Solvation Thermodynamic Costs of forming Cognate Binding Site Formation

Yeonji Ji[1], Vjay Molino[1], Steven Ramsey[2], Tom Kurtzman[1,2,3]

[1] Ph.D. program in Biochemistry, The Graduate Center, City University of New York, New York, USA
[2] Department of Chemistry, Lehman College, City University of New York, Bronx, USA
[3] Ph.D. Programs in Biology & Chemistry, The Graduate Center, City University of New York, New York, USA

1 Abstract

Proteins are inherently flexible which complicates the identification of lead molecules that are shape and charge complementary to target proteins. While significant effort has been dedicated to exploring alternate protein conformations, solvation thermodynamics has typically not been integrated into these studies. Here, we study how solvation thermodynamics fluctuate as proteins adopt different conformations. We analyze solvation thermodynamics within the binding cavities of conformations for which side chains are mobile in molecular dynamics simulations and compare these to conformations for which they remain restrained about the cognate bound structure. We identify structural motifs that present significant costs to the sampling of cognate ligand bound structures. We find that the reorganization of protein side chains has a significant effect on the structure and thermodynamics of binding site solvation. We discuss how understanding the interplay between solvation thermodynamics and protein structural fluctuations is crucial for both discovering alternative binding pockets, estimating the contribution to binding affinity of displacing water upon ligand binding, and assessing revealed cryptic pocket bindability.

2 Introduction

The solvation contribution to the binding affinity between a small molecule and protein is fully described by the difference in solvation free energy between an initial state in which the protein and ligand are unbound a final bound state in which the ligand and protein are bound.[1] In the initial state, the protein and ligand are independently solvated and have conformational flexibility that is greater than in the final state in which the protein and ligand are limited to conformations that are complementary to each other (Figure 1). Solvation mapping methods based upon Inhomogeneous fluid Solvation Theory[2,3] (IST) such as WaterMap[4] and Grid Inhomogeneous Solvation Theory[5,6] (GIST)have been used to estimate the free energy of displacing solvent from the binding cavity upon ligand recognition. The initial formulation of IST relies upon Percus' source particle method which applies to a rigid conformation of the solute[7] and most applications of IST such as GIST and WaterMap have relied upon this rigid solute approximation using molecular dynamics simulations with heavy atoms restrained as in the *rigid* simulations here. Correspondingly, applications that estimate the solvent displacement of water from the binding cavity generally use estimates of the thermodynamics of binding site solvation of *rigid* systems as opposed to *flexible* systems therefore approximating the solvation contribution as the difference between panels c and d in Figure 1.

Based on our prior work[8] which describes the strong coupling between host system conformations and the thermodynamics of solvation, we hypothesized that solvation thermodynamics is strongly coupled to structural fluctuations of protein binding sites. Here, using GIST and Solvation Structural and Thermodynamic Mapping[9] (SSTMap) methods, we computationally investigate the solvation thermodynamics of forming cognate (ligand bound) structures of proteins and estimate the thermodynamic cost of forming such conformations in protein models. We focus this study on part of a thermodynamic end-states process in which the initial state is an unbound state in which the protein has already adopted the cognate bound backbone structure but has freely moving side chains (*flexible*) to an end-state in which the side chains are held rigid in the cognate bound configuration (*rigid*). This allows us to investigate the validity of the rigid receptor approximation by calculating the difference of the solvation free energy between the *rigid* and *flexible* receptors. We find that the solvation free energy penalties to cognate structure formation for these end-states are significant for all 34 systems investigated with average energetic, entropic, and free energetic penalties estimated to be 14.97 kcal/mol, 20.72 kcal/mol, and 35.69 kcal/mol, respectively. We further investigate how water structure varies between the *rigid and flexible* systems and describe how water complementarity to the protein surface (water's ability to form complementary hydrogen bonds with the protein surface) and enhancement or frustration of water networks solvating the binding site contributes to the substantial solvation free energetic penalties of ligand bound protein conformations. Consistent with our prior studies, we identify cognate binding site topographies in these systems that frustrates binding site water structure which contributes to these large solvation thermodynamic penalties.[10]

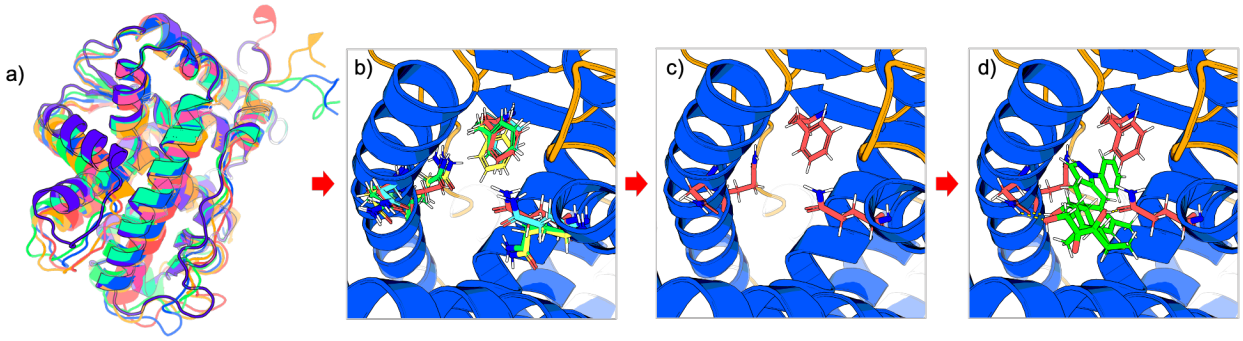

Figure 1. A thermodynamic path of the binding between Glucocorticoid receptor (GCR) and ligand DAY broken down into four states: (a) An initial state in which the ligand and protein are fully flexible and unbound (ligand not shown, (b) an intermediate state in which the protein backbone is restrained about cognate bound conformation but the sidechains are flexible, (c) a 2nd intermediate state in which both the backbone and sidechains of GCR are restrained about their cognate bound conformation, and (d) The protein-ligand complex in its bound structure. (pdbid:3bqd). Differently colored backbone and ligands are representative of their respective conformations at each state.

Finally, a number of methods have been developed to estimate or map out solvation thermodynamics on the surfaces of proteins including 3D-RISM[11–13], GIST, WaterMap, Szmap[14,15], and others.[16–20] Though these tools have provided valuable insight into target binding sites and enabled drug design ideation, they often rely on the assumption that the target binding site is rigid. This assumption is pervasive in solvation mapping methods and we discuss in the final section how the strong coupling between solvation thermodynamics and protein binding site fluctuations may impact the application of these methods in drug discovery settings.

The structure of the paper is as follow:

In the methods section, we first describe details of the protein preparation, molecular dynamics simulations, solvation thermodynamic mapping, and analyses of solvation structural feature. The results section encompasses a comparison of solvation energetic, entropic, and free energetic costs to the formation of *rigid* structures from the *flexible* structures. Following that, we investigate how water structure and water complementarity to the protein surface varies in *rigid* and *flexible* systems by examining the solute–water and water–water interactions within the binding sites. We conclude the present paper by discussing the solvation thermodynamic costs to cognate binding site formation and how this may impact existing computational approaches of exploring protein conformational space.

3 Methods

3.1 Methods

We performed explicit solvent molecular dynamics (MD) simulations using Amber20[21] and Amber22[22] on a subset of 34 proteins from the Database of Useful Decoys -- Enhanced (DUD-E).[23] In one set of simulations, all heavy atoms were harmonically restrained about a protein structure that was minimized in the presence of the cognate ligand (termed *rigid*) and in a second set of simulations, the side chain heavy atoms were left unrestrained (termed *flexible*). In both sets of simulations the proteins were simulated fully solvated with no ligand.

3.1.1 System Preparation and MD simulation

a) Selection and Preparation of Proteins:

We investigate the subset of protein-ligand complexes from the DUD-E database for which the cognate ligand made three or more hydrogen bonds with binding site residues (34 of 102 systems). The structures of the 34 systems were obtained from the Protein Data Bank (PDB).[24] All hetero atoms except for the cognate ligand and water were removed. For proteins with multiple domains, we retained only the chains necessary to investigate the binding site. The PDB ID of each system, chains, and cognate ligand ID used for further steps can be found in Table S1 of the Supplementary Information. The complexes were then prepared using the Protein Prep Wizard[25,26] in Schrodinger Maestro[27] with default settings. This step assigned protonation states, filled gaps in the structure, optimized side chain orientations, and capped the protein termini with N-acetyl and N-methyl amide groups. Water molecules farther than 5 Å from the ligand were removed. Protein, Ligand, and Water forcefields:

Protein atoms were parameterized with tleap using the ff14SB forcefield[28] and solvated using the OPC[29] water model with a minimum water butter of 10 Å.

