitors show an ordered water molecule donating two hydrogen bonds to the inhibitor and accepting two hydrogen bonds from the protein "flaps"(figure 1).[22; 24; 25; 26] NMR studies showed that this water molecule has a long residence time.[23] In the work reported by Lam et al., cyclic urea inhibitors were designed to displace and mimic the interactions of this water molecule and were found to bind more strongly to HIV-1 protease.[27] The crystal structure of the HIV-1 protease-DMP450 complex shows the oxygen atom of the cyclic urea carbonyl group accepting two hydrogen bonds from the protein flaps.

We calculated the contributions of this water molecule to the energy, entropy and heat capacity of solvation using the inhomogeneous fluid theory. MD simulations were first performed at 300 K, followed by calculation of the translational correlation function g_{sw}^{tr} (r, θ', ϕ'), where r, θ' and ϕ' are spherical coordinates of the water oxygen with respect to its average position. The radial distribution function (g_{sw}^{tr}(**r**) averaged over θ' and ϕ') is shown in figure. 2. This function gives the local density relative to bulk water, which can be seen to be very high. From the Euler angles of this water molecule in each frame we also calculated an average orientational correlation function g_{sw}^{or}(θ, ϕ, ψ) over the region occupied by this water molecule (figure. 3 shows the distribution of each angle). The calculation of the integrals in equation 10 was done using either the full three-

dimensional functions g_{sw}^{tr} (r, θ', ϕ') and g_{sw}^{or}(θ, ϕ, ψ) or the factorization approximations g_{sw}^{tr} (r, θ', ϕ')=g_{sw}^{r} (r)g_{sw} (θ')g_{sw} (ϕ') and g_{sw}^{or}(θ, ϕ, ψ)= g_{sw} (θ)g_{sw} (ϕ)g_{sw} (ψ), with similar results. The values below are those obtained by the factorization approximations (table 1). The translational and the orientational contribution to the solute-solvent entropy were found to be –12.5 cal/mol K each, giving S_{sw} = –25.0 cal/mol K. The solute-solvent energy E_{sw} was calculated directly from the simulation to be –28.2 kcal/mol.

The calculation of the solvent reorganization terms in this case is facilitated by the fact that this water molecule does not interact significantly with other water molecules when bound to the protease (the calculated interaction energy E_{ww} is –0.3 kcal/mol). Therefore, the ΔE_{ww} and ΔS_{ww} are simply the energy and entropy of removing a water molecule from bulk water, i.e. ΔE_{ww} = +10.1 kcal /mol and ΔS_{ww} = +15.2 cal/mol K.[103] Therefore ΔE = –18.1 kcal/mol and ΔS = -9.8 cal/mol K. This value is larger than the upper bound estimated by Dunitz, showing the extremely high degree of ordering of this water molecule. The contribution of the water molecule to the solvation free energy at 300 K is –15.2 kcal/mol (using equation 1 with PΔV term negligible).

To examine the dependence of the calculated solvation properties on the protein/inhibitor configuration, we repeated the calculation on the structure obtained after 1 ns of MD simulation, with the inhibitor and the residues of the protein within a radius 13 Å free to move. We obtained S_{sw} = -23.3 cal/mol, E_{sw} = -28.4 kcal/mol, and ΔG_{solv} = -15.9 kcal/mol, quite similar to the values above.

To calculate the contribution of this water molecule to the heat capacity, we repeated the MD simulation at 330 K. We obtained S_{sw} = -24.5 cal/mol K and E_{sw} = -28.0 kcal/mol. From the properties of bulk water, we obtained ΔE_{ww} = +9.6 kcal /mol and

ΔS_{ww} = +13.5 cal/mol K. Therefore, ΔE= -18.4 kcal/mol and ΔS = -11.0 cal/mol K and the contribution of the water molecule to the heat capacity of solvation is,

$$\Delta C_p = \left(\frac{d\Delta H}{dT} \right)_p \cong \frac{\delta E_{sw}}{\delta T} + \frac{\delta \Delta E_{ww}}{\delta T} = 7 - 17 = -10 \text{ cal/molK} \quad (27)$$

An alternative calculation gives,

$$\Delta C_p = T \left(\frac{d\Delta S}{dT} \right)_p = -12 \text{ cal/molK} \quad (28)$$

The negative value of ΔC_p means that the decrease of solvent reorganization energy and entropy with temperature is faster than the decrease of the protein-water interaction energy and entropy. In other words, the interactions of the water molecule in the binding site are less susceptible to temperature than bulk water. This finding is in agreement with the proposal that ordered water molecules contribute to the negative heat capacity observed in protein-DNA complexes.[54]

For the protease/DMP450 complex, we calculated the interaction energy of the carbonyl group with the protein to be –16.9 kcal/mol. This value is more negative than ΔG_{solv} for the water molecule. However, the cost of desolvating a carbonyl group is about +5 kcal/mol. Therefore, the displacement of the key crystal water molecule by insertion of a carbonyl group per se should be unfavorable. On the other hand, two carbonyl groups interacting with the key water molecule are eliminated in going from KNI 272 to DMP450, reducing the desolvation cost. Consideration of the above contributions leads

to the conclusion that the specific way water displacement was achieved in this case was favorable for binding. Of course, because these two inhibitors are entirely different, one would need a comprehensive account of all other differences for a reliable attribution of the gain in binding affinity.

In conclusion, the highly ordered water molecule in the HIV-1 protease-KNI 272 complex has a large entropic penalty of ordering which, however, is outweighed by the favorable water-protein interactions, leading to a large negative value of the contribution to the solvation free energy. This water molecule also has a large negative contribution to the heat capacity. This work also shows that the displacement of the ordered water with a carbonyl group might improve the binding affinity although this is not clear in the present case due to the totally different structures between KNI 272 and DMP450.

3.2 Thermodynamic contributions of the key water in the binding interface of Con A-trimannoside 1 complex

This section is based on the publication (Li, Z. and Lazaridis, T. *J. Phys. Chem. B* **2005**, *109*, 662-670).

The MD simulations performed on Con A-trimannoside **1** and **2** complexes gave very similar conformations for the two ligands, consistent with the results of Clarke et al. (see figure 5, the RMS differences between the 8ns time-averaged MD simulation structures and the experimental structures for ligand **1** and **2** were 0.6 Å and 0.5 Å, respectively).[36]

We first calculated the contribution of the ordered water molecule to the energy, entropy and heat capacity of solvation using IFST (see figure 14). Similar to the calculations and analysis performed on the water molecule in HIV 1-protease-KNI 272

complex, the translational correlation functions of this water were first obtained as a product of the radial distribution function $g_{sw}^{tr}(r)$ ($g_{sw}^{tr}(\mathbf{r})$ averaged over θ' and ϕ'), $g_{sw}^{tr}(\theta')$ ($g_{sw}^{tr}(\mathbf{r})$ averaged over r and ϕ') and $g_{sw}^{tr}(\phi')$ ($g_{sw}^{tr}(\mathbf{r})$ averaged over r and θ') during different portions of the trajectory at 300 K (figure 6). The similarity of the distributions over the first and second half of the simulation suggests good convergence of these functions in 8 ns. We next calculated the orientation of the water molecule in each frame, described by three Euler angles, and from these the average orientational correlation function $g_{sw}^{or}(\theta, \phi, \psi)$ over the region occupied by this water molecule. Fig. 7 shows the distribution of each angle.

The calculation of the integrals in Eq. 10 was done using either the full three-dimensional functions $g_{sw}^{tr}(r, \theta', \phi')$ and $g_{sw}^{or}(\theta, \phi, \psi)$ or the factorization approximations $g_{sw}^{tr}(r, \theta', \phi') = g_{sw}^{r}(r) g_{sw}(\theta') g_{sw}(\phi')$ and $g_{sw}^{or}(\theta, \phi, \psi) = g_{sw}(\theta) g_{sw}(\phi) g_{sw}(\psi)$. The results are similar. For example, with bin sizes dr = 0.06 Å, dθ' = π/6, dϕ' = π/3 for integration of $g_{sw}^{tr}(r, \theta', \phi')$, we obtained S_{sw}^{r} = -11.2 cal/mol K using the factorization approximation and –10.4 cal/mol K without it, and with bin sizes dθ = π/36, dϕ = π/36, dψ = π/36 for integration of $g_{sw}^{or}(\theta, \phi, \psi)$, we obtained S_{sw}^{or} = -10.9 cal/mol K using the factorization approximations and –11.6 cal/mol K without it. The values of the thermodynamic parameters discussed below were obtained using the factorization approximations and the bin sizes dr = 0.06 Å, dθ' = π/6, dϕ' = π/3 and dθ = π/36, dϕ = π/36, dψ = π/36.

The contributions to solvation free energy from the ordered water molecule are listed in table 2. At 300 K, the translational contribution to the solute-solvent entropy is –10.4 cal/mol K and the orientational contribution –11.6 cal/mol K, giving S_{sw} = -22.0 cal/mol K. The solute-solvent energy E_{sw} was calculated directly from the simulation as -30.2 kcal/mol, of which -23.6 kcal/mol is due to h-bonds and -6.6 kcal/mol is due to longer range interactions (beyond first neighbors). The latter includes the interaction of water with the two ions, which is -9.2 kcal/mol, showing that the ions enthalpically stabilize the bound water, but other long range interactions destabilize it. The ions also lower the entropy of the bound water, but this reduction is much smaller than the favorable interaction energy (data not shown). The solvent reorganization energy is the water-water interaction energy in the complex plus the enthalpy of removing a water molecule from bulk water at 300 K, i.e. ΔE^{0}_{ww} = +10.1 kcal /mol. For the calculation of the solvent reorganization entropy, we neglect the contribution of water-water correlations in the complex (indeed, the calculated interaction energy E_{ww} is only +0.9 kcal/mol at 300 K). Therefore, ΔS_{ww} is simply the entropy cost of removing a water molecule from bulk water, ΔS^{0}_{ww} = +15.2 cal/mol K.[103] Thus, ΔE_{solv} = -19.2 ± 0.1 kcal/mol and ΔS_{solv} = - 6.8 ± 0.1 cal/mol K. This value is very close to the upper bound (7cal/mole K) estimated by Dunitz[85]. Noting that the PΔV term is negligible, the contribution of the ordered water molecule to the solvation free energy at 300 K is -17.2 ± 0.1 kcal/mol (from equation. 1). Clearly, the negative value of ΔG_{solv} originates mainly from the interaction of the bound water molecule with the protein and ligand.

To calculate the contribution of this water molecule to the heat capacity, we repeated the MD simulation at 330 K. We obtained S_{sw} = -21.3 cal/mol K, E_{sw} = -30.0 kcal/mol,

and E_{ww} = + 0.9 kcal/mol. From the properties of bulk water, ΔE^{o}_{ww} = +9.6 kcal /mol and ΔS^{o}_{ww} = +13.5 cal/mol K[103]. Therefore, ΔE= -19.5 ± 0.1 kcal/mol, ΔS = -7.8 ± 0.1 cal/mol K. The contribution of the water molecule to the heat capacity of solvation from simulation was calculated as –10±3 cal/mol K (using equation 27) or –10±1 cal/mol K (using equation 28) The negative value of ΔC_p is again consistent with the proposal that ordered water molecules contribute to the negative heat capacity of binding in some complexes[54]. Apparently, the decrease of the energy and entropy with temperature of the protein-water interaction in the Con A-trimannoside **1** complex is slower than that of the solvent reorganization energy and entropy. In other words, the interactions of the water in the binding site are less susceptible to temperature than bulk water.

In conclusion, the interfacial water molecule in Con A-trimannoside **1** complex, which is isolated and fully buried, as the water in the HIV-1 protease-KNI 272 complex, also has a large entropic penalty of ordering, which is outweighed by the favorable water-protein interactions, leading to a large negative value of the contribution to the solvation free energy. This water molecule also has a large negative contribution to the heat capacity.

3.3 Thermodynamic consequences induced by the displacement of interfacial water in Con A-trimannoside 1

This section is also based on the publication (Li, Z. and Lazaridis, T. *J. Phys. Chem. B* **2005**, *109*, 662-670).

The contribution of the ordered water molecule is only one of the factors affecting binding. Other factors include the desolvation of ligand (L) and protein (P), the

translational, rotational and vibrational entropy of L and P, the direct interactions between L and P, etc. In the case of Con A binding trimannoside **1** and **2**, the protein was found to undergo no conformational change.[107] NMR and MD simulations show that these two ligands have very similar conformational properties and the same binding modes of both complexes except for the substituent on the central mannose.[36] Because the two ligands are very similar, the difference in free energy of binding can be decomposed into four terms

$$\Delta\Delta G = \Delta\Delta G_{desolv} + \Delta\Delta G_{int} + \Delta E_{L\text{-}P} + \Delta\Delta G_{solv}\,(\text{ordered water}) \quad (29)$$

where $\Delta\Delta G_{desolv}$ is the difference of the desolvation free energy of the two ligands, $\Delta\Delta G_{int}$ is the "internal" entropic contributions (changes in the translational, rotational, and vibrational degrees of freedom of the protein and ligands), $\Delta E_{L\text{-}P}$ is the difference of the direct interactions between the two ligands and the protein, and $\Delta\Delta G_{solv}$ (ordered water) is the contribution of the ordered water molecule to the solvation free energy.

It is reasonable to assume that, in this case, the $\Delta\Delta G_{desolv}$ originates only from the difference in desolvation free energy of the hydroxyethyl group in **2** and the hydroxyl group it replaced in **1**. We also assume that $\Delta\Delta G_{int}$ originates only from the entropy cost ($-T\Delta S_{config}$) of constraining the hydroxyethyl moiety of trimannoside **2** in the complex. As a result, with equation 11, the $\Delta\Delta G$ can be expressed as:

$$\Delta\Delta G = \Delta\Delta H - T\Delta\Delta S \quad (30)$$

where

$$\Delta\Delta H = \Delta H_{solv}\,(\text{ordered water}) + \Delta E_{L\text{-}P} + \Delta\Delta H_{desolv} \quad (31)$$

$$\Delta\Delta S = \Delta S_{solv}\,(\text{ordered water}) + \Delta S_{config} + \Delta\Delta S_{desolv} \quad (32)$$

We further assume that only the central mannose contributes to the interaction energy difference between trimannoside **1** and **2**. The difference in interaction energy of the C-2 substituents of the two ligands with the protein ($\Delta E_{L2OH\text{-}P}$) is -18.4 kcal/mol. The interaction energies of the ions (Ca^{2+} and Mn^{2+}) with the C-2 substituents of the two ligands $\Delta E_{L2OH\text{-}Ions}$ were also calculated. This interaction is more favorable for trimannoside **2** by 2.3 kcal/mol (table 3). This value is much smaller than the interaction of the ions with the ordered water in the first complex (-9.2 kcal/mol).

Another contribution to the difference in binding free energy is the different desolvation enthalpy and entropy of trimannoside **2** relative to trimannoside **1**. We assume that the solvation effects of polar and apolar groups of trimannoside **1** and **2** are additive and proportional to their solvent accessible surfaces[108]:

$$\Delta S_{apl}^{dehy} = \sum_i \Delta\widehat{S_{apl.i}^{dehy}}\Delta ASA_{apl.i} \quad (33)$$

$$\Delta S_{pol}^{dehy} = \sum_k \Delta\widehat{S_{pol.k}^{dehy}}\Delta ASA_{pol.k} \quad (34)$$

where $\Delta\widehat{S_{apl}^{dehy}}$ and $\Delta\widehat{S_{pol}^{dehy}}$ are the dehydration entropies of the apolar and polar groups per unit of surface area. Similar equations apply to the dehydration enthalpies and free

energies. The analytic surface area method in CHARMM was used to calculate the polar (oxygen) and apolar (carbon) accessible surface area (ASA) of mannose 2 of the two ligands in the complexed and free states. Table 4 shows that trimannoside **2** buries not only more apolar but also more polar surface area than trimannoside **1**. This is because the hydroxyl group in trimannoside **2** is more solvent exposed than that in trimannoside **1**. For the normalized dehydration enthalpies and entropies we used the values given by Makhatadze and Privalov for aliphatic surface and the polar parts of the serine side chain, i.e., 0.03 and 0.25 kcal/mol $Å^2$, respectively, for the enthalpy, 0.14 and 0.24 cal/mol K $Å^2$, respectively, for the entropy, and -0.01 and 0.18 kcal/mol $Å^2$, respectively, for the free energy.[108; 109] The differences in desolvation entropy, enthalpy and free energy of mannose 2 for the two ligands were calculated as +5.1 cal/mol K, +2.9 kcal/mol and +1.4 kcal/mol, respectively. The positive value of these quantities is due to the larger amount of polar surface area buried for trimannoside **2**. Employing equations 27 and 28, the heat capacity of dehydration of –OH and $–CH_2$-CH_2-OH as –9 cal/mol K[103] and –33.4 cal/mol K[110; 111], respectively, we calculated the desolvation entropy and enthalpy of the two ligands at 330 K as $\Delta\Delta H_{desolv}$(330 K) = +3.6 kcal/mol and $\Delta\Delta S_{desolv}$ (330 K) = +7.5 cal/mol K (table 5).

We also calculated the entropy of conformational restriction of the hydroxyethyl side-chain of trimannoside **2**. Empirical scales give an average value of conformational entropy of about 1.5 cal/mol K per rotatable bond[112; 113; 114; 115; 116; 117]. In this work, the conformational entropy change for the hydroxyethyl side-chain of ligand **2** upon protein-ligand binding was estimated by comparing the distributions of the dihedral angles in the complex and in the free ligand. We performed rigid rotations of the three dihedral angles

in the hydroxyethyl side-chain of ligand **2** and one dihedral angle of ligand **1** and obtained the probability (p) from the Boltzmann expression. Then we used the integrals of $p\ln(p)$ to estimate the entropy change, which was ΔS_{config} = -2.9 cal/mol K, consistent with the empirical scales.