For the *rigid* simulations, the protein structure was energetically minimized in the presence of the ligand which was parameterized with the OpenFF Sage Forcefield 2.0.0.[30]

Protein-ligand complex minimization:

For the *rigid* simulations, the protein configuration was first energetically minimized in the presence of the ligand. The complexes were first solvated in OPC water with a minimum buffer of 10 Å in tleap. The water was then minimized using 1500 steps of steepest descent with all heavy atoms of the ligand protein restrained with restraint weight of 100 kcal/mol·Å^{-2}. This was followed by a 2nd energetic minimization of 1500 steps in which only the backbone heavy atoms of the protein were restrained with the same force constant. The water and ligand were then removed and the resulting protein configurations were used for the solvated protein minimizations for the

rigid simulations. The remainder of the preparations steps (below) were identical for both the *rigid* and *flexible* simulations.

b) Solvated Protein Simulation Energy Minimization:

The proteins were first solvated in a box of OPC water with a minimum buffer of 10 Å in tleap. The systems were then energetically minimized using 1500 steps of steepest descents with the water being unrestrained and all heavy atoms of the protein being restrained with a force constant of 100 kcal/mol/Å^2. In the preparation of the *flexible* simulations, the systems were then energetically minimized using 1500 steps of steepest descents with only the backbone heavy atoms of the protein being restrained with the same force constant.

c) Equilibration:

The energetically minimized systems were then equilibrated with molecular dynamics simulations. First, the systems were heated from 0 K to 300 K over 240 ps at constant volume and temperature using the Langevin thermostat with a collision frequency of 2 ps^{-1}.[31–33] This was followed by a 10 ns MD simulation at a constant temperature of 300 K and pressure of 1 bar using the same thermostat and the position scaling barostat[34] with a relaxation time of 0.5 ps^{-1} in which the protein atom restraints were decreased gradually from an initial value of 100 kcal/mol/Å^2 to the production run restraint strength of 2.5 kcal/mol/Å^2. This was followed by a second equilibration MD run of 10 ns at constant temperature and pressure with 2.5 kcal/mol/Å^2 restraints and the same thermostat and barostat.

d) MD Production Runs:

The production MD runs were 80 ns at constant volume and a temperature of 300 K regulated by the Langevin Thermostat with a time constant of 1 ps for the heat bath coupling and a collision frequency of 1 ps^{-1}. The equilibration and production MD runs were performed using GPU-accelerated PMEMD[35,36] with a time step of 2 fs and the SHAKE algorithm[37] was used to constrain bond lengths involving hydrogens. The protein and water configurations were output every 1 ps yielding 80,000 frames.

3.1.2 GIST and SSTMap Solvation Mapping

Grid Inhomogeneous Solvation Theory (GIST) is a computational method that post-processes MD trajectories to estimate and map solvation thermodynamic quantities onto a high-resolution grid. Thermodynamics quantities are estimated using a spatial discretization of the equations of Inhomogeneous Solvation Theory (IST). For the *rigid* and *flexible* simulations of all 34 systems, GIST was applied to the 80 ns production runs using an in-house version of cpptraj GIST. The in-house version of cpptraj GIST varied from the public version in that it had additional code to map out hydrogen bond properties of water on the grid.

SSTMap utilizes a hydration site approach (HSA) to map out solvation and structural properties of water in high density 1 Å radius spherical regions (hydration sites). Structural and thermodynamic quantities of the water in each hydration site are calculated from an analysis of MD trajectories. Hydrations Site analyses using SSTMap were performed on the last 20 ns of the production runs.

a) GIST details:

The grid dimensions were set to ensure the GIST box included the whole system and were set independently for each simulated system. The grid spacing was 0.5 Å along each axis yielding a

voxel volume of 0.125 Å^3. All default quantities were output including the GIST estimated energy, entropy, and free energy densities in each voxel as well as the corresponding per water quantities.

The number density for neat OPC water for GIST entropy calculations was .0333 molecules/Å^3. The total water energy values, $E_{tot,}$ are referenced to the neat OPC water energies (-12.259 kcal/molecule) of the same number of water molecules.

Error analysis on the GIST quantities used block averaging with the 80ns production runs being separated into four 20 ns blocks.

To account for the entropy of correlations between water molecules which GIST does not explicitly calculate, we implemented a simple scaling correction of 0.6 to the GIST calculated entropy.[38–40]

b) Binding cavity Sub-volumes and Integrated Thermodynamic Quantities:

The binding cavity sub-volumes were defined by the set of voxels whose center was within a specified distance of any heavy atom of the system's cognate ligand. The MD simulations were run without the ligand presents so the volume is with reference to the aligned co-crystal ligand coordinates. The properties of water were calculated in sub-volumes ranging from 3 Å to 10 Å from the cognate ligand, with increments of 0.5 Å (Figure 2). GIST post processing (GISTPP)[6] was used to define the sub-volumes and for calculating the thermodynamic properties of water within each sub-volume.

The total values for the GIST quantities in a sub-volume are simply the sum of the voxel quantities for all voxels in the sub-volume. The per water quantities are simply this value divided by the number of water molecules. As the number of water molecules in the cavities varied between the *flexible* and *rigid* systems, we also estimated what the total value would comparing the same number of water molecules in the binding cavity for the *flexible* and *rigid* systems. This is simply the number of water molecules in the cavity multiplied by the average value of water molecules in the cavity. For these comparisons the number of water molecules was chosen to the be the lesser value of the number found in the binding cavity in the *flexible* and *rigid* systems. We refer to these values collectively as *integrated* values.

The lower limit of 3 Å from any ligand heavy atom for the sub-volume was chosen because it has often been used to approximate the volume from which water is displaced upon ligand binding.[41,42] We refer to this volume as V_{disp}. The upper limit of 10 Å was chosen based on our prior study, Chen et al., which showed that integrating the solvation energy, entropy, and free energy over this sub-volume was sufficient to closely estimate the solvation free energy of small molecules.[38] We refer to the 10Å sub-volume as V_{cav}.

c) Solvent Accessible Surface Area

GISTPP was used to generate the solvent accessible surface areas using the default density inputs.

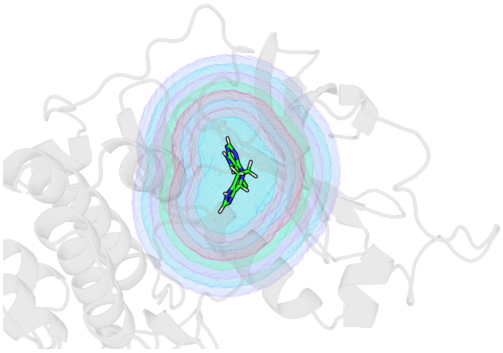

Figure 2. A depiction of binding site sub-volumes about the cognate ligand. Though the ligand is not present in the simulations, the sub-volumes are defined by the regions within a certain distance of any heavy atom the aligned co-crystal ligand coordinates. The inner blue region is the smallest sub-volume defined by all GIST voxels whose center is within 3 Å of the ligand. Each differently shaded region here is demarcated at 1 Å intervals. The full shaded region is the largest sub-volume which includes all voxels within 10 Å of the ligand. The integrated quantities are the sum of all voxel quantities over the region.

d) Geometric Definition of Hydrogen Bonds & Neighbors:

All hydrogen bond data reported used a geometric definition of a hydrogen bond in which a non-covalent polar interaction is considered a hydrogen bond when the distance between the two heavy atoms is less than 3.5 Å and the angle of acceptor–donor–hydrogen is less than 30°.[43] Our in-house version of cpptraj GIST mapped out the densities of water-protein and water-water hydrogen bonds on the GIST grid. Water molecules and/or heavy atoms are considered to be "neighbors" if they are within 3.5 Å of each other.

e) Hydrogen Site Analyses

SSTMap with default settings was applied to the last 20 ns of the *rigid* and *flexible* simulations for all 34 systems investigated. Hydration sites were characterized as either an acceptor or donor if 60% or more of the water molecules found in the hydration site were making the appropriate acceptor or donor interaction. If both values were above 60% then the hydration site was characterized as both donor and acceptor.

4 Results

4.1 Solvation Thermodynamics in the Binding Cavity.

We used GIST to estimate the binding cavity solvation energy, entropy, and free energy for 34 *rigid* and *flexible* proteins. As detailed in the methods section, the term "*rigid"* refers to simulations of systems in which all heavy atoms are restrained to their positions in the cognate crystal structure during the simulation, and the term "*flexible"* refers to results from simulations for systems in which the side chain heavy atoms are left unrestrained. We find that, on average, the GIST estimated energy, entropy, and free energy of solvating the binding cavities is significantly more favorable for *flexible* systems than for the *rigid* systems (Figure 3). For the 10 Å sub-volume of the binding cavity, the GIST estimated difference in average energy is 14.43 kcal/mol, in entropic contribution to the free energy ($T\Delta S$) is 17.96 kcal/mol, and in free energy is 32.39 kcal/mol. On a per water molecule basis, the average difference in solvent energy, entropy

and free energy are 0.17 kcal/mol, 0.20 kcal/mol, and 0.38 kcal/mol, respectively with the *flexible* systems being more favorable for all three.

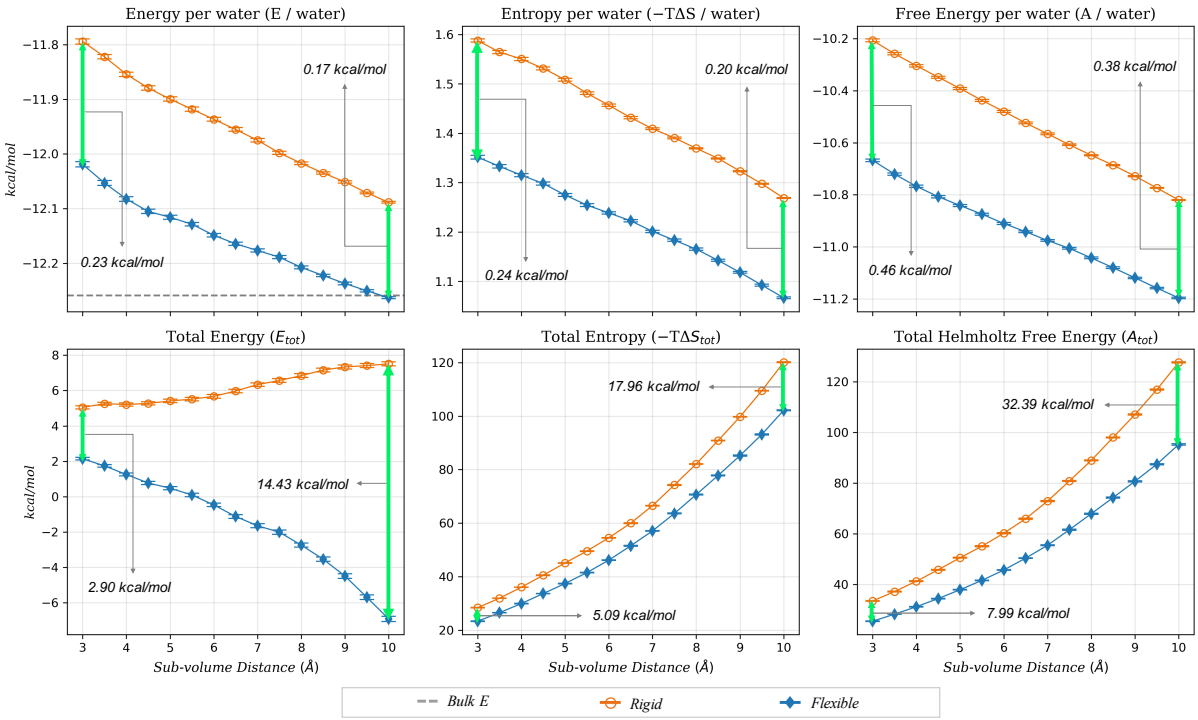

Figure 3. The total solvent energy (E_{tot}), entropy (-TΔS), and Helmholtz free energy (A_{tot}) per water values (y-axes) averaged over the 34 systems within each sub-volume (x-axes). Data from the rigid simulations are hollow orange circles. The total energy is referenced to the same number of water molecules as found in the appropriate sub-volume in a neat system at the same thermodynamic conditions. The error bars are the error in the mean.

In the following sections, we will provide a detailed report of the data, separated into sections for the binding cavity solvation energy, entropy, and free energy. Our analysis will mainly focus on the 10 Å and the 3 Å sub-volume. In our previous work investigating the solvation of small molecule hydration, we found that a GIST analysis of water in the 10 Å sub-volume fully accounted for the full thermodynamics of solvation[38] and, for ligand binding, includes both the region from which solvent is displaced and the region in which water restructures itself around a complex respectively referred to as V_{disp} and V_{rest} in Gilson et al.[44] Based on this prior work, we will use the 10 Å sub-volume around the cognate ligand as an estimate of the full solvation free energy of the binding site for each protein and refer to the volume as V_{cav}. We also focus on the 3 Å region as, in prior work[4,41,45], displacement of water from this region has commonly used this as an estimator of the thermodynamics of water displacement upon ligand binding; we refer to this volume as V_{disp}.

4.2 Solvation Energy

The average total energy and energy per water for the V_{disp} and V_{cav} regions are summarized in Table 1. Although our analysis is focused on the 10 Å sub-volume (V_{cav}), the average values in

the smaller volume (V_{disp}, 3 Å sub-volume), which refers to the solvent displacement region, are also significantly different for the *flexible* and *rigid* systems. In this region (V_{disp}), the average difference in the total energy and energy per water between two conformations is 2.90 kcal/mol and 0.23 kcal/mol, respectively.

The average energy of water molecules in the binding cavity is more favorable in the *flexible* than in the *rigid* cavities and this trend holds for every one of the 34 systems (Figure 4) that we investigated, however, the range of per water energy differences is from almost negligible (0.02 kcal/mol for XIAP, PDB id: 3HL5[46]) to considerable (0.58 kcal/mol for SAHH, PDB id: 1LI4[47]). The total solvent energy in the binding cavity volume (V_{cav}) is more favorable in the *flexible* cavities for 33 of the 34 systems with Peroxisome proliferator-activated receptor delta (PPARD, PDB id: 2ZNP[48]) being an outlier and the total solvent energy has a considerable range from -3.82 kcal/mol for PPARD to 47.00 kcal/mol for Angiotensin-converting enzyme (ACE, PDB id: 3BKL[49]). When comparing the solvent energy in sub-volumes between the flexible and rigid system, the volumes are the same however the number of water molecules in that volume differ. For PPARD, the *flexible* cavity (V_{cav}) has 11.98 more water molecules on average than the *rigid* cavity and, as the water molecules on average in both systems are unfavorable compared to neat water (to which the energies are referenced) this leads to the total solvent energy somewhat misleadingly being less favorable in the *flexible* system compared to the *rigid*. When the same number of water molecules in the cavity are compared, however, the solvent energy in the *flexible* cavity of PPARD is 6.06 kcal/mol more favorable than for the *rigid* system.

Table 1. The average total energy (E_{tot}) and energy per water (E/water) for the 34 systems in sub-volumes V_{disp} (3 Å) and V_{cav} (10 Å). The total energy is referenced to neat water. Units are kcal/mol

Volume	E_{total}			E / water		
	Rigid	*Flexible*	ΔE_{total} *(Rigid – Flexible)*	*Rigid*	*Flexible*	ΔE / water *(Rigid – Flexible)*
V_{cav}	7.51	-6.92	14.43	-12.09	-12.26	0.17
V_{disp}	5.05	2.14	2.90	-11.79	-12.02	0.23

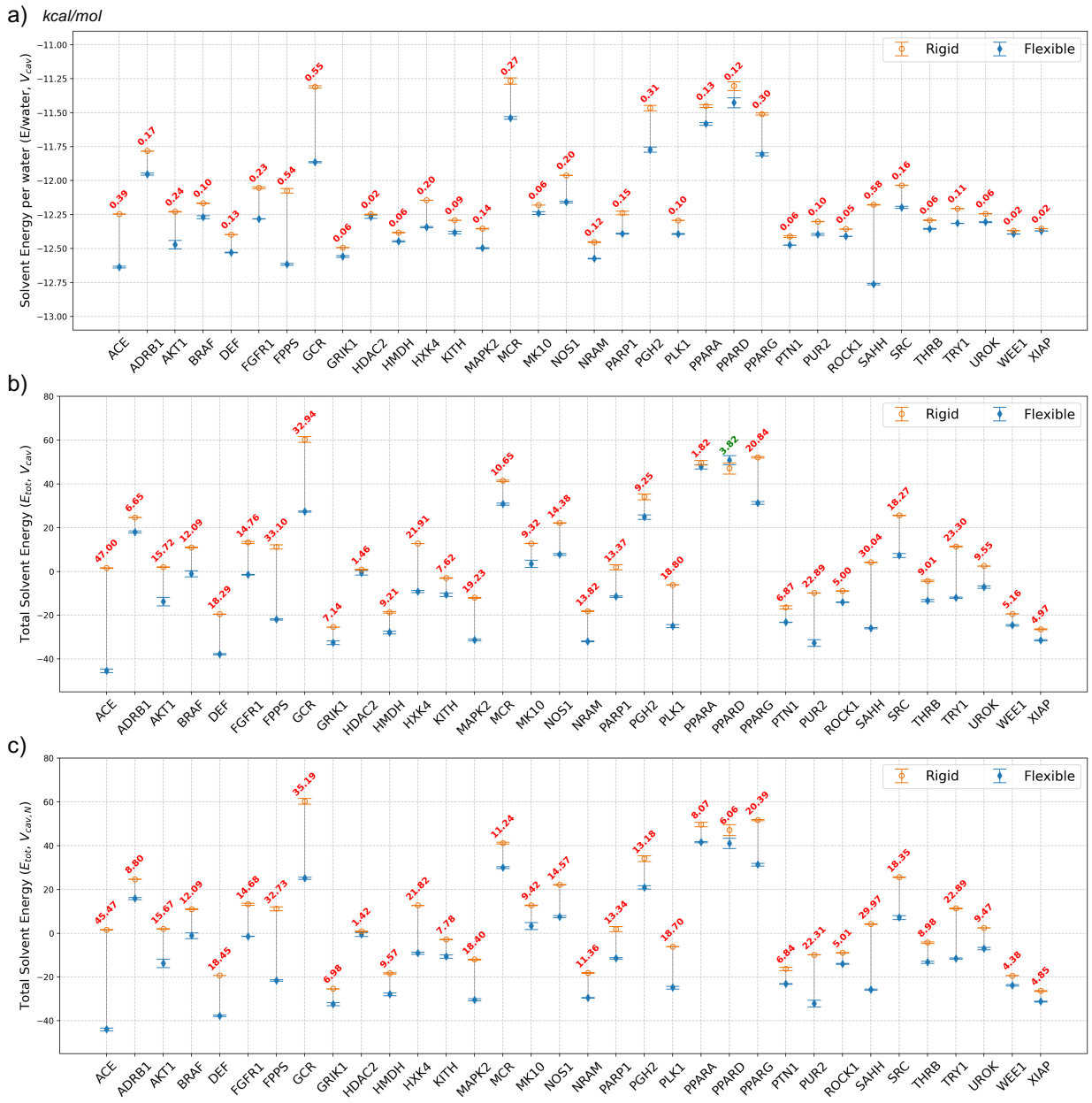

Figure 4. The solvent energy per water molecule (a) and total (b) values for the rigid and flexible binding pockets in V_{cav} (10 Å sub-volume) for 34 systems. Differences between the values of *rigid* and *flexible* pockets for each system are indicated in red text. The total values are referenced to the energy of the same number of neat water molecules. Data for other sub-volumes (3.0 – 9.5 Å) are reported in supplementary information (Figures S3 and S4).