Table 5 lists all contributions involved in the binding enthalpy and entropy that are different between the two ligands. The calculated difference in thermodynamic parameters of binding and the corresponding experimental values are summarized in table 5. The entropy contribution of the displacement is +9.0 cal/mol K, qualitatively comparable with the experimental data (+3.7 cal/mol K). The enthalpy contribution (+3.7 kcal/mol) also compares favorably with the experimental data (+2.3 kcal/mol). The overestimates of the enthalpy and entropy compensate to give a difference in binding affinity very close to experiment (+1.0 kcal/mol, vs. +1.2 kcal/mol from experiment, table 6). The positive value for $\Delta\Delta C_{p,}$ which is also in agreement with experiment[36], is largely due to the negative contribution of the ordered water molecule in trimannoside **1**.

In conclusion, except for the contribution of the water molecule to the solvation free energy, other contributions to the binding affinity difference of the two ligands, including desolvation, entropy of conformational restriction, and interaction between the ligand and protein, are also significant. The final outcome of the water displacement is sensitive to each these contributions and the details of the binding site.

3.4 Thermodynamic contributions of water molecules in the binding interface of Cyp A-ligand complexes

This section is based on the publication (Li, Z. and Lazaridis, T. *J. Phys. Chem. B* **2006**, *110*, 1464-1475).

The crystal structure of CypA-**1** complex shows four water molecules found in the cavity called "Abu pocket" interacting with each other with hydrogen bonds, i.e., WTR5 hydrogen bonds with Ala101, Gln111, Gly109 and WTR6; WTR6 hydrogen bonds with Thr107 WTR 5, 7 and 133; WTR7 hydrogen bonds with WTR6, WTR129, Gln111 (or SER110) and Gly74; WTR133 hydrogen bonds with WTR6 and other solvent water molecules around it. The crystal structure of CypA-**2** complex shows that WTR6 moves slightly out of the Abu pocket and presumably induces the displacement of WTR133; the location of the other two bound water molecules (WTR5 and 7) did not change much; no direct hydrogen bond was observed among these water molecules. [37]

In the MD simulations performed on complex **1** and **2**, WTR6 appears less frequently around its location in complex **1** than in complex **2**, where it is observed hydrogen-bonding with Thr107, the hydroxyl group of ligand **2** and some other solvent molecules, while WTR5 and 7 are more stable in both complexes. All these bound water molecules are found to hydrogen-bond to the protein and ligand as in the crystal structures (see figure 9).[37; 118] The number of hydrogen bonds formed between each bound water molecule and the protein and ligand observed from the last snapshot of the MD simulation is listed in tables 7 and 8. In the MD simulation, no hydrogen bond is formed between any of the four bound water molecules and ligand **1** in complex **1**, due to the hydrophobic nature of the Abu side chain in ligand **1**, while in complex **2**, either WTR 6 or 7 is found to make one hydrogen bond with ligand **2**.

The region occupied by each bound water molecule is defined as a spherical region of radius 1.2 Å centered at its average position obtained from MD simulation. In this way, we split the region occupied by a water cluster into separate spherical regions and studied the thermodynamic properties of water molecules in each region separately. From the MD simulations performed on CypA-**1**, the calculated occupancy (O) of the four bound water molecules in the Abu pocket decreases as: $O_{WTR5} > O_{WTR7} > O_{WTR6} > O_{WTR133}$ (Table 7). This order correlates with the number of hydrogen bonds formed between these water molecules and the protein mentioned above.

We first calculated the solute-solvent entropy of each water molecule in their corresponding regions. The translational correlation function with respect to the average position ($g_{sw}^{tr}(\mathbf{r})$) and orientational correlation function ($g_{sw}^{or}(\omega)$) were calculated as products of one-dimensional functions (figure 11). The results calculated with and without this approximation are similar. The radial distribution function for each water (figure 10) shows that WTR133 (with the lowest peak) is the least ordered with respect to the solute among 4 bound water molecules in complex **1**. We also find that WTR7 is much less ordered in complex **1** than in complex **2**, presumably due to the additional hydrogen bond formed between WTR7 and the hydroxyl group of ligand **2** (figure 9b).

Integrations over these correlation functions employing equation 10 give the solute-solvent contributions of each bound water molecule to the entropy $S_{sw(i)} = S_{sw(i)}^{trans} + S_{sw(i)}^{or}$ (tables 7 and 8). In complex **1**, WTR5 exhibits the most negative solute-solvent entropy S_{sw} and solute-solvent energy E_{sw}. E_{sw} and S_{sw} correlate with the occupancy and the number of hydrogen bonds but the correlation for S_{sw} is not perfect. In complex **2**, all bound water molecules considered have a more negative value of S_{sw} than WTR5 in

complex **1**, indicating that the involvement of the additional hydroxyl group in ligand **2** stabilizes the remaining bound water molecules in the Abu pocket.

To identify water pairs for the solvent-solvent entropy calculations, we determined the distribution of solvent molecules within a 5 Å sphere around each bound water molecule. The sphere was split into several subregions (*j*), each region corresponding to a specific water molecule (table 9). For instance, the water molecules around WTR7 within 5.0 Å in complex **1** approximately concentrate into 5 subregions whose corresponding water molecules are WTR5, WTR129, WTR6, solv54 and WTR133, respectively (fig. 12). The solvent occupancies in these subregions are 0.984, 0.909, 0.771, 0.711 and 0.302 respectively. Some of the subregions are not fully occupied, especially when the subregion is farther away from the corresponding bound water. Subregions with lower occupancy make a smaller contribution to the solvent-solvent entropy. Every bound water molecule [w(*i*)] together with a corresponding water molecule [w(*j*)] in a specific subregion were treated as a solvent pair and their solvent-solvent correlation functions were calculated.

Applying Eqs. 17-23, we calculated the solvent-solvent contributions of each bound water molecule to the entropy in complex **1** and **2**. The pure water pair correlation function ($g_{ww}^{or,o}(\omega^{rel}\,|R)$) was calculated by a 8ns MD simulation, employing eq. 20. This function was used to approximate $g_{ww}^{inh,o}(\omega,\omega'|\mathbf{r},\mathbf{r}')$ at the corresponding bin of Euler angles ω, ω' and R (R = $|\mathbf{r}-\mathbf{r}'|$). Eq. 19 was used to calculate the solvent orientational entropy in the complex $S_{ww}^{m}(\mathbf{r},\mathbf{r}')$. The solvent pair translational correlation function $g_{ww}^{r,inh}(\mathbf{r},\mathbf{r}')$ in the complex was approximated by the radial distribution function $g_{ww}^{or,o}(R)$ in the bulk (eq. 17). The solvent-solvent entropy was then calculated for each solvent pair

by eqs. 22 and 23. The results are listed in table 10. The solvent-solvent entropy of each pair of water molecules varies from –3.78 to +0.15 cal/mol K and reveals how correlated these two water molecules are. The slightly positive values are probably an artifact of the KSA approximation and the numerical integration. Small values of the pair entropy mean that the two water molecules are constrained by the protein or ligand in a conformation that disrupts interactions and correlations between them. S_{ww} in bulk water was calculated as –15.2 cal/mol K, and originates mainly from the water-water correlations in the first neighbor shell. For a given water molecule in bulk, there are about 4 solvent molecules in its first neighbor shell, therefore, the value of S_{ww} for each water-water pair in the bulk is approximately –3.8 kcal/mol K, which is very close to the most negative value of the solvent-solvent entropy of a water-water pair in the complex.

Summing up the pair entropies $S_{w(i)w(j)}$, we obtained the solvent-solvent entropy for each bound water molecule $S^{trans}_{w(i)w}$ and $S^{or}_{w(i)w}$, as listed in table 7 and 8. For instance, for WTR7 in complex **1**, $S^{or}_{w(7)w} = \sum_j S^{or}_{w(7)w(j)}$ =-3.5 cal/mol K, $S^{trans}_{w(7)w} = \sum_j S^{trans}_{w(7)w(j)}$ =-0.3 cal/mol K and $S_{w(7)w} = S^{or}_{w(7)w} + S^{trans}_{w(7)w}$ = -3.8 cal/mol K. The calculated results reveal that in complex **1**, WTR5 is relatively more correlated to other solvent molecules than WTR6, 7 and 133 and all of them have a more negative value of solvent-solvent entropy than those of the bound water molecules in complex **2**, where the values of solvent-solvent entropy for WTR5 and 7 are close to 0. These results show that WTR 5 and 7 in complex **2** are more isolated than the corresponding water molecules in complex **1**.

The contribution of each bound water molecule to the solvation entropy (ΔS_{solv}, see table 7 and 8) indicates that transferring WTR6 and 7 from the bulk to complex **1** is

accompanied by a very small entropy penalty compared to WTR5. The transfer of WTR133 is even entropically favorable, showing that it is even less ordered in complex **1** than in the bulk, which is related to the fact that it interacts weakly with the protein. In contrast, all three bound water molecules in complex **2** have a large negative contribution to the solvation entropy (ΔS_{solv}), showing that each of them is highly ordered in the complex. Notably, the total entropy of the three water molecules in **2** is more negative than that of the four water molecules in **1**. This is because the excess hydrogen bonds formed between the methyl hydroxyl group of ligand **2** with WTR6 and 7 constrain these water molecules, inducing a more negative entropy for each of them.

The solute-solvent energy $E_{sw(i)}$ and solvent-solvent energy $E_{w(i)w}$ of each bound water calculated directly from the simulations are listed in table 7 and 8. Part of the solute-solvent energy $E_{sw(i)}$ is due to long range interactions (about -4 kcal/mol), and the rest is due to hydrogen bonds formed with the protein and ligand. For example, the h-bond energy of WTR5, 6, 7, 133 is about –20, -5, -10, and 0 kcal/mol, respectively in complex **1**. These values reflect the number of hydrogen bonds formed between each bound water molecule and the protein-ligand complexes. The solute-solvent energy $E_{sw(i)}$ for WTR 5, 6, 7, 133 in complex **1** and WTR5 in complex **2** mainly originates from their interaction with the protein; their interaction with the ligand **1** or **2** is relatively weak (-0.1, -0.6, -0.1, -2.4 and –0.7 kcal/mol, respectively), showing that these water molecules essentially stabilize the conformation of the protein residues at the binding site, rather than bridging protein and ligand. WTR6 and 7 in complex **2** were found to interact more strongly with the ligand (-5.4 and -5.5 kcal/mol, respectively) than in complex **1**, due to their additional hydrogen bond with the hydroxyl group of ligand **2**.

The solvent-solvent energy of each bound water molecule shows strong interactions of WTR6 and WTR133 in complex **1** (-15.4 and -15.6 kcal/mol respectively) and WTR6 (-11.0 kcal/mol) in complex **2**, which are even comparable to the solvent-solvent interactions in bulk water (-20.2 kcal/mol). The solvent-solvent interactions of other bound water molecules are much weaker. The solvent-solvent energy between WTR5 and WTR7 is even positive in complex **2**, showing that these two bound water molecules are tightly fixed by the protein and ligand, leading to unfavorable water-water interaction. The contribution of a bound water molecule [w(*i*)] to the solvation enthalpy was calculated as

$$\Delta E_{solv} = E_{sw(i)} + 1/2E_{w(i)w} + 10.1 \text{ kcal/mol.} \quad (35)$$

where 10.1 kcal/mol is half of the solvent-solvent interaction energy in bulk water, which is equal to the energy cost of transferring one water molecule from bulk to the gas phase.

The contribution of each bound water molecule to the solvation free energy was calculated as the sum of the entropy and energy terms employing equation 1(the $P\Delta V$ term is negligible): The contribution of WTR5 to the solvation free energy in complex **1** was obtained as –14.8 kcal/mol, and originates mainly from its interactions with the protein. The contributions of WTR6, 7 and 133 to the solvation free energy is much smaller, –5.4, -6.4, -1.9 kcal/mol, respectively. The small magnitude of ΔG_{solv} of WTR133 shows that it does not contribute to the binding between Cyp A and CsA **1** as favorably as the other three bound water molecules in complex **1**. The contributions of WTR5, 6 and 7 to the solvation free energy in complex **2** are -11.1, -11.1 and –10.6

kcal/mol, respectively, and also originates mainly from their interactions with the protein. The difference of the contributions of the water clusters in complex **1** and **2** is:

$$\Delta\Delta G_{solv}^{2\text{-}1} = [\Delta G_{solv}(WTR5) + \Delta G_{solv}(WTR6) + \Delta G_{solv}(WTR7)]_{2} - [\Delta G_{solv}(WTR5) + \Delta G_{solv}(WTR6) + \Delta G_{solv}(WTR7) + \Delta G_{solv}(WTR133)]_{1} = -4.3 \text{ kcal/mol} \quad (36)$$

i.e., there is an *improvement* in the free energy of solvation going from **1** to **2** despite the reduction in the number of bound water molecules from 4 to 3.

In conclusion, this work reveals that the contributions of interfacial water molecules to the thermodynamic properties vary in a wide range and their different contributions to the solvation free energy mainly originate from the water-protein, water-ligand and water-water interaction energies. Furthermore, this work also shows that the rearrangement of the water molecules in the Cyp A-ligand complexes contributes favorably to the binding affinity, even though one of them is displaced going from **1** to **2**.

3.5 Theoretical studies of the thermodynamic consequences induced by the displacement and reorganization of interfacial water in Cyp A-ligands complexes

This section is also based on the publication (Li, Z. and Lazaridis, T. *J. Phys. Chem. B* **2006**, *110*, 1464-1475)

In addition to solvation free energy, the different affinity of **2** and **1** is also due to the change in protein-ligand interactions, the different desolvation energy and entropy, and the different conformational restriction entropy of the two ligands.

Based on the energy-minimized crystal structures, the direct interactions of the ethyl group of **1** and the hydroxypropyl group of **2** with the protein were –4.1 and –9.1 kcal/mol, respectively, giving $\Delta E_{Abu\_side\_chain\text{-}protein}^{2\text{-}1}$ = -5.0 kcal/mol. However, the total difference of the direct interactions of ligand **2** and **1** with the protein, is $\Delta E_{ligand\text{-}protein}^{2\text{-}1}$ = -1.4 kcal/mol. The extra CH_2OH of ligand **2** probably pushes it about 0.2 Å farther away from the protein than ligand **1**, and this reduces the interaction of the rest of the ligand with the protein.

Another contribution to the difference in binding affinity is the different desolvation enthalpy and entropy of the two ligands. The polar (oxygen and nitrogen) and apolar (carbon) solvent accessible surface areas in the bound and free states are shown in table 11. With the dehydration enthalpy and entropy parameters given by Makhatadze and Privalov, the desolvation entropy and enthalpy of the two ligands at 300 K were obtained as $\Delta\Delta H_{desolv}^{2\text{-}1}$= +8.6 kcal/mol and $\Delta\Delta S_{desolv}^{2\text{-}1}$ = +10.1cal/mol K, giving $\Delta\Delta G_{desolv}^{2\text{-}1}$= +5.7 kcal/mol.

The entropy of conformational restriction of the hydroxypropyl side chain of ligand **2** and the ethyl side chain of ligand **1** also needs to be considered. Empirical scales give an average value of conformational entropy of about 1.5 cal/mol K per rotatable bond. Similar to the method employed in the study of Con A-ligand complexes, the conformational entropy change of the ligands between the free and the complex state was estimated by comparing the probability distributions (p) of the dihedral angles in these two states. There are four dihedral angles in the hydroxypropyl group of ligand **2** and two in the ethyl group in ligand **1**. The probabilities (p) (from the Boltzmann expression) of the dihedral angles were obtained with rigid rotations and the integrals of $p\ln(p)$ were

used to estimate the entropy change, which is $\Delta S_{config}^{\mathbf{2\text{-}1}}$ = -2.7 cal/mol K, or $-T\Delta S_{config}^{\mathbf{2\text{-}1}}$ = +0.8 kcal/mol at 300 K.

The contributions to the difference in binding affinity between the two ligands are summarized in table 12. The difference in binding affinity of the two ligands with the protein CypA is +0.8 ±0.5 kcal/mol (equation 29) (see table 12). This value is consistent with experiment (+1.3 kcal/mol). No experimental data for the other thermodynamic properties have been reported so far.

Crystallography revealed two different conformations of the ligand in complex **2**. The above results pertain to the major conformer (occupancy 0.62). We repeated the calculations on the structure of complex **2** with ligand **2** in the minor conformation and obtained $\Delta\Delta G_{total}^{\mathbf{2'\text{-}1}}$ = +0.7±0.5 kcal/mol, quite similar to the result for the major conformer. However, there are significant differences in the contributions, such as those of certain water molecules. WTR5 and 7 in the minor conformer have larger contributions to the solvation free energy than those in the major conformer, while WTR6 shows the opposite result, giving $\Delta\Delta G_{solv}^{\mathbf{2'\text{-}1}}$ = -6.4 kcal/mol.

Part IV. Discussion

The thermodynamics of either isolated water molecules at the binding interface of HIV-1 protease-inhibitor complex and concanavalin A-trimannoside complex or water clusters at the binding interface of CypA-**1** and CypA-**2** complexes were explored in our work. Comparison of all results obtained (see table 13) leads to the following observations.

4.1 Enthalpic contributions of bound water molecules.

The isolated water molecules in the complexes of HIV-1 protease-inhibitor and concanavalin A-trimannoside have the largest magnitudes of E_{sw} and $\Delta E_{solv,}$. The interaction energy of a bound water molecule with a protein-ligand complex in CypA-**1** and CypA-**2** complexes varies in a wide range. The solute-water interaction energies of WTR5 in CypA-**1** and WTR5 and WTR6 in CypA-**2** are very negative and even comparable to the isolated water molecules in the complexes of HIV-1 protease-inhibitor and concanavalin A-trimannoside, while significant difference was observed for WTR133 in CypA-**1**. Similarly. The water-water interaction energy also varies in a wide range. The isolated water molecules in the complexes of HIV-1 protease-inhibitor and concanavalin A-trimannoside have the smallest magnitude of water-water interaction energy. WTR6 and WTR133 in CypA-**1** and WTR6 in CypA-**2** have very negative E_{ww} and even comparable to the water-water interaction energies in bulk water. E_{ww} is less negative for WTR5 and WTR7 in CypA-**1**, and even positive for WTR5 and WTR7 in CypA-**2**., which approximately corresponds to the number of hydrogen bonds formed between the bound water molecules and the protein-ligand complex. We noticed that the bound water molecules with more negative value of solute-solvent energy usually interact

weakly with other water molecules. This is because when a bound water molecule is strongly constrained by the protein and ligand by several hydrogen bonds, it will be difficult for other solvent molecules to simultaneously approach and interact strongly with it. Nevertheless, E_{sw} and E_{ww} do not fully compensate each other.