4.3 Cavity Size

In this section, we use the average number of water molecules in the cavity region, V_{cav}, as a proxy for the effective volume of the cavity (Figure 5). Allowing flexible side chains effectively allows the cavities to partially open, allowing more water molecules into the cavity, or partially close, effectively displacing water molecules from the cavity. For the 34 systems studied we observed that the binding cavity of 24 of the systems opened to accommodate more water and 10 of the

systems closed to accommodate fewer water molecules. There was a significant variation in the magnitude of opening or closing and no clear trends emerged with regards to solvation thermodynamic quantities being correlated with the cavities opening or closing. For example, there was no significant correlation between whether a rigid system would open or close (the change in the average number of cavity water molecules) and the energy of the water in the rigid cavity (R^2 =0.21) or the entropy (R^2 =0.02) or the free energy (R^2 =0.10) or the changes in these quantities Energy (R^2 =0.001), Entropy (R^2 =0.01), or Free Energy (R^2 =0.01).

Other than noting that the binding cavities sometimes open and close when the side chains are flexible as opposed to rigid, we were unable to identify a clear predictor of this other than noting that the energy for all the systems becomes more favorable regardless of whether the cavity opens or closes. It is important to highlight this, however, because the number of water molecules in the cavities vary making direct comparisons of the water in the *rigid* and *flexible* difficult due to the fact that there are different numbers of water molecules being compared. This motivates the use of per water quantities which we report when direct comparisons of the thermodynamic quantities are complicated by this.

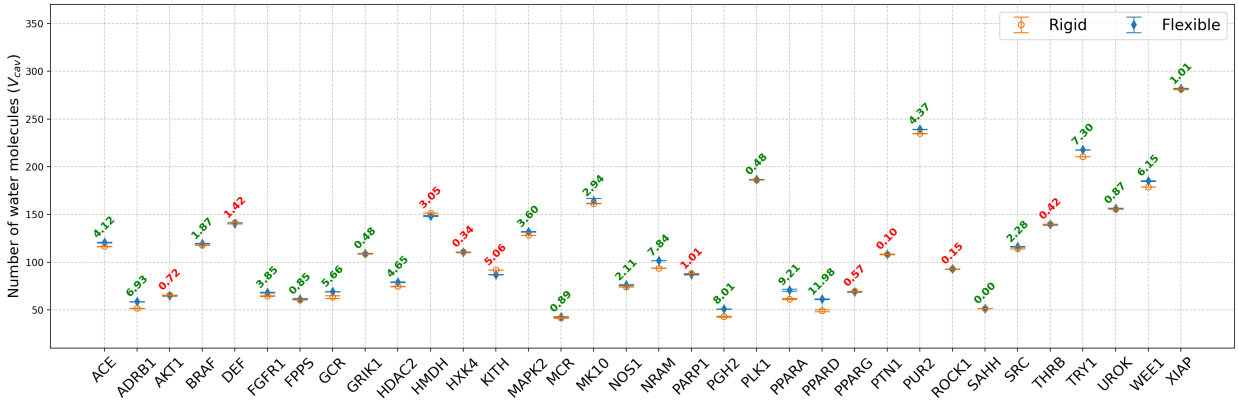

Figure 5. Number of water molecules in V_{cav}, for the rigid and flexible binding pockets for all 34 systems. The text shows the difference between *rigid* and *flexible* with red text denoting that the flexible cavities have fewer water molecules and green text denoting that the rigid cavities have fewer. Data for other sub-volumes (3.0 – 9.5 Å) are reported in supplementary information (Figure S2).

4.4 The protein–water and water–water interaction energy and hydrogen bonds

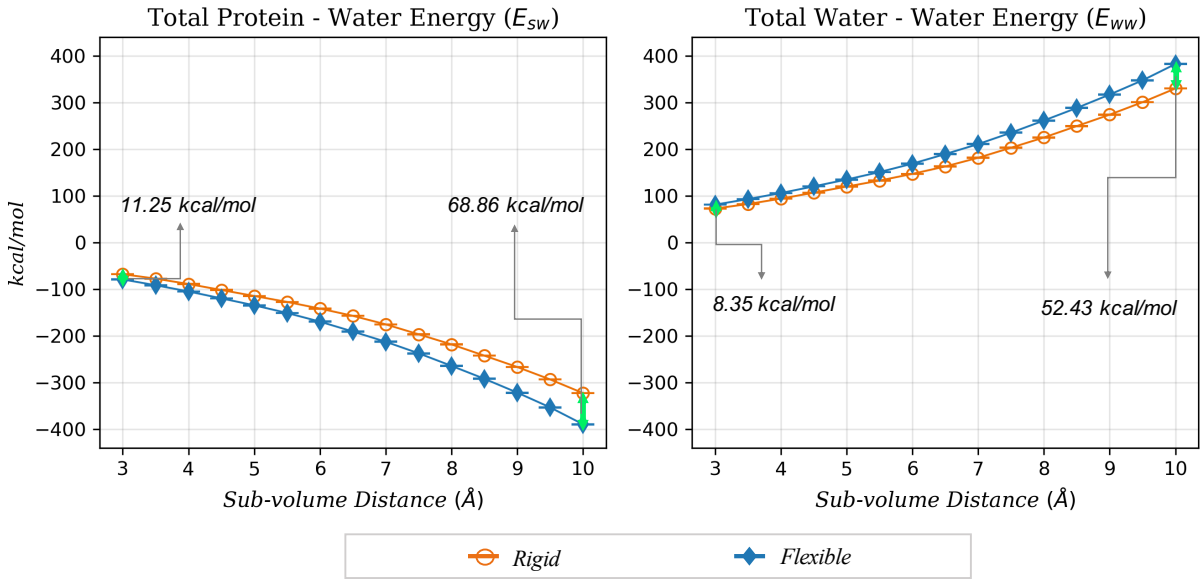

Figure 6. The protein – water and water – water energies per water molecule averaged over 34 systems studied within each sub-volume (x-axis). The water–water energy is referenced to the average water-water energy per water in a neat system.

The total solvation energy can be broken down into a sum of two terms: one from the contribution of water-water interactions (E_{ww}) and another from the protein-water (E_{sw}) interaction. Figure 6 shows that, while the total water energy is more favorable for the *flexible* systems, this total energy difference has two opposing contributions: a protein-water contribution that is more favorable for *flexible* systems and water-water contribution that is less favorable for *flexible* systems. Figure 7 and Figure 8 shows that this trend holds for all 34 systems of the systems investigated in this study with the magnitude of the protein-water differences being larger than the water-water difference for all systems investigated.

The more favorable water-protein energy in the *flexible* systems can be understood in part by the differences in water-protein hydrogen bonds observed in our simulations. Figure 9 shows that on average water forms 8.51 more hydrogen bonds with the protein in V_{cav} of the *flexible* systems than for the *rigid*. Figure 10 shows that in every *flexible* system water forms more hydrogen bonds with the protein than in the *rigid* systems and that the differences in number of protein-water hydrogen bonds ranges from a difference of 2.78 hydrogen bonds for UROK (PDB id: 1SQT[50]) to 18.46 hydrogen bonds for HMDH (PDB id: 3CCW[51]). We also investigated whether the difference in the number of protein-water hydrogen bonds between the *rigid* and *flexible* systems could be related to the differences in numbers of water molecules observed in the binding cavities. We found that regardless of whether there were greater or fewer water molecules in the binding cavity of the *flexible* systems, the water formed more protein-water hydrogen bonds than with the *flexible* systems. Additionally, the difference in the number of water molecules in the binding cavity between the *rigid* and *flexible* systems was uncorrelated with the difference in the number of water-protein hydrogen bonds (R^2=0.00008).

A pattern of water and protein restructuring from the *rigid* to *flexible* systems emerges from this data. In the *rigid* systems, the protein side chains adopt conformations that are suitably

complementary to the ligand, however, water does not make fully complementary h-bonding interactions with the proteins when they are in these conformations. When the restraints on the side chain motions are removed, the protein and water restructures in a manner that lowers the energy of the binding site water and allows the formation of more complementary water-protein interactions as illustrated by the protein-water h-bond data (Figure 9Figure 10). In turn, in the *rigid* systems, water is unable to make as many favorable interactions with the protein and instead following Le Chatelier's principle prioritizes forming more favorable interactions with the neighboring water molecules as illustrated by the lower solvation energy and water-water energy for *rigid* systems (Figure 8). Conversely, while the *flexible* systems restructure to allow more favorable water-protein interactions, this comes at the cost of frustrating the water-water interactions in the *flexible* systems compared to the *rigid* systems.

The frustration of water in *rigid* systems is a topic that we have previously investigated[10] and discussed how water in binding site topographies is often unable to pack into the binding cavity in such a way as to form favorable interactions with both the protein and neighboring water molecules. In section 4.7 below, we give several examples from the systems investigated here as to how water is unable to make complementary interactions with the protein surface in *rigid* systems and illustrate how the side-chains restructure in the *flexible* systems to make more complementary interactions with water.

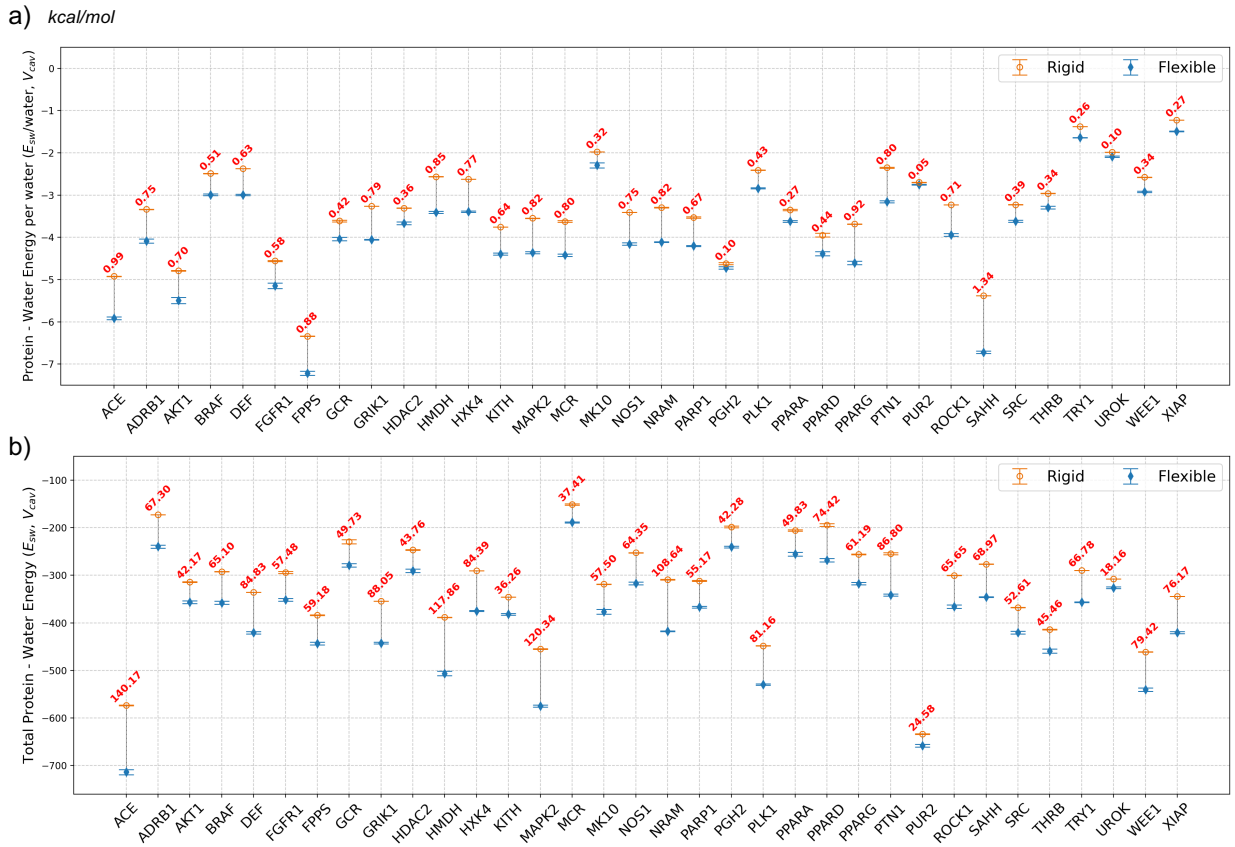

Figure 7. The total protein-water interaction energies per water molecule (a) and total values (b) in the rigid and flexible binding pockets (V_{cav}) for all 34 systems. Differences between the two values are in the text with red text

denoting that the flexible cavities have lower energies. Data for the other sub-volumes (3 – 9.5 Å) can be found in the supplementary information (Figures S5 and S6).

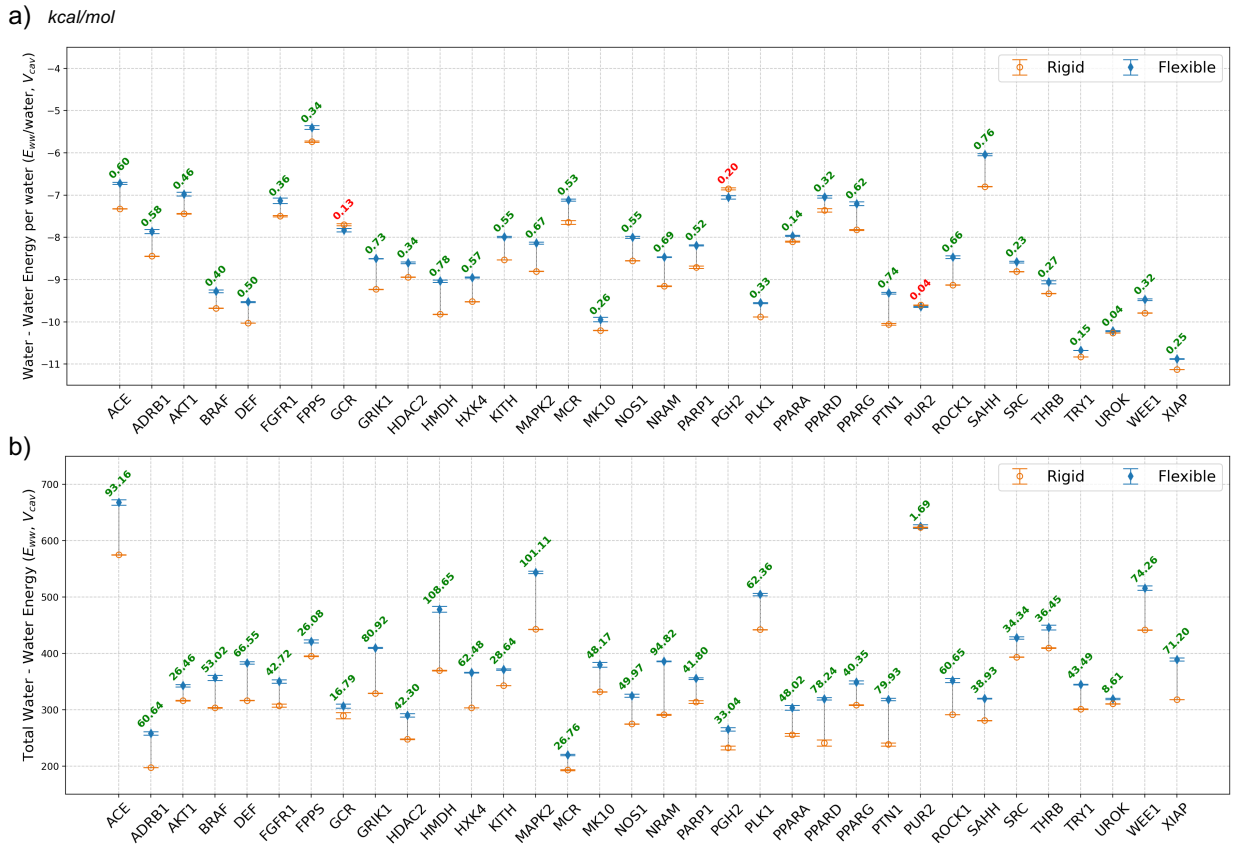

Figure 8. The water-water interaction energy per water molecule (a) and total (b) for the rigid and flexible binding pockets (V_{cav}) for all 34 systems. Differences between the values for the rigid and flexible pockets are in the figure text with green denoting that the water-water energy is more favorable in the rigid cavity and red text denoting that the water-water energy in the flexible cavity is more favorable. Data for other sub-volumes are reported in supplementary information (Figures S7 and S8).

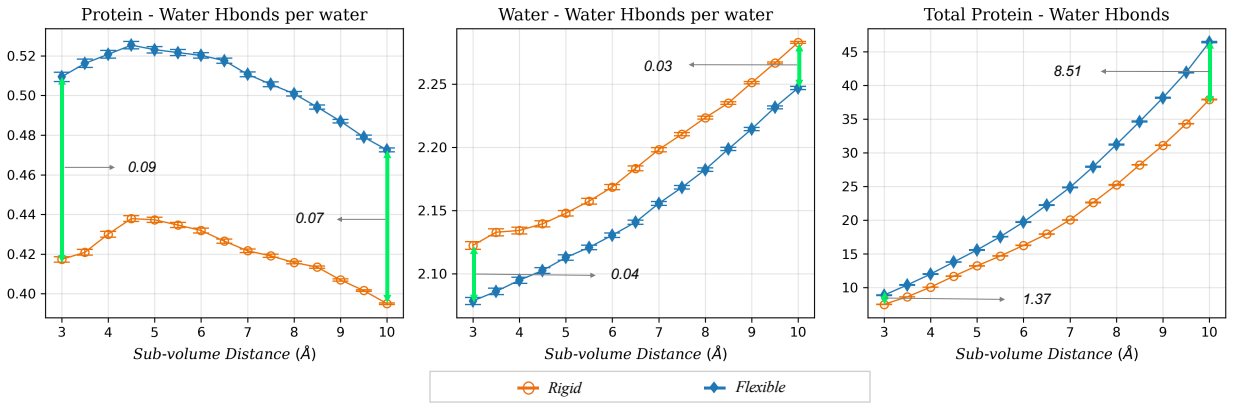

Figure 9. The number of protein-water, water-water, and total hydrogen bonds per water molecule for the rigid and flexible binding cavities in each sub-volume averaged over all 34 systems.