4.2 Entropic contributions of bound water molecules.

Similar to the enthalpic contribution, the isolated water molecules in the complexes of HIV-1 protease-inhibitor and concanavalin A-trimannoside, their solvent-solvent entropic contributions are also negligible. And the solvent-solvent contributions to the solvation entropy in CypA-**1** and CypA-**2** complexes vary in a wide range (from about 0 to-5.5 cal/mol K). Among all of the bound water molecules in these complexes, WTR5 in CypA-**1** has the most negative solvent-solvent entropies, as it is highly correlated with other water molecules (i.e., WTR6 and WTR7, see table 10) at the binding interface. WTR6, 7 and 133 in CypA-**1** are close, while WTR5 or 7 in CypA-**2** have a similar solvent-solvent entropy as the isolated water molecules in HIV-1 protease-inhibitor and Con A-trimannoside complexes, showing that they are fully buried at the binding interfaces and kept far from other solvent molecules. As with the energy, the largest value for the solute-solvent entropy (S_{sw}) are observed for the isolated water molecules in HIV-1 protease-inhibitor and Con A-trimannoside complexes. From table 13, we notice that a stronger interaction between the protein and the bound water usually corresponds to a more negative value of the solute-solvent entropy. However, this is not always the case. For instance, in complex **1**, WTR7 interacts more strongly with the protein than WTR6, but the solute-solvent entropies show the opposite result. The broader distributions of the

Euler angles, especially the broader distribution of g(θ) of WTR7 than that of WTR6 in complex **1** (see figure 5) implies that WTR7 in complex **1** adopts multiple favorable orientations and appears less ordered than WTR6 at the binding interface although it forms one more h-bond with the protein than WTR6.

4.3 Contributions of bound water molecules to solvation free energy

The values of solvation free energy of the water molecules in CypA-**1** and CypA-**2** reveal that almost all of the water molecules in the water clusters contribute less to protein-ligand binding than those isolated bound water molecules in HIV-1 protease-inhibitor and concanavalin A-trimannoside complexes and that WTR133 has the smallest favorable contribution, which mainly originates from the different water-protein and ligand interactions (see figure 13).

The additional hydroxyl group in complex CypA-**2** forms two more hydrogen bonds with the remaining water molecules in the water cluster (WTR6 and 7) and stabilizes them, which leads to a more favorable contribution to the solvation free energy (about 7 kcal/mol) even though the longer side chain replaces one less tightly bound water molecule in the water cluster. The additional interactions of **2** with the protein are also favorable for binding. These favorable energy contributions, however, are counterbalanced by the more positive desolvation enthalpy of ligand **2** relative to ligand **1**, leading to a less negative binding energy for ligand **2**. The overall entropy change is also unfavorable, leading to an unfavorable free energy change (see table 12).

4.4 Connecting the thermodynamic contributions of interfacial water or water cluster to the protein-ligand binding thermodynamics

No connection to the binding thermodynamics could be made in the displacement of the bound water going from HIV 1 protease-KNI 272 to HIV 1 protease DMP450, because the inhibitor DMP450 that displaced the ordered water molecule was entirely different from KNI-272. This complicates the attribution of the gain of binding affinity.

The similar ligand structure and binding modes in Con A- trimannoside complexes make it possible to connect the thermodynamic contributions of interfacial water to the protein-ligand binding thermodynamics. From table 6, we noticed that the elimination of the contribution of the ordered water molecules going from trimannoside **1** to **2** is the main origin of the loss of enthalpy by the displacement. This large contribution is almost compensated by the stronger direct interaction of trimannoside **2**, leading to a small change in binding enthalpy. Another small contribution to the more positive enthalpy in binding of trimannoside **2** to Con A is the more positive desolvation enthalpy of trimannoside **2** relative to trimannoside **1**.[119] This value is much smaller than the other two large contributions: $\Delta E_{L\text{-}P}$ and ΔE_{solv} (ordered water), but its contribution to the total change of enthalpy is not negligible.

The entropy of releasing the ordered water molecule in Con A-trimannoside **1** to bulk solvation is favorable and it is only partly offset by the constraint of the hydroxyethyl moiety of trimannoside **2** in the complex. The positive difference of the desolvation entropy of the two ligands add to the overall favorable entropy change. But this favorable entropy change is outweighed by the unfavorable enthalpy change, leading to the unfavorable free energy change $\Delta\Delta G$ = +1.0 kcal/mol. The final outcome on binding

affinity is a sensitive result of compensation of large contributions. Therefore, it is difficult to predict a priori whether water displacement will be favorable or not; it depends on structural details.

The above analysis differs in some ways from that proposed by Clarke et al..[36] In that work, the loss of enthalpy from trimannoside **1** to trimannoside **2** was proposed to result from differences in hydrogen bonding between the interactions of the two ligands with Con A and was based on several assumptions, some of which do not agree with our findings. First, the desolvation enthalpy of protein and the two ligands in their work was assumed to be the same for the two complexes, which is not what we find (although our estimate of this quantity is also very approximate). Second, counting hydrogen bonds neglects long range contributions to the difference in ligand-protein interactions, which we find to be substantial, especially the interaction with the ions. Third, the interaction energy of the water molecule in complex **1** with the protein was assumed to be identical to the interaction energy of the hydroxyl group in the hydroxyethyl moiety of trimannoside **2**, but our calculation gave an energy difference of 2.4 kcal/mol, showing a stronger interaction of the ordered water molecule with the protein. Finally, their assumption that the fractional occupancy of the hydrogen bond of an additional, dynamic water molecule with the C2 hydroxyl was the same as that of hydrogen bonds in bulk water is also in conflict with our simulations. The dynamic water appeared very scarcely in our simulations (the fractional occupancy of the additional hydrogen bond was about 0.01, very different from 0.7 in bulk water). In addition, the interaction energy of the C2 hydroxyl in trimannoside **1** and **2** with the bulk water was unfavorable (+1.0 and +2.1 kcal/mol, respectively).

As with the Con A-trimannoside complexes, the similar ligand structure and binding modes in Cyp A-ligand complexes also make it possible to connect the thermodynamic contributions of interfacial water cluster to the protein-ligand binding thermodynamics.

The additional hydroxyl group in complex CypA-**2** forms two more hydrogen bonds with the remaining water molecules in the water cluster (WTR6 and 7) and stabilizes them, which leads to a more favorable contribution to the binding energy (about 7 kcal/mol) even though the longer side chain replaces one less tightly bound water molecule in the water cluster. The additional interactions of **2** with the protein are also favorable for binding. These favorable energy contributions, however, are counterbalanced by the more positive desolvation enthalpy of ligand **2** relative to ligand **1**, leading to a less negative binding energy for ligand **2**. The overall entropy change is also unfavorable, leading to an unfavorable free energy change (see table 12).

The above analysis differs in some ways from that proposed by Mikol et al..[37] In that work, the contributions of the bound water molecules WTR5 and 7 to the binding affinity were assumed the same in the two complexes, the contribution of WTR133 was neglected and the free energy cost by the displacement of WTR6 together with the desolvation cost of **2** were assumed less than 2 kcal/mol. The favorable interaction of CypA-**2** relative to Cyp-**1** was considered as $\Delta E_{ligand\text{-}protein}^{\mathbf{2\text{-}1}}$ = -0.5 kcal/mol, which was presumably counterbalanced by contributions due to changes in the translational, rotational and vibrational degrees of the protein and ligand. Our findings give a different picture. First, the free energy contributions of WTR5 in the two complexes are different (ca. 3.6 kcal/mol), which presumably originates from the disruption of the hydrogen bond between WTR5 and WTR6 going from **1** to **2**. A similar difference is also found for

WTR7 but in the opposite direction, due to the formation of a new hydrogen bond between WTR7 and ligand **2**. These two terms compensate each other and add up to $\Delta\Delta G_{solv(w5+w7)}^{2-1}$ = -0.6 kcal/mol. Second, WTR133 does make a small contribution to the free energy in Cyp-1 ($\Delta G_{solv(w133)}$= -1.9 kcal/mol). Third, the free energy cost by the displacement of WTR6 was obtained as $\Delta\Delta G_{solv(w6)}^{2-1}$ = -5.8 kcal/mol, combined with the different desolvation free energy $\Delta\Delta G_{desolv}^{2-1}$ = +9.6 kcal/mol, giving +3.8 kcal/mol, which is even larger than the corresponding upper bound proposed by Mikol et al..[37] The calculated difference in conformational entropy loss of the Abu side chain of **1** and **2** ($-T\Delta S_{config}^{2-1}$ = +0.8 kcal/mol) is similar to the value they assumed. However, the calculated additional interaction energy of ligand **2** with the protein($\Delta E_{ligand\text{-}protein}^{2-1}$ = -1.4 kcal/mol) is much larger than they assumed.

In our work on Con A-trimannoside and Cyp A-ligand complexes, the good agreement between the calculated and experimental change in affinity is at least somewhat fortuitous, given the numerous approximations made in the calculations. We neglected the conformational entropy difference arising from the same parts of the two ligands. The desolvation enthalpy and entropy of protein in the two complexes were assumed to be the same, and the desolvation terms of the ligands were estimated according to their solvent accessible surface area, which is an empirical model. To calculate the solvent-solvent entropy of interfacial waters in Cyp A-ligand complexes, we assumed that the inhomogeneous pair correlation function is equal to that in bulk solvent. In addition, this work focused exclusively on the 3 or 4 bound water molecules at the binding interface and assumed that the contribution of all other water molecules is the same in the two complexes. Applying this theoretical analysis to all water molecules in

the system is currently not feasible. All these approximations, together with the standard uncertainties related to force field parameters and lack of polarizability etc., make the true uncertainties much larger than the error bars listed in Tables 2, 3, 5, 6, 7, 10 and 12. However, the qualitative picture that arises from the calculations should be reliable.

Our work shows that IFST can yield the thermodynamic contributions of ordered water molecules with rigorous decomposition of the solvation free energy into contributions from different solvent regions[120] One advantage of this method over standard free energy simulations is that it allows us to focus on the vicinity of the solute, especially on the biomolecular interface region, where the largest contributions to solvation thermodynamics originate. Another advantage is that it is not only applicable to highly ordered, isolated water molecules but also to clusters of water molecules, or surface water molecules. In contrast, the removal of water molecules from surface sites upon FE simulations is not meaningful, because another water molecule will immediately take their place. Other advantages include the facility by which it yields all thermodynamic functions, easier assessment of sampling adequacy, and a lower computational cost. The disadvantages of ISFT are that it involves numerous approximations, and requires the evaluation of complicated integrals, which makes it less "user friendly".[121]

This approach could be useful in rational drug design by estimating which bound water molecules would be most favorable to displace. In addition, the insights obtained from this analysis could help improve empirical scoring functions for prediction of protein-ligand binding affinity by taking the thermodynamic contributions of bound water molecules into account. From the results obtained so far, gain of binding free energy from

the displacement of water might be achieved under the following conditions: (i) the new ligand must interact strongly with the protein and the remaining water molecules; (ii) the additional group ideally should not disrupt interactions between bound water molecules; (iii) the conformational entropy loss should be minimized if possible, for example by use of a carbonyl group instead of a hydroxyl group. This is consistent with the rules introduced by Garcia-Sosa et al..[122]

From the published experimental and theoretical studies on interfacial water the following conclusions can be reached: a) interfacial water involved in hydrogen bonds with the protein and ligand makes a negative contribution to the entropy, enthalpy, and heat capacity of binding. However, water that does not form hydrogen bonds can have higher entropy than in the bulk b) Water-mediated interactions can be as strong as direct interactions. Displacement of bound water by ligand modification can increase or decrease binding affinity depending on the detailed balance of the various contributions involved.

A number of methods have been applied to obtain the thermodynamics of interfacial water: FEP simulations,[95] the Gaussian model,[87] and IFST.[123; 124; 125] These methods all have advantages and disadvantages. Furthermore, they are very different from each other. It would be useful in the future to explore the relationship between them by applying them all to a wider range of problems and comparing closely the results.

The biggest challenge is to find practical ways to include water-mediated interactions in protein-ligand docking, scoring, and drug design. One direction is to use discrete, explicit water molecules at key positions in the binding interface. This will require improvements in the methodology of predicting water locations and has the disadvantage

of neglecting entropic effects. The second direction is towards implicit modeling of water. A start has been made with the knowledge-based water-mediated potentials of Wolynes and coworkers,[82; 83] but it would be useful to develop physics-based potentials as well. What is essentially needed is the extension of implicit solvent models, such as ACE,[126] Generalized Born,[127] or EEF1,[128] to make them applicable to water in restricted environments. It is a difficult task, because the thermodynamics of bound water is sensitive to its surroundings. To capture this complexity correctly one would most likely need many-body potentials and an extensive set of data on the dependence of the free energy of the bound water on the configuration of surrounding polar atoms.

Table 1. Thermodynamic contributions of the ordered water molecule in HIV-1 protease

Temperature	300 K	330 K
S_{sw} (cal/mol K)	-25.0	-24.5
E_{sw} (kcal/mol)	-28.2	-28.0
ΔE^{o}_{ww} (kcal/mol)	+10.1	+9.6
ΔS^{o}_{ww} (cal/mol K)	+15.2	+13.5
ΔE_{solv} (kcal/mol)	-18.1	-18.4
ΔS_{solv} (cal/mol K)	-9.8	-11.0
ΔG_{solv} (kcal/mol)	-15.2	-14.8

Table 2. Contributions to solvation from the ordered water molecule in Con A-trimannoside **1** calculated from the MD simulations at different temperature.

Temp.	300K	330K
E_{sw}	-30.2 ±0.1	-30.0±0.2
$E_{w\text{-}ions}$	-9.2 ±0.02	-9.2 ±0.02
E_{ww}	+0.9 ±0.1	+0.9 ±0.2
S_{sw}^{or}	-11.6 ±0.01	-11.4 ±0.01
S_{sw}^{tr}	-10.4 ±0.1	-9.9 ±0.1
$S_{sw} = S_{sw}^{or} + S_{sw}^{tr}$	-22.1 ±0.1	-21.3 ±0.1
$\Delta E_{solv} = E_{sw} + E_{ww} + \Delta E_{ww}^{0}$	-19.2 ±0.1	-19.5 ±0.2
$\Delta S_{solv} = S_{sw} + \Delta S_{ww}^{0}$	-6.8 ±0.1	-7.8 ±0.1
$\Delta G_{solv} = \Delta E_{ww}^{0} - T\,\Delta S_{ww}^{0}$	-17.2 ±0.1	-16.9 ±0.1
ΔCp	-10± 3	

* Note: Units for enthalpy and free energy are kcal/mol

Units for entropy and heat capacity are cal/mol K

All of the error bars listed in table 1 and other tables were calculated by averaging the values of each thermodynamic parameter over 2 ns portions of the MD trajectories, and then taking the standard deviation of the 4 samples.

Table 3. Interaction energies (kcal/mol) of the C2-substituents of trimannoside **1** and trimannoside **2** with Con A and Ca^{2+} and Mn^{2+} obtained directly from the MD simulations at 300 K.

	Con A – trimannoside **1**	Con A – trimannoside **2**	Difference between **1** and **2**
E_{L2OH-P}	-9.4±0.1	-27.8 ±0.02	-18.4 ± 0.1
$E_{L2OH-Ions}$	-10.0 ±0.08	-12.3±0.03	-2.3 ± 0.1

Table 4. The apolar and polar accessible surface areas (Å^2) of mannose 2 in trimannoside **1** and **2** at complex (final structure after 8 ns MDs) and uncomplexed states

	Complexed		Uncomplexed		Difference (ΔASA)	
	Polar	Apolar	Polar	Apolar	Polar	Apolar
Trimannoside **1**	7.5	136.8	65.6	168.2	58.1	31.4
Trimannoside **2**	10.5	125.2	77.3	179.8	66.8	53.6

Table 5. Contributions to the binding enthalpy, entropy and heat capacity in Con A-ligands complexes.

	Con A – Trimannoside **1**		Con A – Trimannoside **2**		Difference between **1** and **2**	
Temp.	300K	330K	300K	330K	300K	330K
E_{L2OH-P}	-9.4 ±0.1	-9.3 ±0.03	-27.8 ±0.02	-27.6±0.05	-18.4 ±0.1	-18.3±0.05
ΔE_{solv} (ordered water)	-19.2 ±0.1	-19.5 ±0.1	0	0	+19.2 ±0.1	+19.5 ±0.1
ΔS_{solv} (ordered water)	-6.8 ±0.1	-7.8 ±0.1	0	0	+6.8 ±0.1	+7.8 ±0.1
ΔH_{desolv} [a]	+15.4	+15.7	+18.3	+19.3	+2.9	+3.6
ΔS_{desolv} [a]	+18.0	+18.9	+23.1	+26.4	+5.1	+7.5
ΔS_{config}	0	0	-2.9	-2.9	-2.9	-2.9
ΔE (total) [b]	-13.2±0.1	-13.1±0.1	-9.5± 0.02	-8.3±0.05	+3.7± 0.1	+4.8±0.1
ΔS (total)	+11.2±0.1	+11.1±0.1	+20.2	+23.5	+9.0±0.1	+12.3±0.1
ΔG (total)	-16.6±0.1	-16.8±0.1	-15.6±0.02	-16.1±0.05	+1.0±0.1	+0.7±0.1
ΔC_p [c]	3 ± 3		40± 1		+37±3	

* Note: Units for enthalpy and free energy are kcal/mol

Units for entropy and heat capacity are cal/mol K

a. ΔH_{desolv} stands for the contribution of enthalpies of dehydration from the hydroxyl group in trimannoside **1** and from the hydroxyethyl group in trimannoside **2;** Similarly to ΔS_{desolv}

b. ΔE (total) = E_{L2OH-P}+ ΔE_{solv} (ordered water) +$\Delta H_{desolv,}$ and ΔG (total) = ΔE (total) -TΔS (total)

c. ΔC_p was calculated with $\Delta\Delta E$ (total)/ΔT

Table 6. Experimental and calculated differences (trimannoside **2** - trimannoside **1**) in binding between the two ligands.