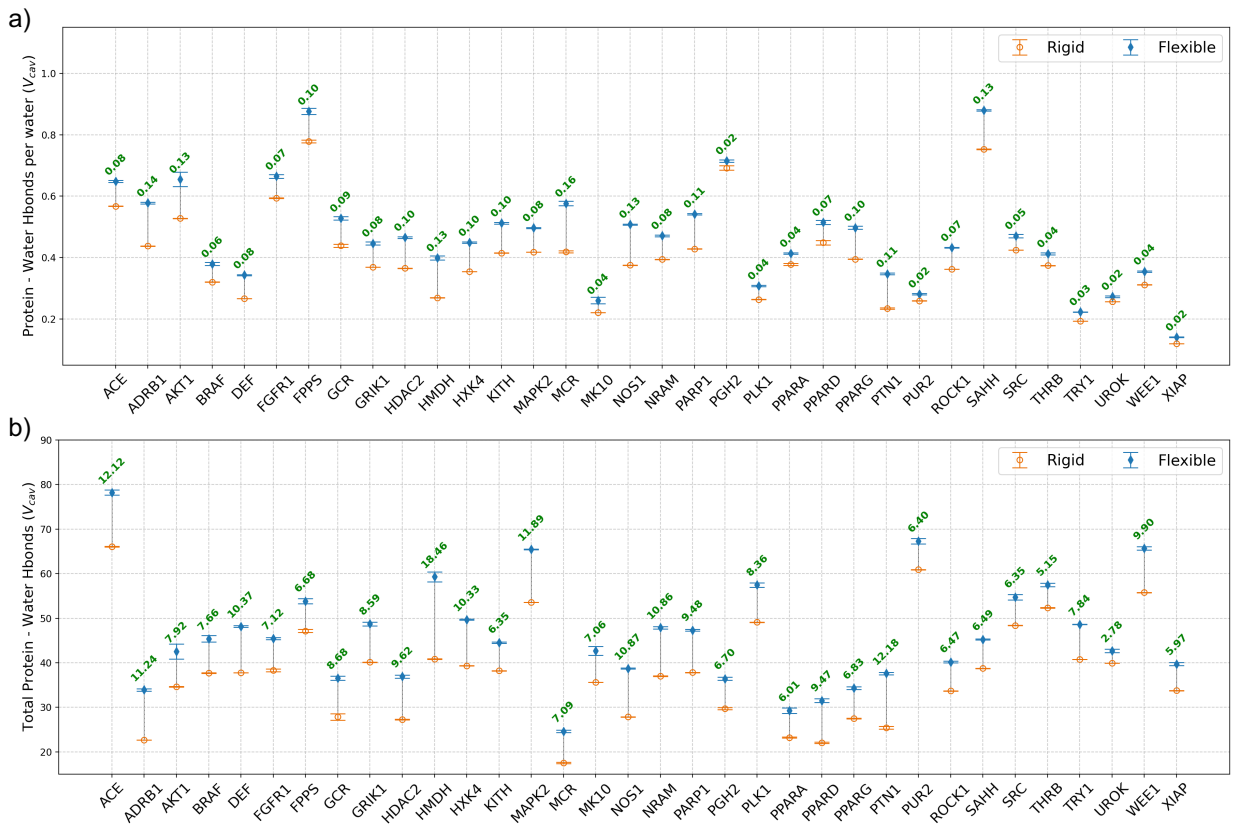

Figure 10. The average number of protein-water hydrogen bonds per water molecule (a) and total (b) values for the *rigid* and *flexible* binding pockets (V_{cav}) for all 34 systems. Differences between the values of *rigid and flexible* pockets for each system are shown in green text when the *flexible* value is higher and red when the *rigid* value is higher. Data for other sub-volumes are reported in supplementary information (Figures S9 and S10).

4.5 Total Entropy

On average the GIST estimated solvation entropy for *flexible* systems is more favorable by 17.96 kcal/mol in the binding cavity V_{cav} (Table 2 and Figure 11). We note that for every system, the entropy per water molecule is more favorable for the *flexible* systems however for four of the 34 systems the total entropy of water (-TS in kcal) in V_{cav} is estimated to be less favorable for four of the 34 systems. This is a result of there being more water molecules in the binding cavities for these systems (Figure 12) with, on average, each water having an unfavorable contribution to the entropy which varies for these four systems from 1.66 kcal/water for PGH2 (PDB id: 3LN1[52]) and 1.48 kcal/water for PPARD. On a per water molecule basis, the water is less structured (higher entropy) by the protein for the *flexible* systems and the entropy is more favorable on a per water basis for all 34 systems investigated. This is consistent with the water having stronger interactions and more hydrogen bonds with the protein surface for the *flexible* simulations. We note that, because of the protein motion in the *flexible* systems, the translational density distributions are smoothed and GIST likely overestimates the entropy for *flexible* systems compared to *rigid.*

Table 2. The average entropy per water and total in Vcav for the 34 systems.

	$-T\Delta S$ / *water*	$-T\Delta S_{total}$
Rigid	1.27	120.13
Flexible	1.07	102.17
Difference (Rigid – Flexible)	0.20	17.96

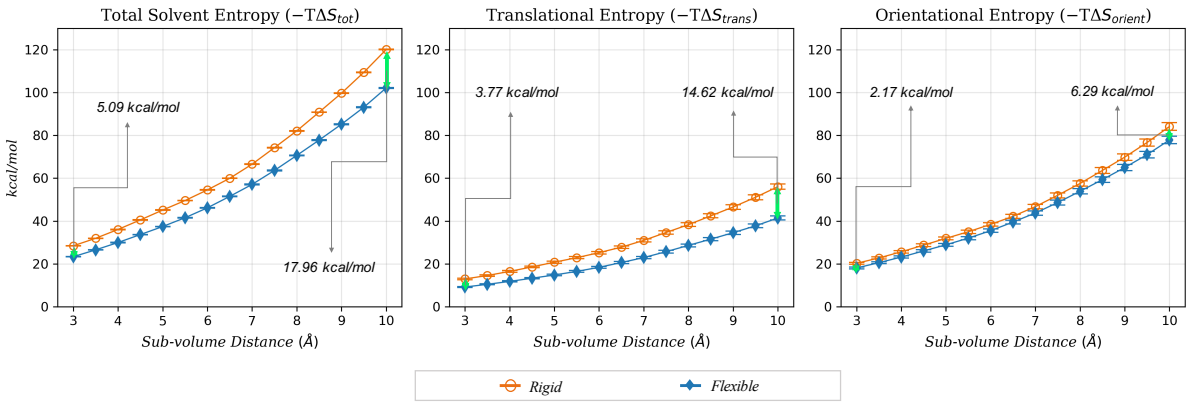

Figure 11. The total, translational and orientational entropy averaged over all 34 systems in each sub-volume.

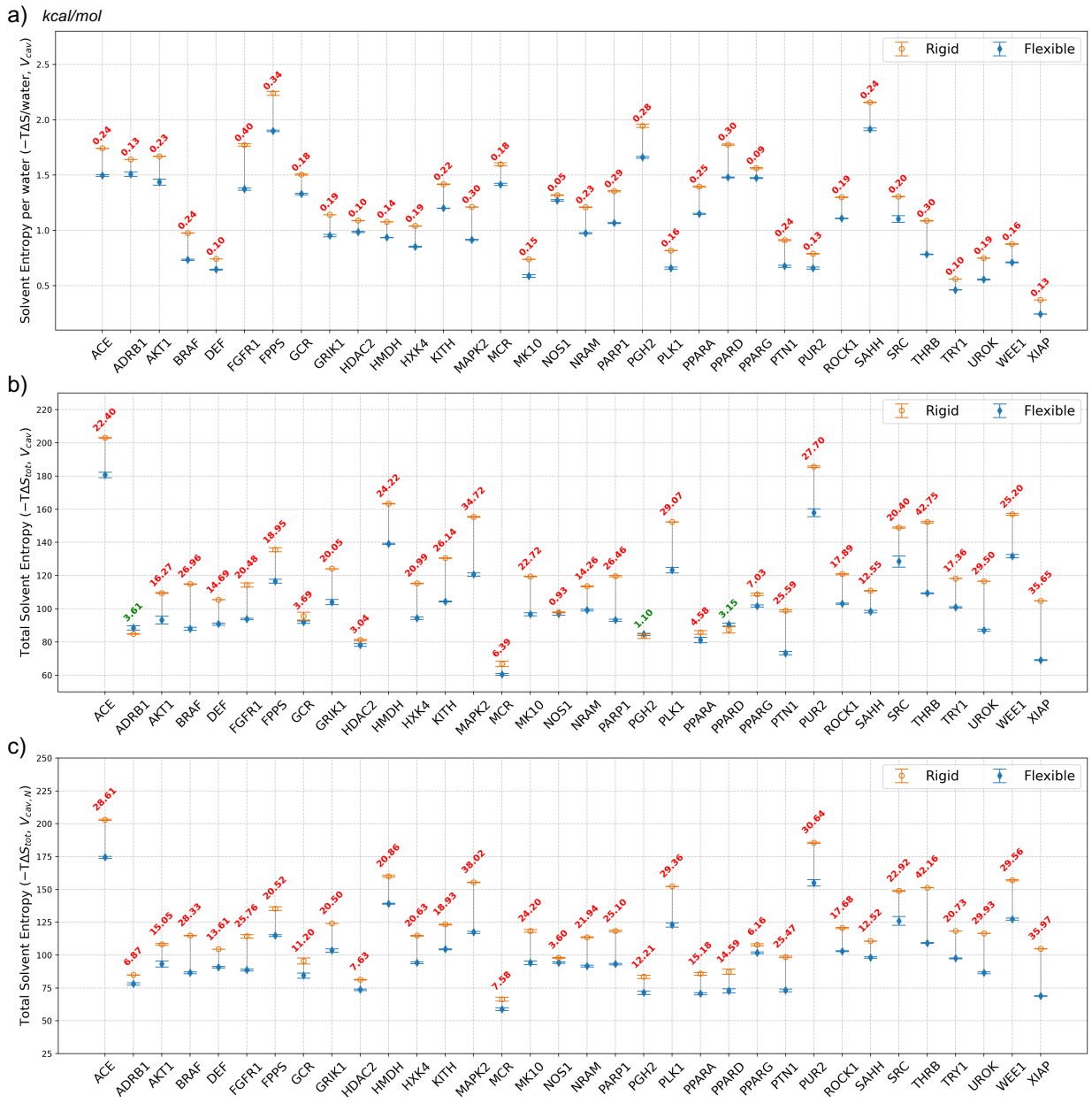

Figure 12. The solvent entropy a) per water molecule, b) total values in the binding cavity, and c) total value in the cavity for the same number of water molecules. Differences between the values of rigid and flexible pockets for each system are in red text for when the value for the flexible system is more favorable, and green when the value for the rigid system is more favorable. Data for other sub-volumes (3.0 – 9.5 Å) are reported in supplementary information (Figures S12 and S13).

4.6 Helmholtz free energy

On average the GIST estimated Helmholtz solvation free energy of the *flexible* systems are more favorable by 32.39 kcal/mol within V_{cav} (Figure 13). On a per system basis, the free energy is lower for the *flexible* simulations for 31 of the 34 systems investigated however, the three systems that have this estimate have more water molecules in V_{cav} in the *flexible* simulations and have total

entropies in V_{cav} that are estimated using different numbers of molecules as noted in Section 4.5. On a per molecule basis, however, the Helmholtz free energy is more favorable in the binding cavity of the *flexible* systems for every system studied (Figure 13).

The magnitude of difference between the free energies of solvation between the *rigid* and *flexible* systems varies significantly when comparing between proteins with values varying from 6.96 kcal/mol more favorable for the *rigid* PPAR-δ (PDB id: 2ZNP[48]) to 69.40 kcal/mol more favorable for the *flexible* ACE (PDB id: 3BKL[49]). Data for all systems is shown in Figure 13. We note that if we compare the same number of water molecules in the *rigid* and *flexible* systems then the free energies are more favorable for the *flexible* cavity for all 34 systems investigated.

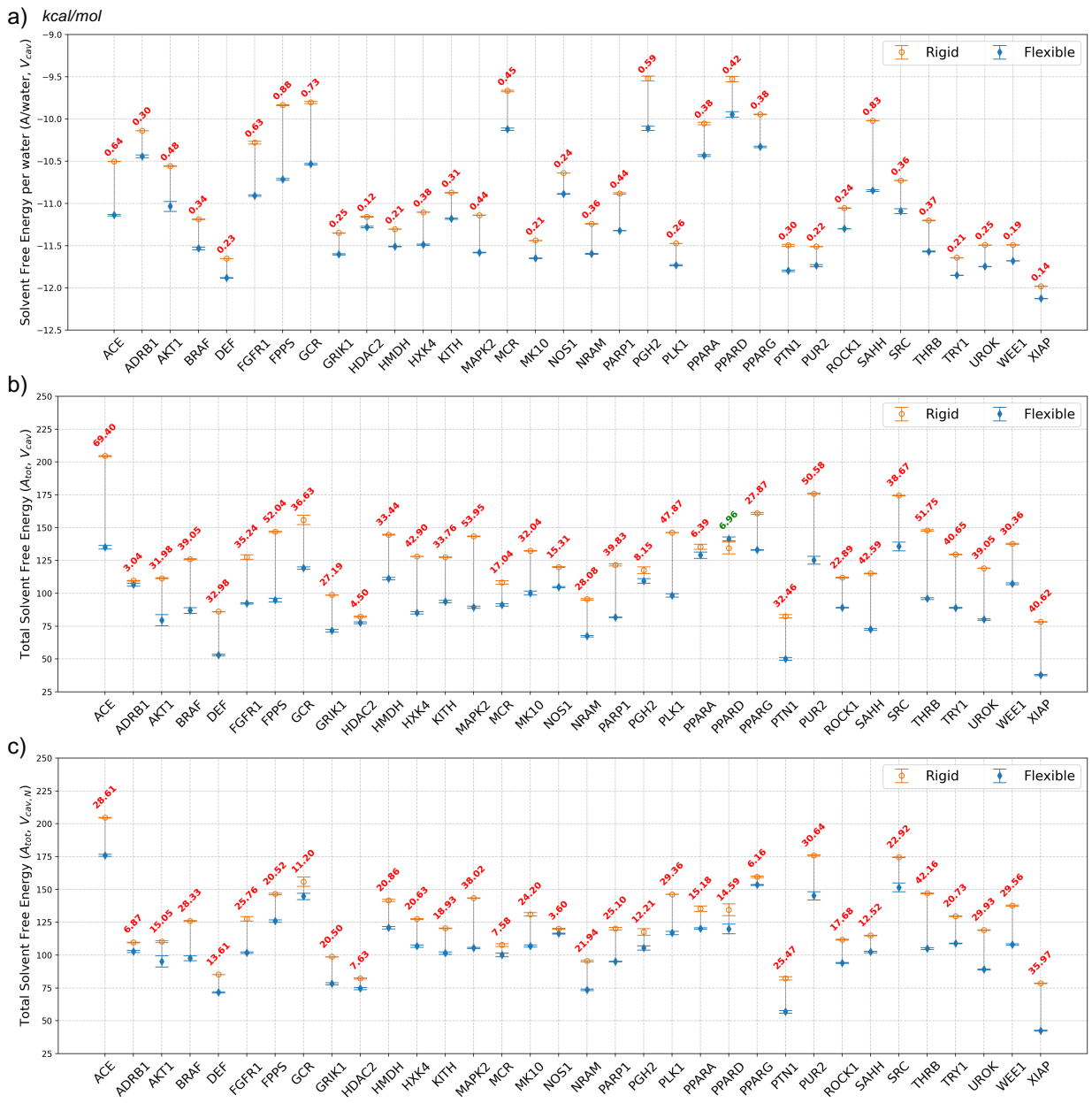

Figure 13. The Helmholtz solvent free energy a) per water molecule, b) total values in the binding cavity, and c) total value in the cavity for the same number of water molecules. Differences between the values of rigid and flexible pockets for each system are in red text for when the value for the flexible system is more favorable, and green when the value for the rigid system is more favorable. Data for other sub-volumes (3.0 – 9.5 Å) are reported in supplementary information (Figures S14 and S15).

4.7 Systems bound to multi-dentate cognate ligands

In a previous study, we investigated the structural frustration of water molecules solvating *rigid* binding sites.[10] Inspired by Frank & Evans iceberg model of hydrophobic hydration,[53] we characterized *optimal* hydration as water molecules' ability to pack into a binding cavity in such a way as to make favorable h-bonding interactions with both the protein surface and water neighbors.

Conversely, we characterized *sub-optimal* hydration as situations in which water is unable to simultaneously form favorable h-bonds with both its neighboring water molecules and the protein surface.

In particular, we investigated the water molecules solvating the residues that form hydrogen bonds with aspartate in the *rigid* binding site of the aspartyl specific protease Caspase-3, a highly enclosed binding pocket. In this cavity, the structure of the water was *frustrated* in that the water in the cavity was unable to form the same number of hydrogen bonds with the surface as the carboxylate of the ligand and the water-water interactions were unfavorable compared to those formed in bulk water. Our physical interpretation behind this had to do with how water molecules pack in the liquid phase. In neat water, the oxygens of two water molecules do not approach a distance closer than 2.4 Å as the radial distribution function is effectively zero until this distance and the optimal (most probable) distance between two water molecules is 2.8 Å. Many functional groups in pharmaceutical compounds, including the aspartate in this prior study, have donor/acceptor heavy atom pairs that are less than 2.4 Å from each other and water due to its inability to pack so close together is unable to place the donors and acceptors at the same positions as the ligands in these cases.

Here, we describe in greater detail the binding site solvation of several systems (PPAR-γ, HMDH, and WEE1) proximal to the region where the bi-dentate ligands form hydrogen bond contacts with the protein surface. For each functional group of the protein that forms hydrogen bonds with the bidentate ligand, we detail metrics of solvation including number of water neighbors (a metric of solvent exposure) and how many hydrogen bonds they are able to make with the solvent and how these values differ in the *rigid* and *flexible* simulations. Our analysis focuses on PPAR-γ (Peroxisome proliferator-activated receptor gamma, PDB id: 2GTK[54]), HMDH (HMG-CoA reductase, PDB id: 3CCW), and WEE1 (Serine/threonine-protein kinase WEE1, PDB id: 3BIZ[55]) each of which has a bi-dentate ligand but have binding site topographies that differ considerably.

4.7.1 System 1: **PPAR-γ (PDB id: 2GTK)**

The binding site of PPAR-γ tightly encloses ligand 208[a]'s carboxylate which accepts four total hydrogen bonds with protein residues TYR-473, HIS-448 (Figure 14, right oxygen) with SER-289, and HIS-323 (Figure 14, left). While the binding cavity in both the *rigid* and *flexible* simulations are large enough to accommodate a carboxylate (Figure 14, middle and right), neither is large enough to accommodate two water molecules and, therefore, when these cavities are solvated only a single water molecule is placed that can accept the hydrogen bonds that the ligand carboxylate makes. This is due to the limited ability of water to pack into tightly enclosed areas. The distance between the two carboxylate oxygens is 2.3 Å which is closer than two water oxygens approach due to their van der Waals radii (e.g. the radial distribution of water is effectively zero at this distance) and the binding cavity in neither the *flexible* nor *rigid* simulations is large enough to accommodate two water molecules.