	Difference between trimannoside **1** and trimannoside **2** from calculation	Difference between trimannoside **1** and **2** from experimental data[36]
Temperature	300 K	298.15 K
ΔΔE (kcal/mol)	+3.7± 0.1	+2.3±0.9
ΔΔS (cal/mol K)	+9.0±0.1	+3.7±3.0
ΔΔG (kcal/mol)	+1.0±0.1	+1.2±0.1
ΔΔCp (a) (cal/mol K)	+37 ± 4	+17.0±14

Note: Units for enthalpy and free energy are kcal/mol; Units for entropy and heat capacity are cal/mol K

Table 7. Contributions to solvation from the ordered water molecules in Cyclophilin A (CypA)-Cyclosporin A (CsA, **1**) calculated from MD simulations at 300 K

Bound water	WTR5	WTR6	WTR7	WTR133
Number of H-bonds with the protein[1]	3	1	2	0
Occupancy[2]	7999/8000	7747/8000	7983/8000	7481/8000
E_{sw}	-25.6±0.1	-8.3±0.3	-14.1±0.1	-3.7±0.1
S_{sw}^{or}	-8.4±0.2	-6.9±0.4	-5.1±0.01	-5.2±0.2
S_{sw}^{tr}	-9.9±0.2	-6.5±0.3	-7.1±0.1	-4.9±0.4
S_{sw}	-18.3±0.3	-13.4±0.5	-12.2±0.1	-10.1±0.5
E_{ww}	-3.8±0.1	-15.4±0.6	-5.1±0.1	-15.6±0.6
S_{ww}^{or}	-5.1±0.1	-2.9±0.3	-3.5±0.2	-3.2±0.4
S_{ww}^{trans}	-0.4±0.01	-0.5±0.03	-0.3±0.01	-0.3±0.1
S_{ww}	-5.5±0.1	-3.4±0.3	-3.8±0.2	-5.6±0.4
ΔE_{solv}	-17.4±0.1	-5.9±0.4	-6.6±0.05	-1.4±0.3
ΔS_{solv}	-8.7±0.3	-1.6±0.5	-0.7±0.2	+1.5±0.5
ΔG_{solv}	-14.8±0.1	-5.4±0.4	-6.4±0.1	-1.9±0.3

* Note: 1. No H-bond is found between the water molecules and the ligand.

2. Probability of finding a water molecule in the specific region, within r =1.2 Å from its average position

3. Units for enthalpy and free energy are kcal/mol; Units for entropy are cal/mol K.

Table 8. Contributions to solvation from the bound water molecules in Cyclophilin A -(5-hydroxynorvalin)-2-cyclosporin (**2**) calculated from the MD simulations at 300 K

Bound water	Wat5	Wat6	Wat7
Number of H-bonds with the protein and ligand[1]	3	2	3
Occupancy	8000/8000	8000/8000	8000/8000
E_{sw}	-23.0±0.04	-17.7±0.01	-24.3±0.06
E_{ww}	+0.4±0.01	-11.0±0.08	+3.4±0.1
S_{sw}^{or}	-9.6±0.05	-9.9±0.01	-9.1±0.02
S_{sw}^{tr}	-11.3±0.2	-10.6±0.1	-12.2±0.05
S_{sw}	-20.9±0.2	-20.5±0.1	-21.3±0.1
S_{ww}^{or}	+0.2±0.001	-0.6±0.5	-0.2±0.02
S_{ww}^{trans}	-0.01±0.01	-0.6±0.02	-0.04±0.01
S_{ww}	+0.2±0.01	-1.2±0.5	-0.2±0.02
ΔE_{solv}	-12.7±0.04	-13.1±0.08	-12.5±0.1
ΔS_{solv}	-5.5±0.2	-6.5±0.5	-6.3±0.1
ΔG_{solv}	-11.1±0.1	-11.1±0.2	-10.6±0.1

Note: 1. For either WTR6 or 7, one of the H-bonds is formed between the water and the ligand.

2. Units for enthalpy and free energy are kcal/mol; Units for entropy are cal/mol K.

Table 9. The regions of neighboring water molecules around each bound water molecule in Cyclophilin A (CypA)-Cyclosporin A (**1**), and Cyclophilin A -(5-hydroxynorvalin)-2-cyclosporin (**2**) complexes

Complex	Bound water	Neighbor	R(Å)	θ (radian)	ϕ (radian)	Corresponding water[1]
Complex **1**	WTR5	1	2.0~4.0	1.1~1.85	0.0~1.6	WTR7
		2	2.0~5.0	1.85~2.35	0.0~1.3	WTR129*
		3	2.0~3.0	2.5~3.14	0.0~1.2	WTR6*
	WTR6	1	2.0~3.0	2.2~2.8	0.9~2.1	WTR133*
		2	2.0~3.0	0.0~0.6	3.1~4.75	WTR5
		3	2.0~5.0	1.2~1.75	0.0~0.9	WTR129*
		4	3.0~4.0	0.4~1.0	0.0~1.1	WTR7
		5	3.0~5.0	1.3~2.2	0.9~1.6	Solv54*
	WTR7	1	2.0~4.0	1.3~1.8	3.3~4.3	WTR5
		2	2.0~5.0	2.05~2.70	0~1.1	WTR129*
		3	3.0~4.0	2.2~2.70	3.3~4.3	WTR6*
		4	3.0~5.0	2.2~2.70	1.3~2.3	Solv54*
		5	4.0~5.0	2.65~3.0	2.4~3.3	WTR133*
	WTR133	1	2.0~3.0	0.3~0.95	4.0~5.3	WTR6*
		2	2.0~3.0	0.8~1.5	0.8~1.5	Solv54*
		3	2.0~5.0	2.05~2.6	1.6~2.5	WTR110*
		4	3.0~5.0	1.55~2.0	0.7~1.9	Solv9*
		5	3.0~5.0	1.0~1.6	0.0~0.7	Solv59*
		6	3.0~5.0	1.7~2.5	0.0~0.7	WTR91*
Complex **2**	WTR5	1	3.0~4.0	0.8~1.2	0.5~1.2	WTR7
	WTR6	1	2.0~3.0	1.3~1.6	5.7~6.28	WTR102*
		2	2.0~5.0	0.85~1.5	1.2~1.9	WTR67*
		3	3.0~4.0	0.3~0.9	0~1.4	WTR114*
		4	3.0~5.0	0.3~0.75	4.4~5.1	WTR7
		5	4.0~5.0	1.5~1.8	1.1~1.9	Solv83*
		6	3.0~5.0	1.15~1.45	0.6~0.9	Solv98*
	WTR7	1	3.0~4.0	2~2.3	3.6~4.3	WTR5
		2	4.0~5.0	2.15~2.5	0~0.8	WTR102
		3	3.0~5.0	1.6~1.9	0.7~1.5	WTR114*
		4	3.0~5.0	2.4~2.7	1.2~1.9	WTR6

** Note: "Corresponding water" means to the water molecule located in that region. "WTR" denotes those water molecules from the crystal structure, "Solv" denotes those water molecule in the solvent sphere we added around the binding sites. The "*" means the water molecule in that region exchanged at least once during the MDs.*

Table 10. The solvent-solvent entropy (cal/mol K) of each bound water molecule in Cyclophilin A (CypA)-Cyclosporin A (**1**), and Cyclophilin A -(5-hydroxynorvalin)-2-cyclosporin (**2**) complexes

Complex	Bound water [w(*i*)]	Neighbor	Corresponding water	Solvent-solvent entropy $S_{w(i)w(j)}$[a]
Complex **1**	WTR5	1	WTR7	-3.50
		2	WTR129*	-0.01
		3	WTR6*	-2.05
	WTR6	1	WTR133*	-1.70
		2	WTR5	-1.60
		3	WTR129*	-0.05
		4	WTR7	-0.10
		5	Solv54*	+0.02
	WTR7	1	WTR5	-3.63
		2	WTR129*	-0.03
		3	WTR6*	-0.04
		4	Solv54*	-0.02
		5	WTR133*	-0.03
	WTR133	1	WTR6*	-1.51
		2	Solv54*	-2.03
		3	WTR110*	-0.02
		4	Solv9*	-0.01
		5	Solv98*	-0.03
		6	WTR91*	-0.03
Complex **2**	WTR5	1	WTR7	+0.15
	WTR6	1	WTR102*	-0.37
		2	WTR67*	-0.17
		3	WTR114*	-0.41
		4	WTR7	-0.26
		5	Solv59*	+0.01
		6	Solv83*	+0.01
	WTR7	1	WTR5	+0.14
		2	WTR102	-0.12
		3	WTR114*	+0.02
		4	WTR6	-0.26

*Note: a. Good convergence of the results of $S_{w(i)w(j)}$ was obtained over two portions of MD trajectory and the values listed here were obtained over the whole MD trajectory. $S_{w(i)w(j)}$ and $S_{w(j)w(i)}$ give similar values.

b. Units for entropy and heat capacity are cal/mol K

Table 11. The apolar and polar accessible surface areas ($Å^2$) of CsA (**1**) and (5-hydroxynorvaline)-2-cyclosporin (**2**) at free and bound states

	Bound		Free		Difference (ΔASA)	
	Polar	Apolar	Polar	Apolar	Polar	Apolar
CsA (**1**)	95.1	597.0	178.9	1046.2	83.8	449.2
(5-hydroxynorvaline)-2-cyclosporin (**2**)	98.9	597.6	215.1	1063.5	116.2	465.9

Table 12. Enthalpic and entropic contributions to the difference in the binding affinity between the two ligands in Cyclophilin A (CypA)-Cyclosporin A (**1**), and Cyclophilin A -(5-hydroxynorvalin)-2-cyclosporin (**2**) complexes

	CypA-**1**	CypA- **2**	Difference between **1** and **2**
$E_{ligand\text{-}protein}$	-113.6	-115.0	-1.4
ΔE_{solv} (water cluster)	-31.3±0.5	-38.3±0.1	-7.0±0.5
$-T\Delta S_{solv}$ (water cluster)	+2.8±0.2	+5.5±0.1	+2.7±0.2
ΔH_{desolv} [a]	+34.4	+43.0	+8.6
$-T\Delta S_{desolv}$ [a]	-25.0	-27.9	-2.9
$-T\Delta S_{config}$	+0.9	+1.7	+0.8
ΔH (total) [b]	-110.5±0.5	-110.3±0.1	+0.2±0.5
$-T\Delta S$ (total)	-21.3±0.2	-20.7±0.1	+0.6±0.2
ΔG (total)	-131.8±0.5	-131.0±0.1	+0.8±0.5

* Note: Units for enthalpy, $T\Delta S$ and free energy are kcal/mol

ΔH_{desolv} and ΔS_{desolv} are the differences in dehydration enthalpy and entropy

ΔH (total) = $E_{ligand\text{-}protein}$+ ΔE_{solv} (water cluster) +$\Delta H_{desolv,}$ ΔS (total) = $\Delta S_{ligand\text{-}protein}$+ ΔS_{solv} (water cluster) +$\Delta S_{config,}$ and

ΔG (total) = ΔE (total) -$T\Delta S$ (total)

Table 13. Thermodynamic parameters of the bound water molecules in the HIV-1 protease-inhibitor complex, Con A-trimannoside complex, cyclophilin A-CsA **1**, complex and cyclophilin A-CsA **2** complex.

Complex	Bound Water	E_{ww}	E_{sw}	ΔE_{solv}	S_{ww}	S_{sw}	$-T\Delta S_{solv}$	ΔG_{solv}
HIV-1 protease	WTR1	0	-28.2	-18.1	0	-25.0	+2.9	-15.2
Con A-**1**	WTR1	0	-30.2	-19.2	0	-22.1	+2.0	-17.2
Cyclophilin A-**1**	WTR5	-3.8	-25.6	-17.4	-5.5	-18.3	+2.6	-14.8
	WTR6	-15.4	-8.3	-5.9	-3.4	-13.4	+0.5	-5.4
	WTR7	-5.1	-14.1	-6.6	-3.8	-12.2	+0.2	-6.4
	WTR133	-15.6	-3.7	-1.4	-3.5	-10.1	-0.5	-1.9
	Cluster[1]	-39.9	-51.7	-31.3	-18.2	-54.0	+2.8	-28.5
Cyclophilin A-**2**	WTR5	+0.4	-23.0	-12.7	+0.2	-20.9	+1.6	-11.1
	WTR6	-11.0	-17.7	-13.1	-1.2	-20.5	+2.0	-11.1
	WTR7	+3.4	-24.3	-12.5	-0.2	-21.3	+1.9	-10.6
	Cluster[2]	-7.2	-65.0	-38.3	-1.2	-62.7	+5.5	-32.8

Note: Units for the energy are kcal/mol;

Units for the entropy are cal/mol K.

Figure 1. Binding between HIV 1-protease-KNI 272 including a water molecule bridging the protein and the ligand.

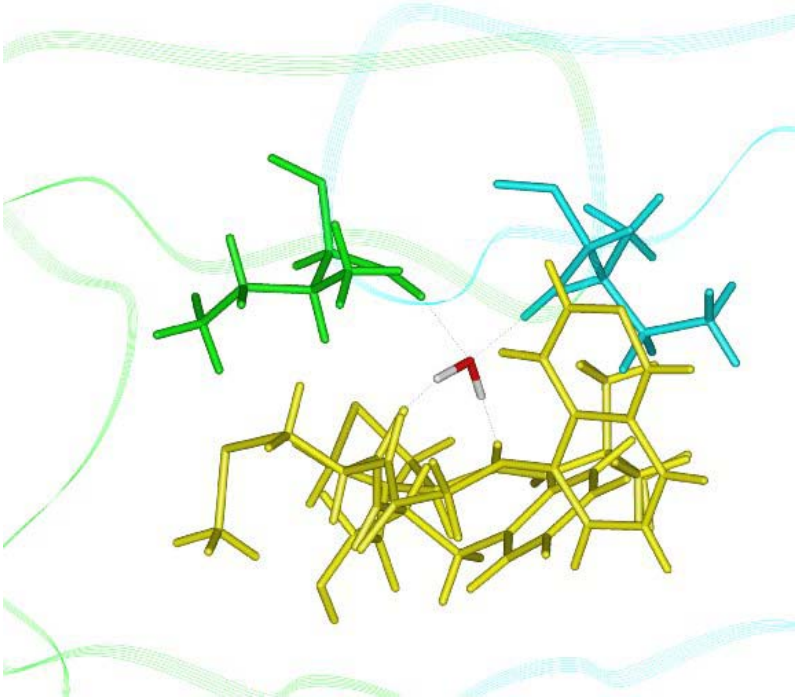


Figure 2. Radial distribution function at 300 and 330 K. r is the distance of the key water molecule in HIV-1 protease-KNI 272 complex from its average position.

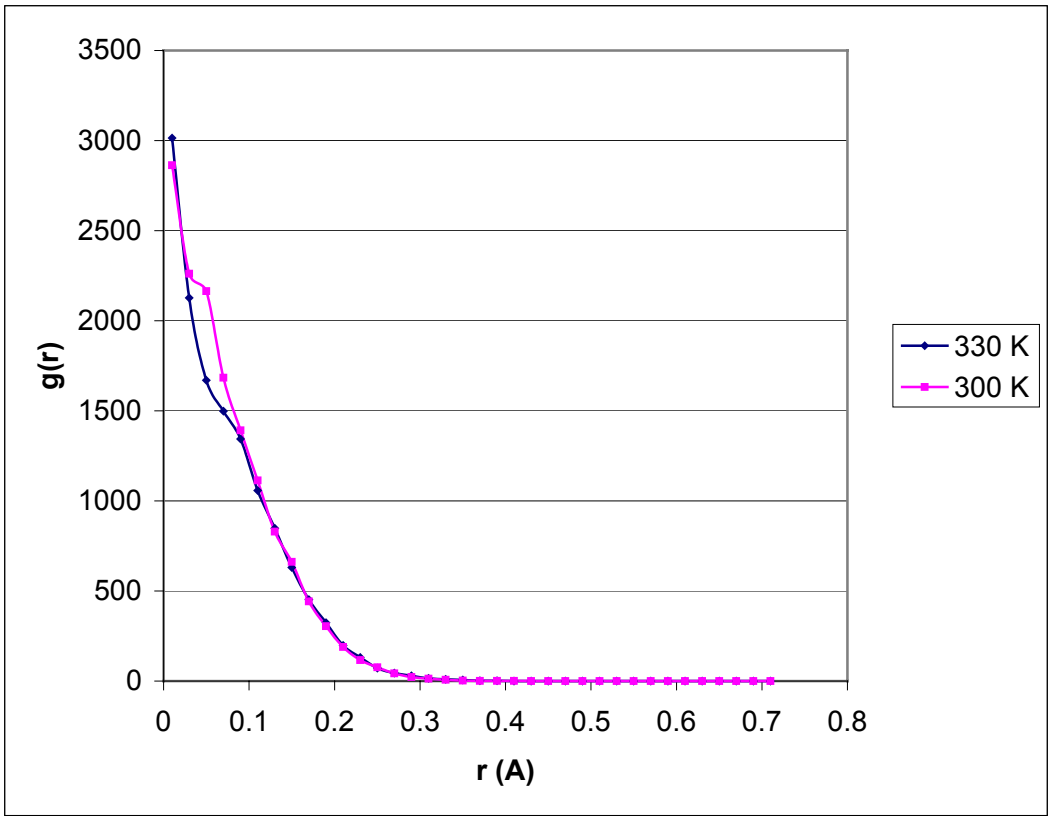


Figure 3. Probability distribution of the Euler angles around their average values for the water in HIV-1 protease-KNI 272 complex.

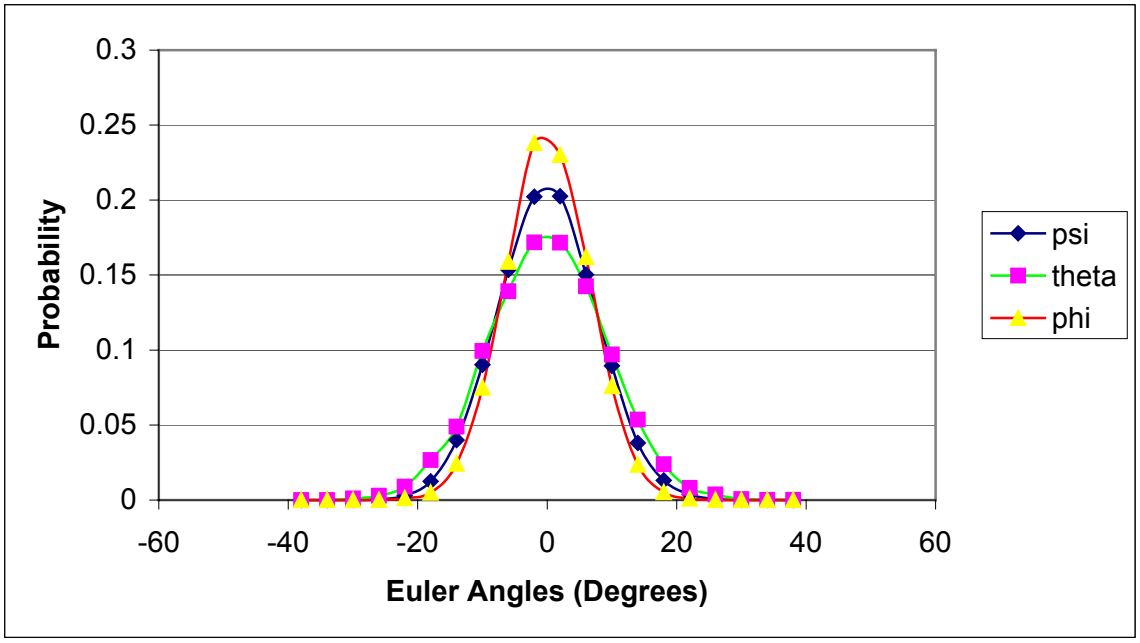


Figure 4. Structure of trimannoside **1** (left) and **2 (**right**)**;

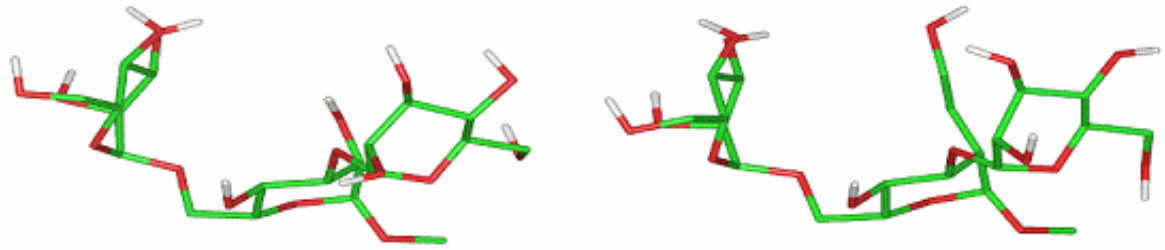
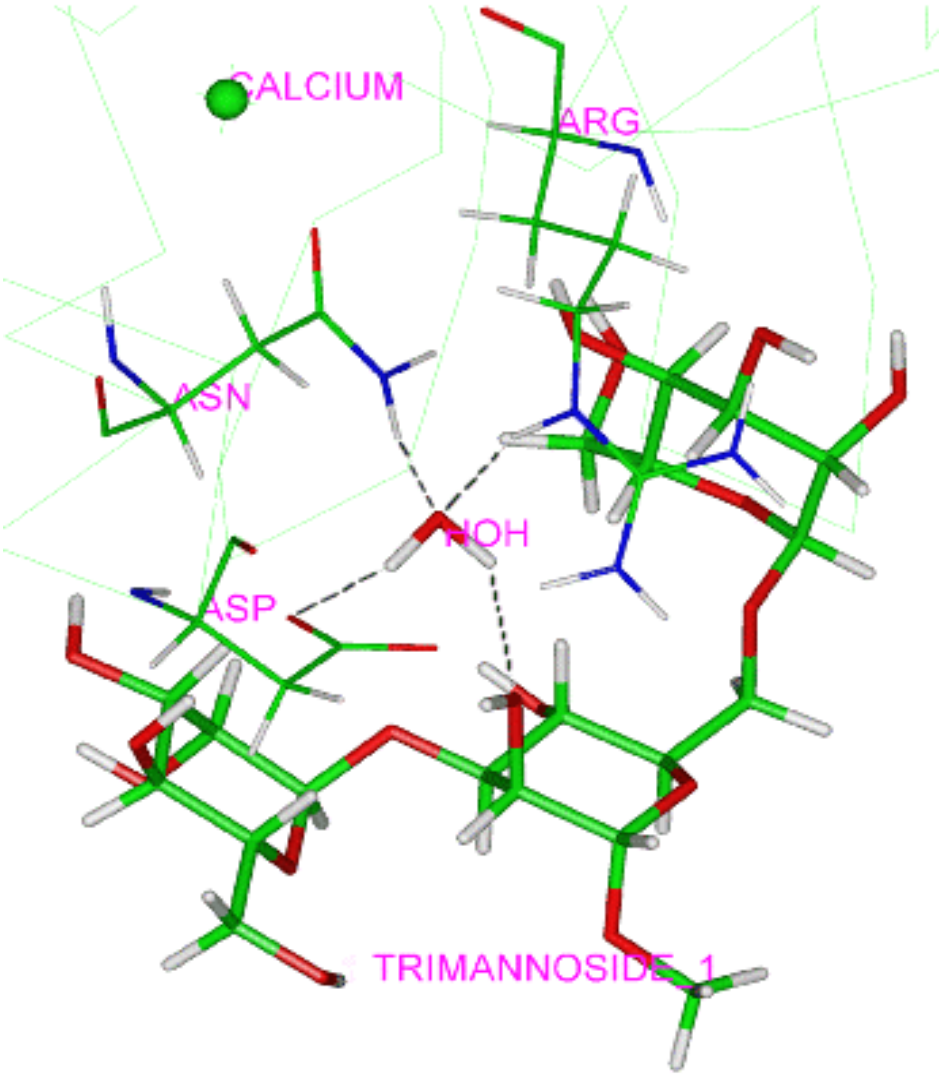


Figure 5. Binding of trimannoside **1** and **2** to Con A (structures obtained after 8ns MD simulation). a. The ordered water molecule bridging trimannoside **1** (in thicker sticks) and the 3 residues (Asp14, Asn16, and Arg228 in thinner sticks) at the binding site of Con A with 4 hydrogen bonds (dashed black line); Oxygen atoms are in red, nitrogen atoms in blue, carbon atoms in green and hydrogen atoms in gray; Calcium is shown as a sphere.



b. The similar hydrogen bonds (three hydrogen bonds in dashed black line) of the C2 hydroxyl in trimannoside **2** with the protein. Oxygen atoms are in red, nitrogen atoms in blue, carbon atoms in green and hydrogen atoms in gray; Calcium is shown as a sphere.

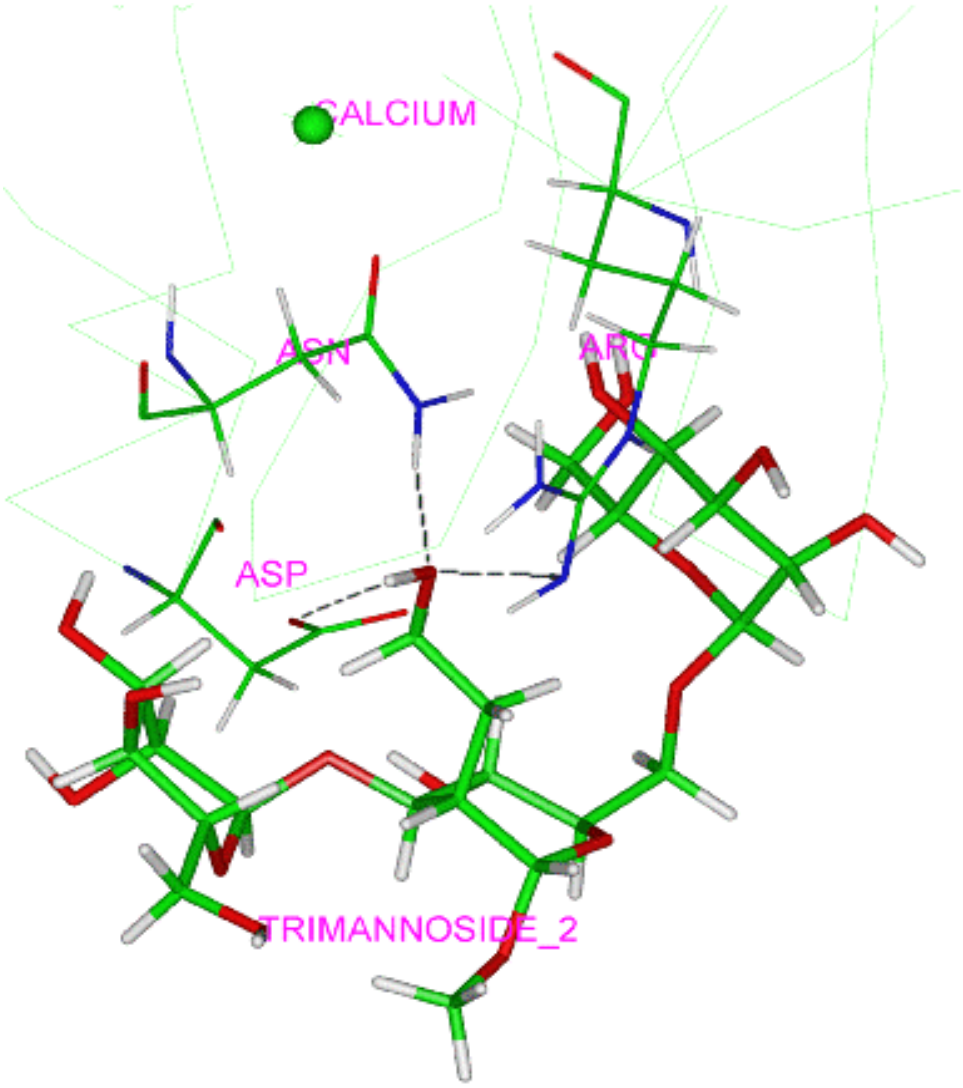


Figure 6. Translational distribution function of the bound water in Con A-trimannoside 1 complex $g_{sw}^{r}(r)$, $g_{sw}^{tr}(\theta')$ and $g_{sw}^{tr}(\phi')$ during different portions (1st & 2nd half) of the simulations. r is the distance of the bound water molecule from its average position.

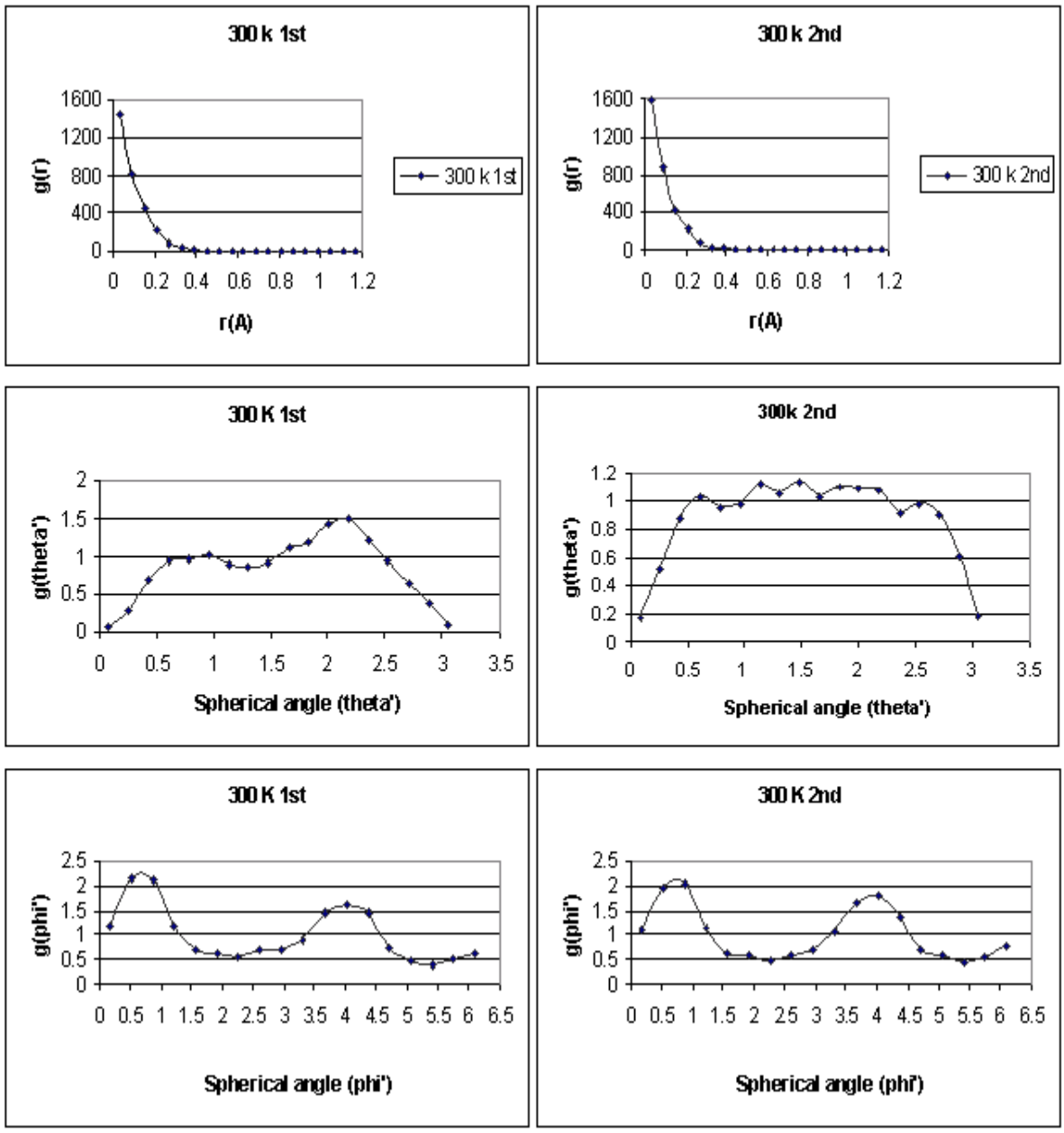


Figure 7. Probability distribution of the Euler angles of the bound water in Con A-trimannoside 1 complex around their average values at 300 K, during different portions (1^{st} & 2^{nd} half) of the simulations.

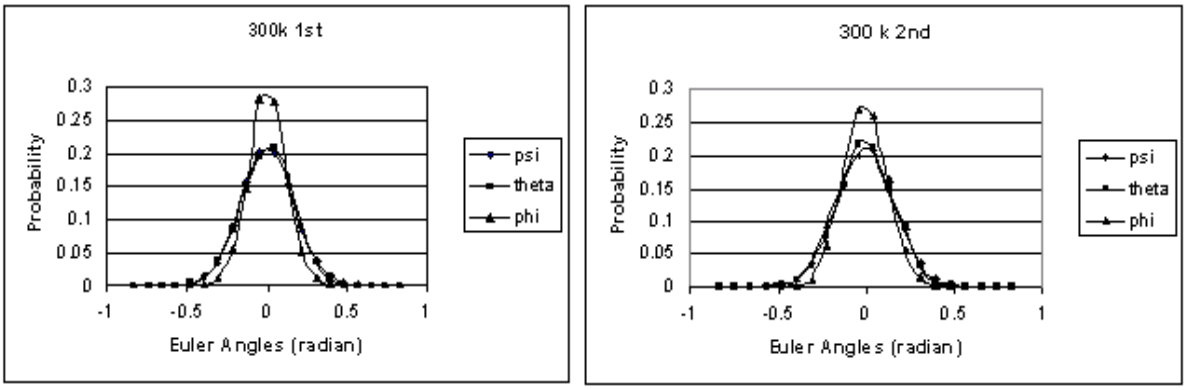


Figure 8. Structure of cyclic undecapeptides cyclosporin A (CsA) and (5-hydroxynorvaline)-2-cyclosporin. The only difference between these two molecules is highlighted by the purple rectangle. CsA includes: MeBmt1 (*N*-methyl-(4R)-4-[(E)-2-butenyl]-4-methyl-l-threonine), Abu2 (l-α-amino-butyric acid), Sar3 (sarcosine), MeLeu4 (*N*-methyl-leucine), Val5, MeLeu6, Ala7, d-Ala8, MeLeu9, MeLeu10, and MeVal11 (*N*-methylvaline).

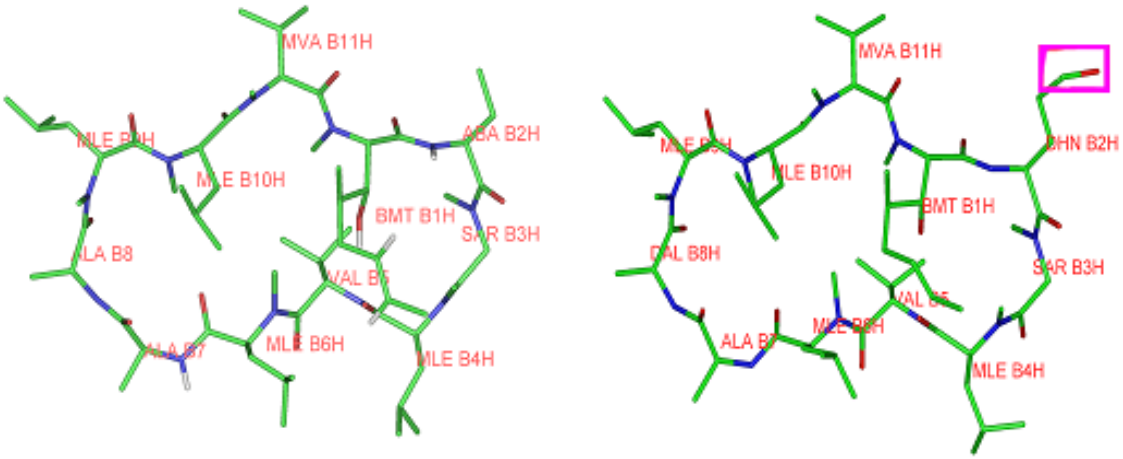
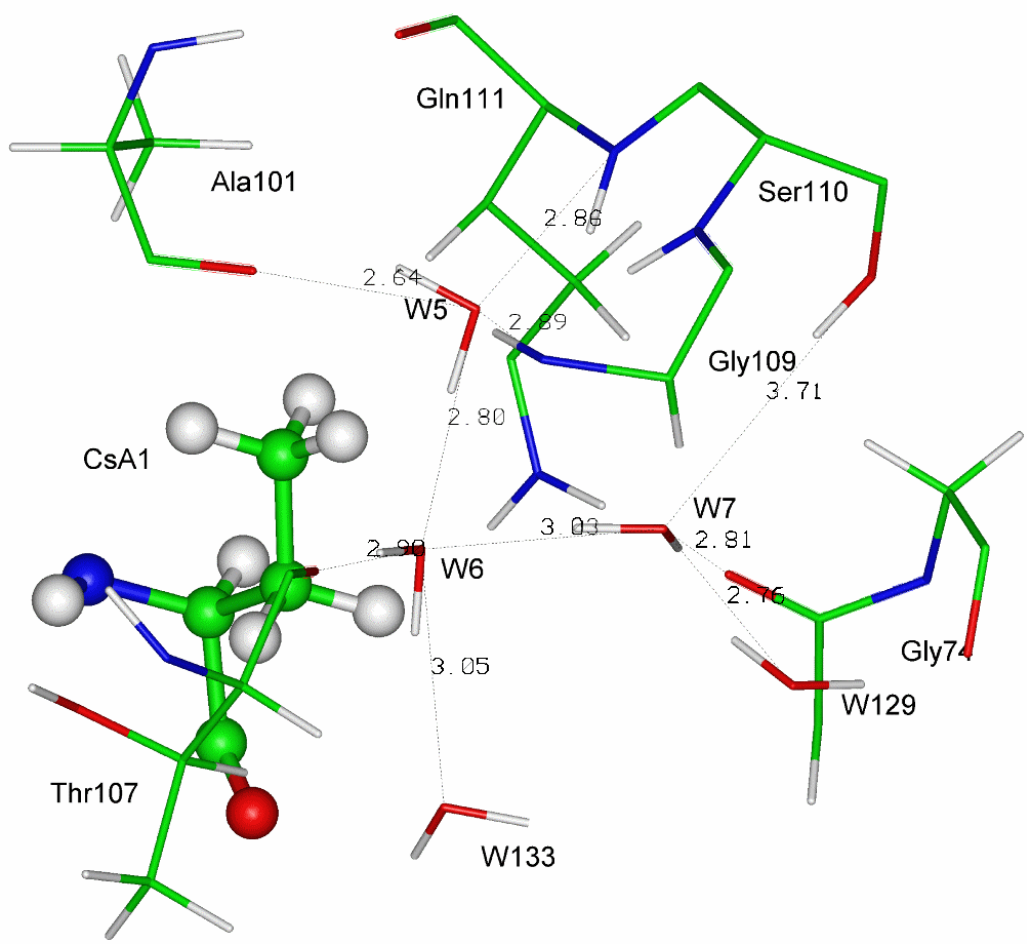


Figure 9. a. Bound water molecules in the Abu pocket in the CypA-**1** complex. CsA **1** is drawn in ball and sticks (only the residue Abu2 is depicted). CypA is drawn in thinner sticks. All potential h-bonds are shown in dashed black lines. Oxygen atoms are in red, nitrogen atoms in blue, carbon atoms in green and hydrogen atoms in gray.



b. Bound water molecules in the Abu pocket in the CypA-**2** complex. CsA **2** is drawn in ball and sticks (only the residue Abu2 is depicted). CypA is drawn in thinner sticks. All potential h-bonds are shown in dashed black lines. Oxygen atoms are in red, nitrogen atoms in blue, carbon atoms in green and hydrogen atoms in gray.

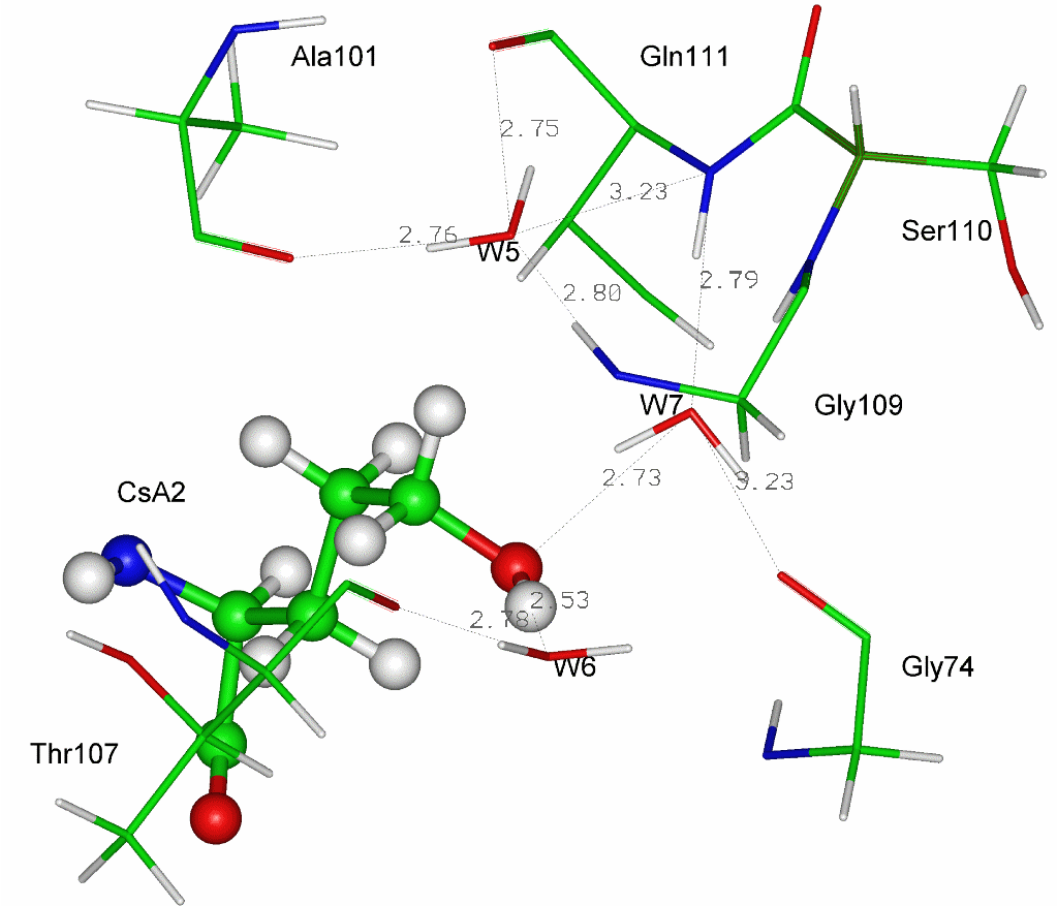
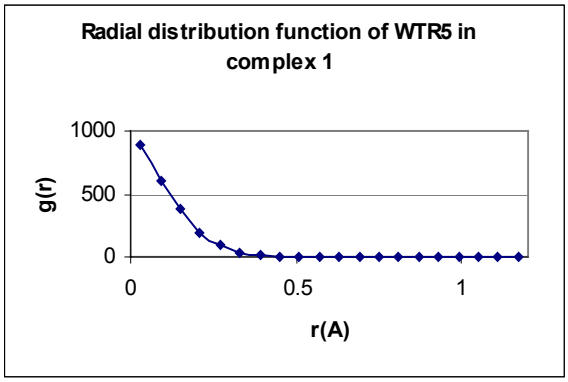
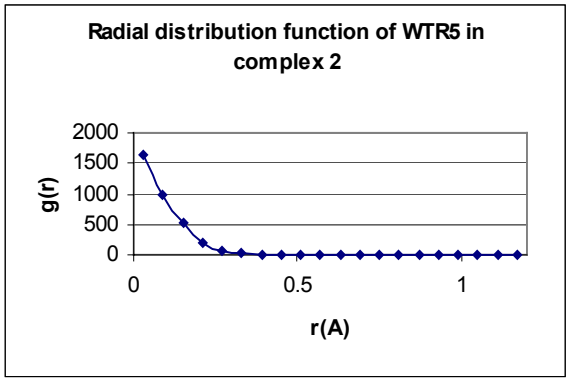
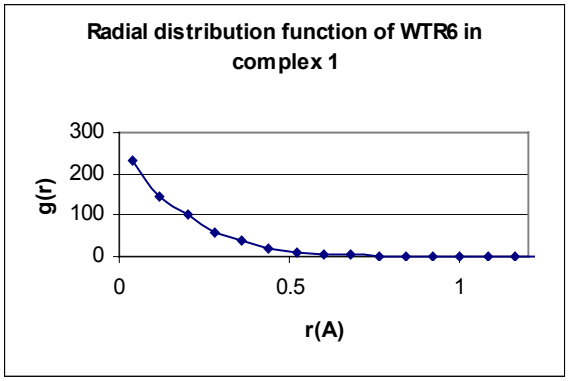
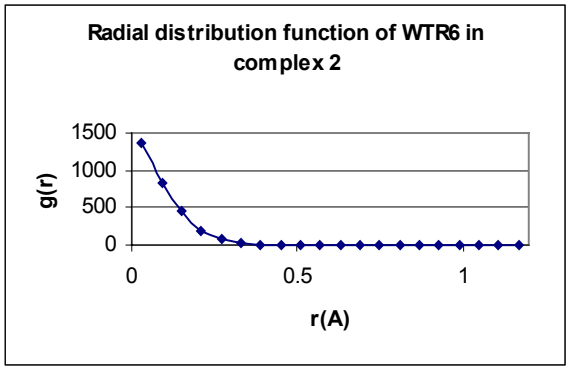


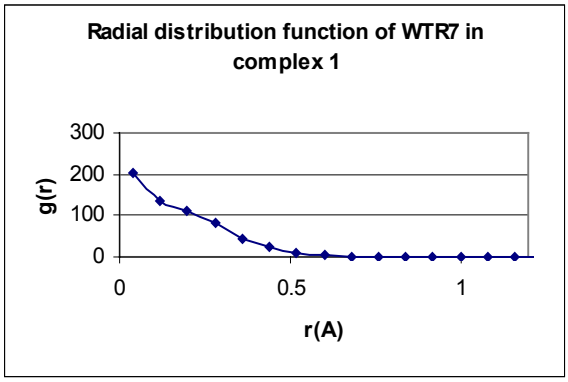
Figure 10. Radial distribution function of different water molecules with respect to their average position in Cyclophilin A (CypA)-Cyclosporin A (**1**), and Cyclophilin A -(5-hydroxynorvalin)-2-cyclosporin (**2**) complexes at 300 K.

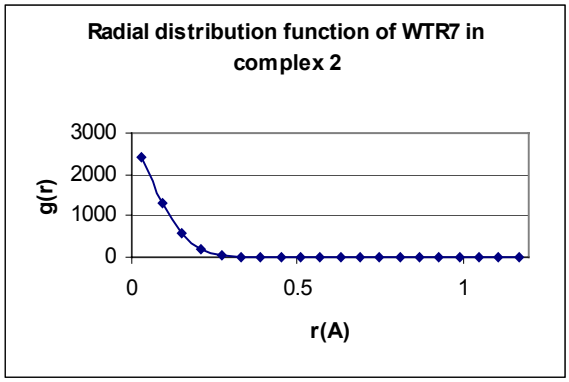

Radial distribution function of WTR5 in complex 1
1000
500
0
g(r)
0
0.5
1
r(A)


Radial distribution function of WTR5 in complex 2
2000
1500
1000
500
0
g(r)
0
0.5
1
r(A)


Radial distribution function of WTR6 in complex 1
300
200
100
0
g(r)
0
0.5
1
r(A)


Radial distribution function of WTR6 in complex 2
1500
1000
500
0
g(r)
0
0.5
1
r(A)


Radial distribution function of WTR7 in complex 1
300
200
100
0
g(r)
0
0.5
1
r(A)


Radial distribution function of WTR7 in complex 2
3000
2000
1000
0
g(r)
0
0.5
1
r(A)

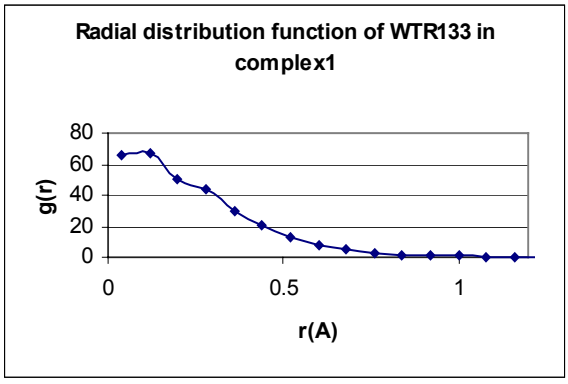
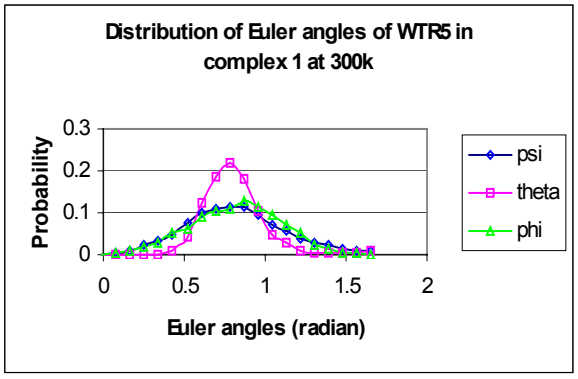
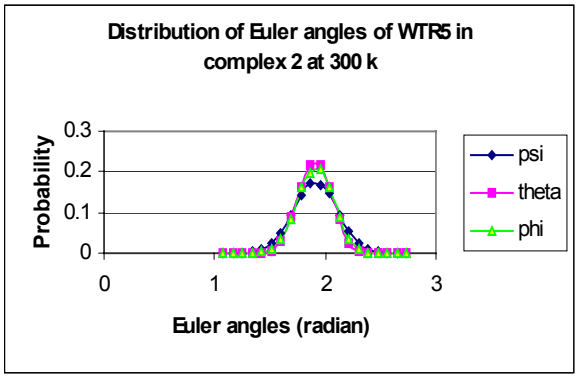
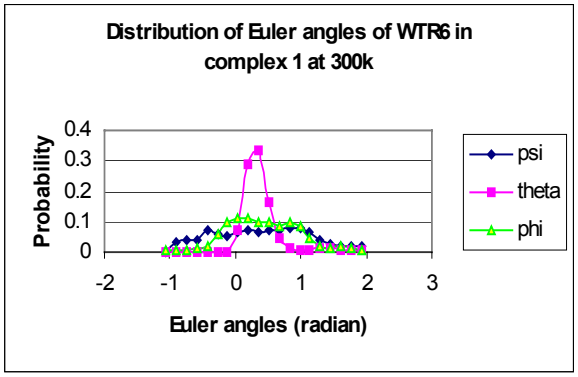
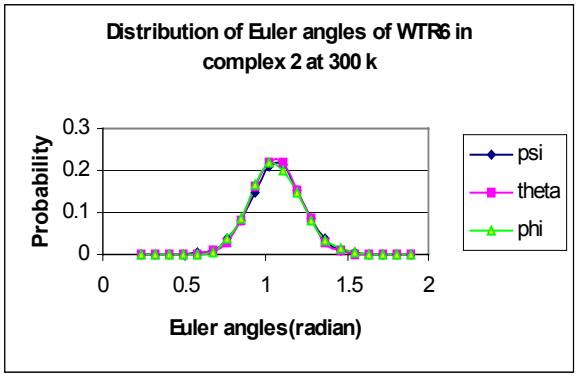

Radial distribution function of WTR133 in complex1
80
60
40
20
0
g(r)
0
0.5
1
r(A)

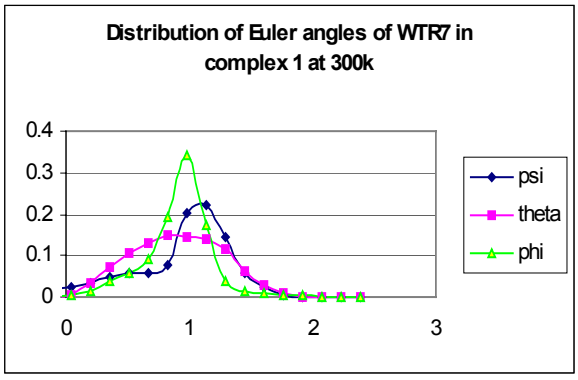
Figure 11. Probability distribution of the Euler angles of different bound water molecules in Cyclophilin A (CypA)-Cyclosporin A (**1**), and Cyclophilin A -(5-hydroxynorvalin)-2-cyclosporin (**2**) complexes.

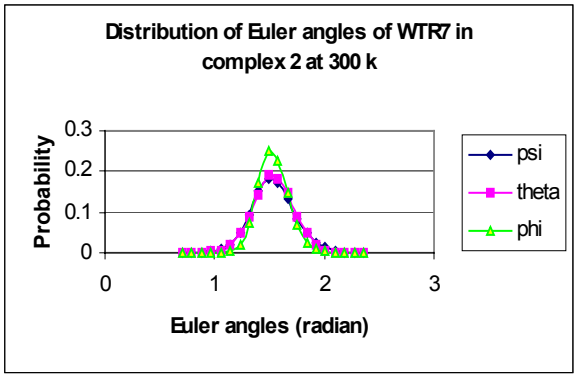

Distribution of Euler angles of WTR5 in complex 1 at 300k
Probability
0.3
0.2
0.1
0
0
0.5
1
1.5
2
Euler angles (radian)
psi
theta
phi


Distribution of Euler angles of WTR5 in complex 2 at 300 k
Probability
0.3
0.2
0.1
0
0
1
2
3
Euler angles (radian)
psi
theta
phi


Distribution of Euler angles of WTR6 in complex 1 at 300k
Probability
0.4
0.3
0.2
0.1
0
-2
-1
0
1
2
3
Euler angles (radian)
psi
theta
phi


Distribution of Euler angles of WTR6 in complex 2 at 300 k
Probability
0.3
0.2
0.1
0
0
0.5
1
1.5
2
Euler angles(radian)
psi
theta
phi


Distribution of Euler angles of WTR7 in complex 1 at 300k
0.4
0.3
0.2
0.1
0
0
1
2
3
psi
theta
phi


Distribution of Euler angles of WTR7 in complex 2 at 300 k
Probability
0.3
0.2
0.1
0
0
1
2
3
Euler angles (radian)
psi
theta
phi

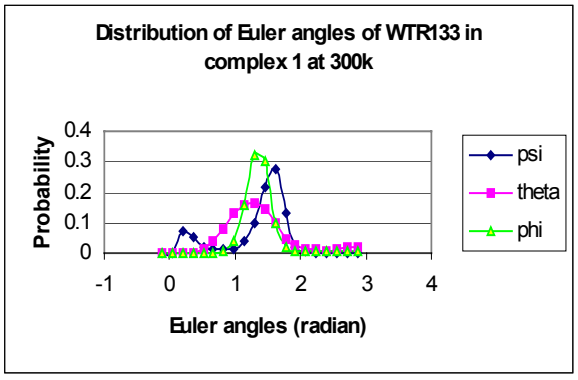

Distribution of Euler angles of WTR133 in complex 1 at 300k
Probability
0.4
0.3
0.2
0.1
0
-1
0
1
2
3
4
Euler angles (radian)
psi
theta
phi

Figure 12. Distribution of probabilities of finding solvent water molecules around the bound water molecule WTR7 in Cyclophilin A (CypA)-Cyclosporin A (**1**) complex during the MDs with different cutoff values of r

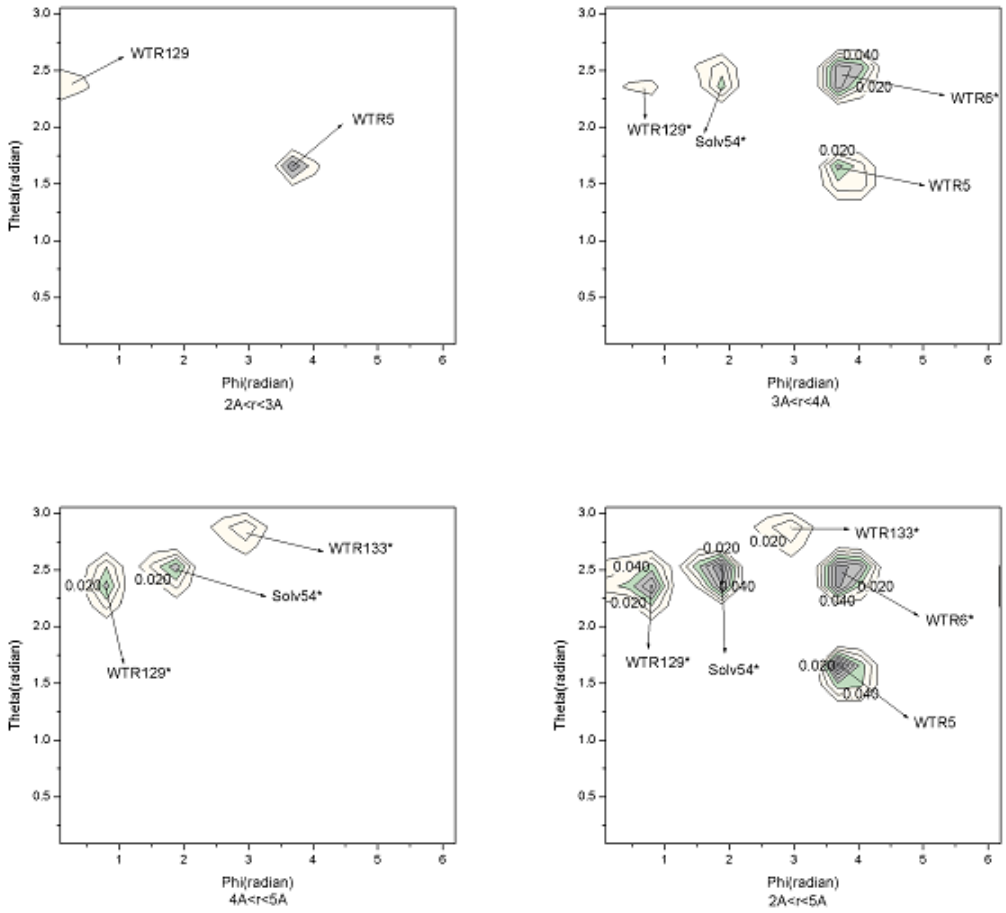


Figure 13. Contribution of water molecules to the solvation free energy in the complexes we studied.

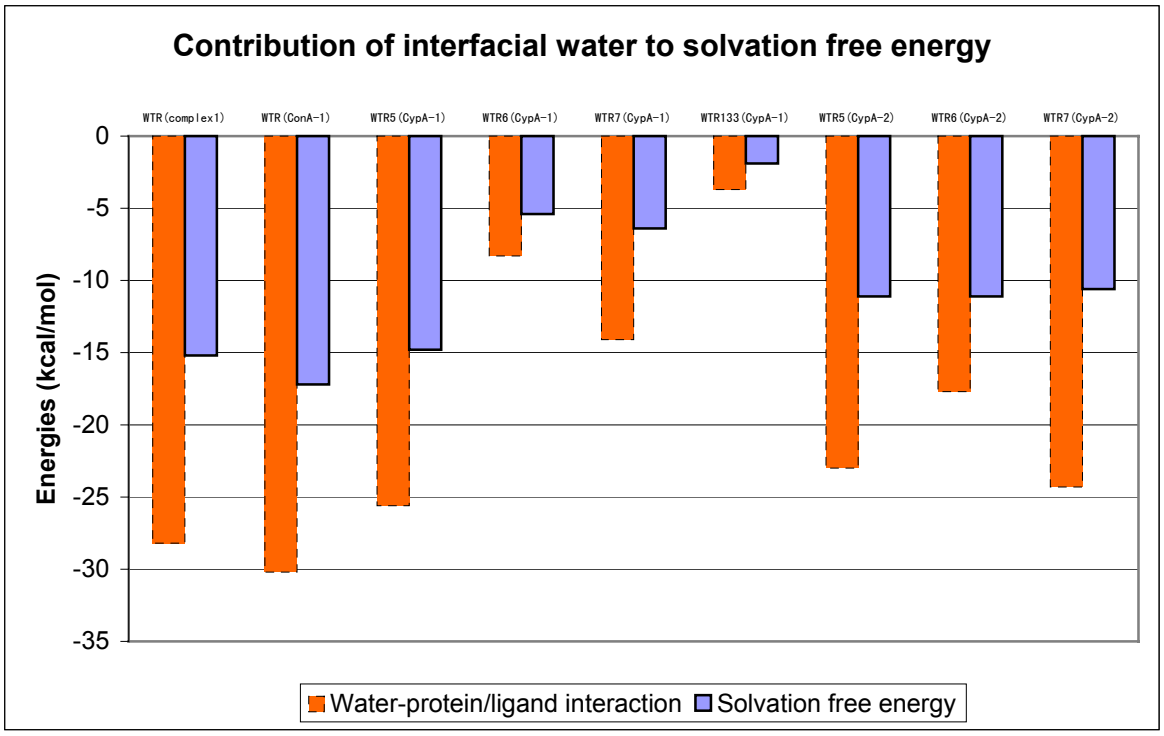
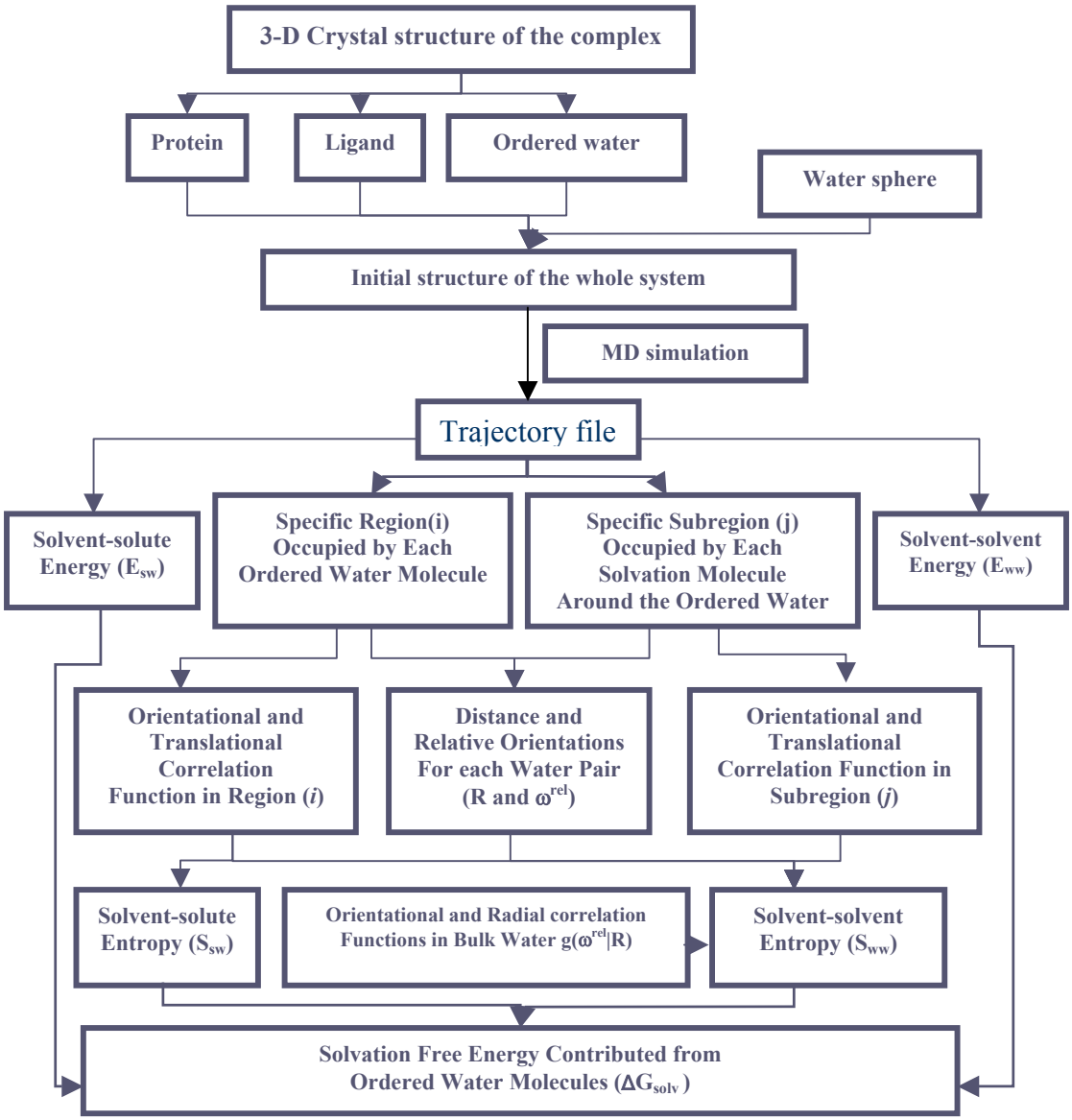


Figure 14. Flow chart of calculating the contributions of ordered water molecules to solvation thermodynamics (COWMST)



References:

1. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335, 321-9.

2. Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1989). Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* 340, 404-407.

3. Mikol, V., Kallen, J., Pflugl, G. & Walkinshaw, D. (1993). X-ray structure of monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 A resolution. *J. Mol. Biol.* 234, 1119-1130.

4. Zheng, J., Trafny, E. A., Knighton, D. R., Xuong, N., Taylor, S. S. & Sowadski, J. M. (1993). 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Cryst. D* 49, 362-365.

5. Bhat, T. N. & Poljak, R. J. (1994). Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* 91, 1089-1093.

6. Joachimiak, A., Haran, T. E. & Sigler, P. B. (1994). Mutagenesis supports water mediated recognition in the Trp repressor-operator system. *EMBO J.* 13, 367-372.

7. Tame, J. R. H., Murshudov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F. & Wilkinson, A. J. (1994). The structural basis of sequence-independent peptide binding by OppA protein. *Science* 264, 1578-1581.

8. Huang, K., Lu, W. Y., Anderson, S., Laskowski, M. & James, M. N. G. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structure of Leu 18, Ala 18 and Gly 18 variants of turkey ovomuciod inhibitor third domain complexed with streptomyces griseus proteinase B. *Prot. Sci.* 4, 1985-1987.

9. Holdgate, G. A., Tunnicliffe, A., Ward, W. H. J., Weston, S. A., Rosenbrack, G., Barth, P. T., Taylor, I. W. F., Pauptit, R. A. & Timms, D. (1997). The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. *Biochemistry* 36, 9663-9673.

10. Sleigh, S. H., Tame, J. R. H., Dodson, E. J. & Wilkinson, A. J. (1997). Peptide binding in OppA , the crystal structures of the periplasmic oligopeptide binding protein in the unliganded from and in complex with lysyllysine. *Biochemistry* 36, 9747-9758.

11. Shaltiel, S., Cox, S. & Taylor, S. S. (1998). Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. *Proc. Natl. Acad. Sci. USA* 95, 484-491.

12. Niefind, K., Putter, M., Guerra, B., Issinger, O. G. & Schomburg, D. (1999). GTP plus water mimic ATP in the active site of protein kinase CK2. *Nature Structural Biology* 6, 1100-1103.

13. Babor, M., Sobolev, V. & Edelman, M. (2002). Conserved positions for ribose recognition: importance of water bridging interactions among ATP, ADP and FAD-protein complexes. *J. Mol. Biol.* 323, 523-532.

14. Reddy, C. K., Das, A. & Jayaram, B. (2001). Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.* 314, 619-632.

15. Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* 285, 2177-2198.

16. Rodier, F., Bahadur, R. P., Chakrabarti, P. & Janin, J. (2005). Hydration of protein-protein interfaces. *Proteins: Structure, Function, and Bioinformatics* 60, 36-45.

17. Levitt, M. & Park, B. H. (1993). Water: Now you see it, now you don't. *Structure* 1, 223-226.

18. Ladbury, J. E. (1996). 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.

19. Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein-protein and protein recognition. *Structure with folding and design* 7, R277-R279.

20. Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D. J., Kellogg, G. E. & Mozzarelli, A. (2004). Free energy of ligands binding to protein: Evaluation of the contribution of water molecules by computational methods. *Current Medicinal Chemistry* 11, 3093-3118.

21. Jayaram, B. & Jain, T. (2004). The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* 33, 343-361.

22. Baldwin, E. T., Bhat, T. N., Gunik, S., Liu, B. S., Topol, I. A., Kiso, Y., Mimoto, T., Mitsuya, H. & Erickson, J. W. (1995). Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* 3, 581-590.

23. Wang, Y.-X., Freedberg, D. I., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., Bhat, T. N., Erickson, J. W. & Torchia, D. A. (1996). Bound water molecules

at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR. *J. Am. Chem. Soc.* 128, 12287-12290.

24. Kervinen, J., Thanki, N., Zdanov, A., Tina, J., Barrish, J., Lin, P. F., Colonno, R., Riccardi, K., Samanta, H. & Wlodawer, A. (1996). Structural Analysis of the Native and Drug-resistant HIV-1 Proteinases Complexed with an Aminodiol Inhibitor. *Protein Pept. Lett.* 3, 399-406.

25. Hong, L., Zhang, X. J., Foundling, S., Hartsuck, J. A. & Tang, J. (1997). Structure of a G48H Mutant of HiV-1 Protease Explains How Glycine-48 Replacements Produce Mutants Resistant to Inhibitor Drugs. *FEBS Lett.* 420, 11-16.

26. Louis, J. M., Dyda, F., Nashed, N. T., Kimmel, A. R. & Davies, D. R. (1998). Hydrophilic peptides Derived from the Transframe Region of GAG-POL Inhibit the HIV-1 Protease. *Biochemistry* 37, 2105-2110.

27. Lam, P. Y. S., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C. H., Weber, P. C., Jackson, D. A., Sharpe, T. R. & EricksonViitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263, 380-284.

28. Grzesiek, S., Bax, A., Nicholson, L. K., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Eyermann, C. J., Torchia, D. A., Hodge, C. N., Lam, P. Y. S., Jadhav, P. K. & Chang, C. H. (1994). NMR evidence for the displacement of a conserved interior water molecule in HIV protease by a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.* 116, 1581-1582.

29. Hodge, C. N., Aldrich, P. E., Bacheler, L. T., Chang, C. H., Eyermann, C. J., Garber, S., Grubb, M., Jackson, D. A., Jadhav, P. K., Korant, B., Lam, P. Y. S., Maurin, M. B., Meek, J. L., Otto, M. J., Rayner, M. M., Reid, C., Sharpe, T. R., Shum, L., Winslow, D. L. & EricksonViitanen, S. (1996). Improved cyclic urea inhibitors of the HIV-1 synthesis, potency, resistance profile, pharmacokinetics and X-ray crystal structure. *Chem. Biol.* 3, 301-314.

30. Chen, J. M., Xu, S. L., Wawrzak, Z., Basarab, G. S. & Jordan, B. D. (1998). Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry* 37, 17735-17744.

31. Weber, P. C., Pantoliano, M. W., Simons, D. M. & Salemme, F. R. (1994). Structure-Based design of synthetic Azobenzene ligands for streptavidin. *J. Am. Chem. Soc.* 116, 2717-2724.

32. Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M. & Schreiber, S. L. (1991). Solution structure of FKBP a rotamase enzyme and receptor for FK506 and rapamycin. *Science* 252, 836-839.

33. Connelly, P. R., Aldape, R. A., Bruzzese, F. J., Chambers, S. P., Fitzgibbon, M. J., Fleming, M. A., Itoh, S., Livingston, D. J., Navia, M. A., Thomson, J. A. & Wilson, K. P. (1994). Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc. Natl. Acad. Sci. USA* 91, 1964-1968.

34. Park, S. & Saven, J. (2005). Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins: Structure, Function, and Bioinformatics* 60, 450-463.

35. Loris, R., Maes, D., Poortmans, F., Wyns, L. & Bouckaert, J. (1996). A structure of the complex between concanavalin A and methyl-3,6-di-O-(a-D-mannopyranosyl)-a-D-mannopyranoside reveals two binding modes. *J. Biol. Chem.* 271, 30614-30618.

36. Clarke, C., Woods, R. J., Gluska, J., Cooper, A., Nutley, M. A. & Boons, G. J. (2001). Involvement water in carbohydrate-protein binding. *J. Am. Chem. Soc.* 123, 12238-12247.

37. Mikol, V., Papageorgiou, C. & Borer, X. (1995). 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.

38. Sharrow, S. D., Edmonds, K. A., Goodman, M. A., Novotny, M. V. & Stone, M. J. (2005). Thermodynamic consequences of disrupting a water-mediated hydrogen bond network in a protein: pheromone complex. *Prot. Sci.* 14, 249-256.

39. Braden, B. C., Goldman, E. R., Mariuzza, R. A. & Poljak , R. J. (1998). Anatomy of an antibody molecule: structure, kinetics, thermodynamics and mutational study of the antilysozyme. *Immunological Reviews* 163, 45-57.

40. Yokota, A., Tsumoto, K., Shiroishi, M., Kondo, H. & Kumagai, I. (2003). The role of hydrogen bonding via interfacial water molecules in Antigen-antibody complexation. *J. Biol. Chem.* 278, 5410-5418.

41. Chacko, S., Silverton, E., Kam-morgan, L., Smith-Gill, S., Cohen, G. & Davies, D. (1995). Structure of an antibody-lysozyme complex unexpected effect of a conservative mutation. *J. Mol. Biol.* 245, 261-274.

42. Watson, K. A. & Papageorgiou, A. (1994). Design of inhibitors of glycogen phosphorylase: a study of a- B-C-Glucosides and 1-thio-B-D-Glucose compounds. *Biochemistry* 19, 5745-5758.

43. Dall'Acqua, W., Goldman, E. R., Lin, W. H., Teng, C., Tsuchiya, D., Li, H. M., Ysern, X., Braden, B. C., Li, Y. L., Smith-Gill, S. & Mariuzza, R. A. (1998). A mutanional analysis of binding interactions in antigen-antibody protein-protein complex. *Biochemistry* 37, 7981-7991.

44. Xu, J., Basse, W. A., Quillin, M. L., Baldwin, E. P. & Matthews, B. W. (2001). Structural and thermodynamic analysis of the binding of solvent at internal sites in T4 lysozyme. *Prot. Sci.* 10, 1067-1078.

45. Langhorst, U., Backmann, J., Loris, R. & Steyaert, J. (2000). Analysis of a water mediated protein-protein interactions within RNase T1. *Biochemistry* 39, 6586-6593.

46. Takano, K., Yamagata, Y., Kubota, M., Funahashi, J., Fujii, S. & Yutani, K. (1999). Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and x-ray analysis of six Ser- Ala mutants. *Biochemistry* 38, 6623-6629.

47. Takano, K., Yamagata, Y., Funahashi, J., Hioki, Y., Kuramitsu, S. & Yutani, K. (1999). Contribution of Intra- and intermolecular hydrogen bonds to the conformational stability of human lysozyme. *Biochemistry* 38, 12968-12708.

48. Denisov, V. P., Venu, K., Peters, J., Horlein, H. D. & Halle, B. (1997). Orientational disorder and entropy of water in protein cavities. *J. Phys. Chem. B* 101, 9380-9389.

49. Robinson, C. R. & Sligar, S. G. (1993). Molecular recognition mediated by bound water - a mechanism for star activity of the restriction-endonuclease ecori. *J. Mol. Biol.* 234, 302-306.

50. Brown, M. P., Grillo, a. O., Boyer, M. & Royer, C. A. (1999). Probing the role of water in the tyrptophan repressor-operator complex. *Protein Sci.* 8, 1276-1285.

51. Fried, M. G., Stickle, D. F., Smirnakis, K. V., Adams, C., MacDonald, D. & Lu, P. Z. (2002). Role of hydration in the binding of lac repressor to DNA. *J. Biol. Chem.* 277, 50676-50682.

52. Bergqvist, S., Williams, M. A., O'Brien, R. & Ladbury, J. E. (2004). Heat capacity effects of water molecules and ions at protein-DNA interface. *J. Mol. Biol.* 336, 829-842.

53. Ladbury, J. E., Wright, J. G., Sturtevant, J. M. & Sigler, P. B. (1994). A thermodynamic study of the trp repressor-operator interaction. *J. Mol. Biol.* 238, 669-681.

54. Morton, C. J. & Ladbury, J. E. (1996). Water mediated protein-DNA interactions: The relationship of thermodynamics to structural detail. *Prot. Sci.* 5, 2115-8.

55. Davis, A. M., Teague, S. J. & Kleywegt, G. J. (2003). Application and limitation of X-ray crystallographic data in structure based ligand and drug design. *Angewante Chemie-international edition* 42, 2718-2736.

56. Pitt, R. W. & Goodfellow, J. M. (1991). Modeling of solvent positions around polar groups in proteins. *Protein Engineering* 4, 531-537.

57. Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849-857.

58. Wade, R. C. & Goodford, P. J. (1993). Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of know structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* 36, 148-156.

59. Pastor, M., Cruciani, G. & Watson, K. A. (1997). A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure-activity relationship analysis. *J. Med. Chem.* 40, 4089-4102.

60. Fornabaio, M., Spyrakis, F., Mozzarelli, A., Cozzini, P., Abraham, D. J. & Kellogg, G. E. (2004). Simple, Intuitive Calculations of Free Energy of Binding for Protein-Ligand Complex. 3. The Free Energy Contributions of Structural Water Molecules in HIV-1 Protease Complexes. *J. Med. Chem.* 47, 4507-4516.

61. Miranker, A. & Karplus, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 11, 29-34.

62. Verdonk, M. L., Cole, J. C. & Taylor, R. (1999). SuperStar: A knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* 289, 1093-1108.

63. Kortvelyesi, T., Dennis, S., Silberstein, M., Brown, l. I. & Vajda, S. (2003). Algorithm for computational solvent mapping of proteins. *Proteins* 51, 340-351.

64. Schymkowitz, J. W. H., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F. & Serrano, L. (2005). Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. USA* 102, 10147-10152.

65. Garcia-Sosa, A. T., Mancera, R. L. & Dean, P. M. (2003). Water Score: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *J. Mol. Mod.* 9, 172-182.

66. Rarey, M., Kramer, B. & Lengauer, T. (1999). The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins: Structure, Function, and Bioinformatics* 4, 17-28.

67. Osterberg, F., Morris, G. M., Sanner, M. F., Olson, A. J. & Goodsell, D. S. (2002). Automated Docking to multiple target structures: incorporation of protein

mobility and structural water heterogeneity in autodock. *Proteins: Structure, Function, and Bioinformatics* 46, 34-40.

68. Wang, J., Chan, S. L. & Ramnarayan, K. (2003). Structure-based prediction of free energy changes of binding of PTP1B inhibitors. *Journal of Computer-Aided Molecular Design* 18, 495-513.

69. Wang, R. X., Lu, Y. P. & Wang, S. M. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* 46, 2287-2303.

70. Lloyd, D. G., Garcia-Sosa, A. T., Alberts, I. L., Todorov, N. P. & Mancera, R. L. (2004). The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *Journal of Computer-Aided Molecular Design* 18, 89-100.

71. De Graaf, C., Pospisil, P., Pos, W., Folkers, G. & Vermeulen, N. P. E. (2005). Binding mode prediction of Cytochrome P450 and Thymidine kinase protein-ligand complexes by considerations of water and rescoring in automated docking. *J. Med. Chem.* 48, 2308-2318.

72. Verdonk, M. L., Chessari, G., Cole, J. C., Hartschorn, M. J., Murray, C. W., Nissink, J. W. M., Taylor, R. D. & Taylor, R. (2005). Modeling water molecules in protein-ligand docking using gold. *J. Med. Chem.* 48, 6504-6515.

73. Amadasi, A., Spyrakis, F., Cozzini, P., Abraham, D. J., Kellogg, G. E. & Mozzarelli, A. (2006). Mapping the energetics of water-protein and water-ligand interactions with the "natural" HINT force field: predictive tools for characterizing the role of water in biomolecules. *J. Mol. Biol.* 358, 289-309.

74. Cozzini, P., Fornabaio, M., Mozzarelli, A., Spyrakis, F., Kellogg, G. E. & Abraham, D. J. (2006). Water: how to evaluate its contribution in protein-ligand interactions. *International Journal of Quantum Chemistry* 106, 647-651.

75. Mancera, R. L. (2002). De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *Journal of Computer-Aided Molecular Design* 16, 479-499.

76. Schnecke, V. & Kuhn, L. A. (2000). Virtual screening with solvation and ligand induced complementarity. *Perspectives in drug descovery and design* 20, 171-190.

77. Friesner, R., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P. & Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. mehtod and assessment of docking accuracy. *J. Med. Chem.* 47, 1739-1749.

78. Pospisil, P., Kuoni, T., Scapozza, L. & Folkers, G. (2002). Methodology and problems of protein-ligand docking: case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterae 4. *J. Recept. Signal Transduction Res.* 22, 141-154.

79. Yang, J. M. & Chen, C. C. (2004). GEMDOCK: a generic evolutionary method for molecular docking. *Protein Sci.* 55, 288-304.

80. Birch, L., Murray, C. W., Hartschorn, M. J., Tickle, I. J. & Verdonk, M. L. (2002). Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Computer-Aided Molecular Design* 16, 855-869.

81. Jiang, L., Kuhlman, B., Kortemme, T. & Baker, D. (2005). A "solvated reotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interface. *Proteins: Structure, Function, and Bioinformatics* 58, 893-904.

82. Papoian, G. A., Ulander, J. & Wolynes, P. G. (2003). Role of water mediated interactions in protein-protein recognition landscapes. *J. Am. Chem. Soc.* 125, 9170-9178.

83. Papoian, G. A., Ulander, J., Eastwook, M. P., Luthey-Schulten, Z. & Wolynes, P. G. (2004). Water in protein structure prediction. *Proc. Natl. Acad. Sci. USA* 101, 3352-3357.

84. Zong, C., Papoian, G. A., Ulander, J. & Wolynes, P. G. (2006). Role of topology, nonadditivity, and water-mediated interactions in predicting the structures of a/B proteins. *J. Am. Chem. Soc.* 128, 5168-5176.

85. Dunitz, J. D. (1994). The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* 264, 670.

86. Covell, D. G. & Wallquist, A. (1997). Analysis of protein-protein interactions and the effects of amino acid mutations on their energetics. The importance fo water molecules in the binding epitope. *J. Mol. Biol.* 269, 281-297.

87. Petrone, P. M. & Garcia, A. E. (2004). MHC-peptide binding is assisted by bound water molecules. *J. Mol. Biol.* 338, 419-435.

88. Fischer, S., Smith, J. C. & Verma, C. S. (2001). 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.

89. Verma, C. S. & Fischer, S. (2005). Protein stability and ligand binding: new paradigms from in-silico experiments. *Biophysical Chemistry* 115, 295-302.

90. Brandl, M., Meyer, M. & Suhnel, J. (2000). Water-mediated base pairs in RNA: A quantum chemical study. *J. Phys. Chem. A* 104, 11177-11187.

91. Tschampel, S. M. & Woods, R. J. (2003). Quantifying the role of water in protein-carbohydrate interactions. *J. Phys. Chem. A* 107, 9175-9181.

92. Wade, R. C., Mazor, M. H., McCammon, J. A. & Quiocho, F. A. (1991). A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers* 31, 919-931.

93. Roux, B., Nina, M., Pomes, R. & Smith, J. C. (1996). Thermodynamic Stability of Water Molecules in the Bacteriorhodopsin Proton Channel: A Molecular Dynamics Free Energy Perturbation Study. *Biophysical Journal* 71, 670-681.

94. Zhang, L. & Hermans, J. (1996). Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Bioinformatics* 24, 433-438.

95. Hamelberg, D. & McCammon, J. A. (2004). Standard free energy of releasing a localized water molecule from the binding pockets of protein: double-decoupling method. *J. Am. Chem. Soc.* 126, 7683-7689.

96. Olano, R. L. & Rick, S. W. (2004). Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* 126, 7991-8000.

97. Tashiro, M. & Stuchebrukhov, A. A. (2005). Thermodynamic properties of internal water molecules in the hydrophobic cavity around he catalytic center of cytochrome c oxidase. *J. Phys. Chem. B* 109, 1015-1022.

98. Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* 72, 1047-1069.

99. Archontis, G., watson, K. A., Xie, Q., Andreou, G., Chrysina, E. D., Zographos, S. E., Oikonomakos, N. G. & Karplus, M. (2005). Gycogen phosphorylase inhibitors: A free energy perturbation analysis of glucopyranose spirohydantionanalogues. *Proteins: Structure, Function, and Bioinformatics* 61, 984-998.

100. Monecke, P., Borosch, T., Brickmann, J. & Kast, S. M. (2006). Determination of the interfacial water content in protein-protein complexes from free energy simulations. *Biophys. J.* 90, 841-850.

101. Lazaridis, T. (1998). Inhomogeneous fluid approach to solvation thermodynamics 1. Theory. *J. Phys. Chem. B* 102, 3531-41.

102. Lazaridis, T. (1998). Inhomogeneous fluid approach to solvation thermodynamics. 2. Application to simple fluids. *J. Phys. Chem. B* 102, 3542-3550.

103. Lazaridis, T. & Karplus, M. (1996). Orientational correlations and entropy in liquid water. *J. Chem. Phys.* 105, 4294-4316.

104. Lazaridis, T. (2000). Solvent reorganization energy and entropy in hydrophobic hydration. *J. Phys. Chem. B* 104, 4964-79.

105. Mackerell, A. D., Bashford, D., Bellott, M., Dunback, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S. N., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S. W., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586-3616.

106. Brooks, C. L. & Karplus, M. (1983). Deformable stochastic boundaries in molecular-dynamics. *J. Chem. Phys.* 79, 6312-6325.

107. Weis, W. I. & Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem* 65, 441-473.

108. Makhatadze, G. I. & Privalov, P. L. (1993). Contribution of hydration to protein-folding thermodynamics .1. The enthalpy of hydration. *J. Mol. Biol.* 232, 639-659.

109. Privalov, P. L. & Makhatadze, G. I. (1993). Contribution of hydration to protein folding thermodynamics .2. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* 232, 660-679.

110. Plyasunov, A. V. & Shock, E. L. (2000). Thermodynamic functions of hydration of hydrocarbons at 295.15 K and 0.1 MPa. *Geochimica et Cosmochimica Acta* 64, 439-468.

111. Cabani, S., Gianni, P., Mollica, V. & Lepori, L. (1981). Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous-solution. *J. Solu. Chem.* 10, 563-595.

112. Wang, J., Szewczuk, Z., Yue, S.-Y., Tsuda, Y., Konishi, Y. & Purisima, E. O. (1995). Calculation of relative binding free energies and configurational entropies: a structural and thermodynamic analysis of the nature of nonpolar binding of thrombin inhibitors based on hirudin. *J. Mol. Biol.* 253, 473-492.

113. Luo, R. & Gilson, M. K. (2000). Synthetic adenine receptors: Direct calculation of binding affinity and entropy. *J. Am. Chem. Soc.* 122, 2934-7.

114. Pickett, S. D. & Sternberg, M. J. E. (1993). Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* 231, 825-839.

115. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Prot. Sci.* 11, 281-296.

116. Creamer, T. P. (2000). Side-chain conformational entropy in protein unfolded states. *Protein Sci.* 40, 443-450.

117. D'Aquino, J. A., Freire, E. & Amzel, L. M. (2000). Binding of small organic molecules to macromolecular targets: evaluation of conformational entropy changes. *Protein Sci.* Suppl 4, 93-107.

118. Mikol, V., Kallen, J., Pflugl, G. & Walkinshaw, D. (1995). X-ray structure of a monomeric cyclophilin A-cyclsporin A crystal complex at 2.1 A resolution. *J. Mol. Biol.* 224, 1119-1130.

119. Garcia, H. E. & Hernadez, A. A. (1999). Structural bases of lectin-carbohydrate affinities: comparison with protein-folding energetics. *Protein Sci.* 8, 1075-1086.

120. Lazaridis, T. & Karplus, M. (2003). Thermodynamics of protein folding: A microscopic view. *Biophys. Chem.* 100, 367-395.

121. Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* 10, 139-145.

122. Garcia-Sosa, A. T., Firth-Clark, S. & Mancera, R. L. (2005). 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.

123. Li, Z. & Lazaridis, T. (2003). Thermodynamic Contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.* 125, 6636-6637.

124. Li, Z. & Lazaridis, T. (2005). The effect of water displacement on binding thermodynamics: Concanavalin A. *J. Phys. Chem. B* 109, 662-670.

125. Li, Z. & Lazaridis, T. (2006). Thermodynamics of buried water clusters at a protein-ligand binding interface. *J. Phys. Chem. B* 110, 1464-1475.

126. Schaefer, M. & Karplus, M. (1996). A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* 100, 1578-1599.

127. Qiu, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. (1997). The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* 101, 3005-3014.

128. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Structure, Function, and Bioinformatics* 35, 133-152.