In both the *rigid* and *flexible* systems, a hydration site is placed to form hydrogen bonds with TYR-473 and HIS-449 and form comparable contacts with the protein that the ligand carboxylate makes (Figure 14, middle and right). This leaves the h-bond contacts with SER-289 and HIS-323 unsatisfied in the *rigid* structure when the ligand is not present. In the *rigid* simulations, the hydroxy of SER-289 and HIS-323 are poorly solvated with few water neighbors and forming almost no hydrogen bonds with water (Table 3Table 4).

[a] (2s)-3-(1-{[2-(2-chlorophenyl)-5-methyl-1,3-oxazol-4-yl]methyl}-1h-indol-5-yl)-2-ethoxypropanoic acid

In the *flexible* simulations, the protein, correspondingly, restructures to a configuration that is more *complementary* to the water. In the *flexible* simulations, the imidazole group flips to form an internal hydrogen bond with LYS-319 (not shown) and is no longer solvent accessible with no water neighbors and correspondingly no hydrogen bond contacts with water (Table 3Table 4). The hydroxy of SER-289 moves to become more solvent accessible (Table 3Table 4) and thus forms better hydrogen bonds with water. Overall, the Serine hydroxy is significantly better hydrated in the *flexible* system with more neighbors (1.4) and more hydrogen bonds (1.14) than in the *rigid* simulation (Table 3Table 4).

The reorganization of the protein in the *flexible* simulations has repercussions with regards to identifying ligands that are complementary to the protein surface. In the cognate structure, the ε-amino of HIS-323 is on the surface and can hydrogen bond with a complementary ligand whereas in the predominant *flexible* structure, the protein surface is neutral. In both the cognate structure and the predominant *flexible* structure of the hydroxy of SER-289 are solvent accessible however in the flexible system the hydrogen bonding position has moved by 2.5 Å. We note that in the *flexible* systems, the side chains can move back and forth. In the *flexible* simulation, the hydroxy of the SER-289 was proximal to the shown position (Figure 14, right) in 93.6% of the sampled frames and closer to the cognate structure in 6.4% of the sampled frames.

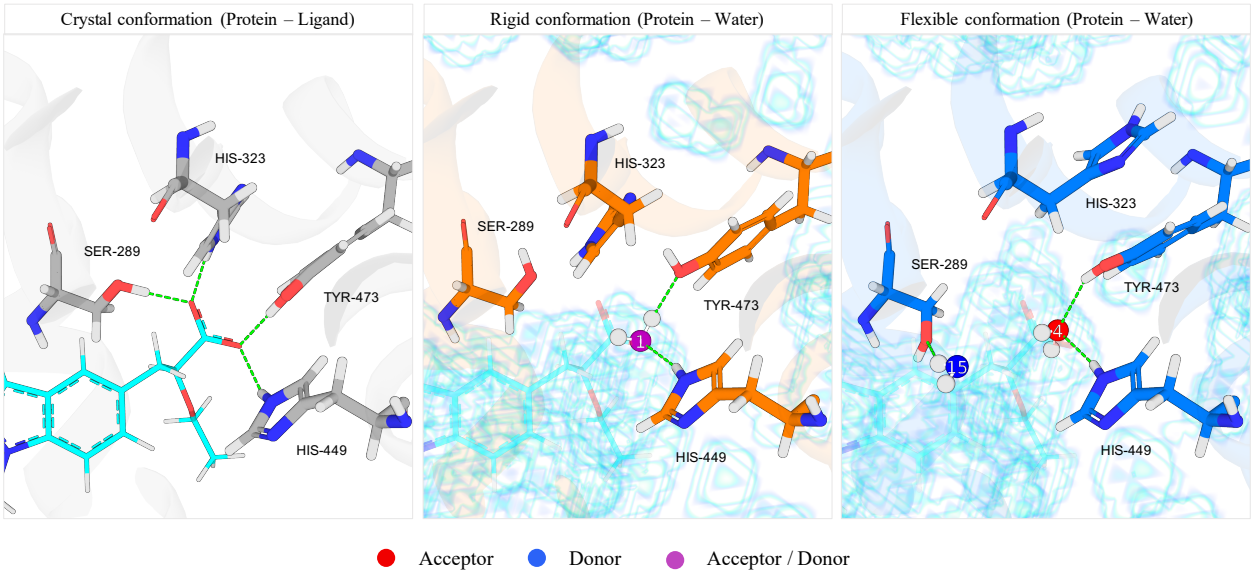

Figure 14. The binding cavity of PPAR-γ (PDB id: 2GTK). (Left) – Ligand 208 and the hydrogen bonds that its carboxylate forms with the protein, (Middle) – The solvent accessible surface (scaffolded) and SSTMap hydration site that interacts comparably to the ligand carboxylate in the rigid simulation. (Right) – The same results for the flexible simulation.

Table 3. The average number of water molecule neighbors (within 3.5 Å) of the functional groups of side chains that formed hydrogen bonds with the ligand in the cognate conformation of PPAR-γ (PDB id: 2GTK). This is a metric of how well solvated each functional group is.

	SER289-OG	HIS323-NE2	TYR473-OH	HIS449-NE2	**Total H_2O Neighbors**
Rigid	0.30	0.43	0.88	1.11	2.73

Flexible	1.73	0.00	2.18	1.30	5.21

Table 4. The average number of hydrogen bonds formed between water and the functional groups of the side chains that formed hydrogen bonds with the ligand in the cognate conformation of PPAR-γ (PDB id: 2GTK).

	SER289-OG	HIS323-NE2	TYR473-OH	HIS449-NE2	**Total H-bonds**
Rigid	0.09	0.00	0.73	0.80	1.62
Flexible	1.23	0	1.78	0.83	3.84

As water structures itself differently in the *rigid* and *flexible* binding cavities, it is often difficult to compare hydration sites between them as they often change positions. However, hydration site 1 (HS1) in the *rigid* system and HS4 in the *flexible* system both bridge hydrogen bonds between the same functional groups of TYR-473 and HIS-449 which makes the comparison more reasonable. The data for these two hydration sites is in Table 9. The water in the *flexible* system has slightly more water neighbors and forms on average slightly more h-bonds (0.28) with its water neighbors. In turn, it makes slightly fewer h-bonds with the protein and overall makes slightly less favorable energetic interactions with the protein but the significantly better energetic interactions with its water neighbors compensate for this and the energy and free energy of the hydration site are more favorable in the flexible system (Table 5).

Table 5. The protein–water interaction energy (E_{sw}), water–water interaction energy (E_{ww}), total energy (E_{tot}), total entropy (TS_{tot}), and total Helmholtz free energy (A_{tot}) of hydration site 1 (*rigid*) and hydration site 4 (*flexible*).

	E_{sw}	E_{ww}	$\boldsymbol{E_{tot}}$	$\boldsymbol{TS_{tot}}$	$\boldsymbol{A_{tot}}$
HS1 (*Rigid*)	-5.82677	-4.83525	-10.662	-3.81948	-14.48148
HS4 (*Flexible*)	-5.5404	-6.08115	-11.6216	-4.01545	-15.63705

4.7.2 System3: **WEE1 (PDB id: 3BIZ)**

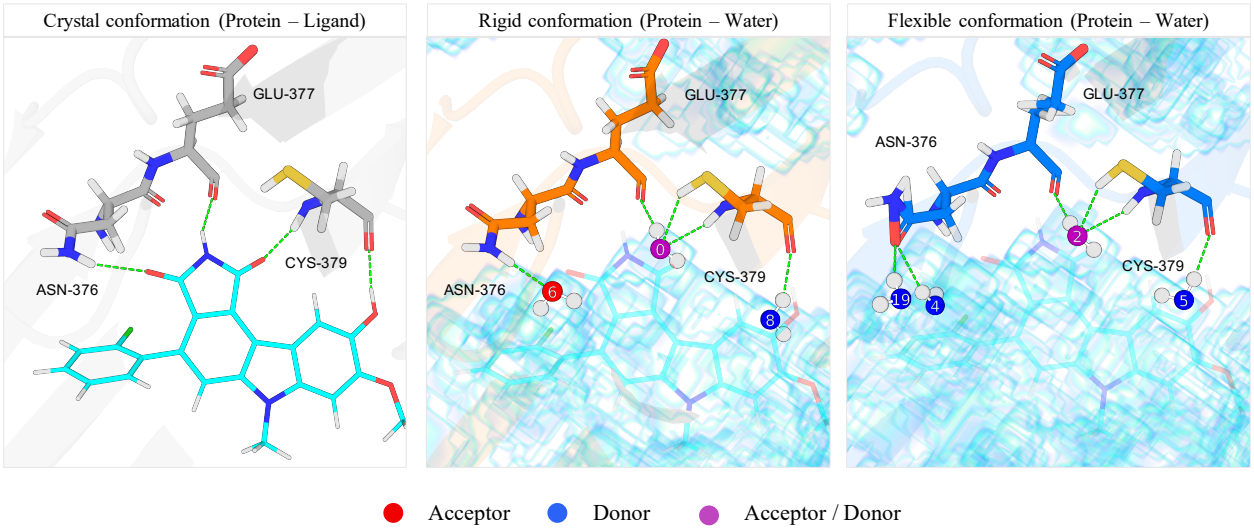

Figure 15. The binding cavity of WEE1. (Left) – Ligand 61E and the hydrogen bonds that it forms with the protein, (Middle) – The solvent accessible surface (scaffolded) and SSTMap hydration sites that interact comparably to the ligand in the rigid simulation. (Right) The same results for the *flexible* simulation with the time averaged structure of the protein.

Figure 15 (left) shows the complex of ligand 61E[b] with WEE1. In contrast to PPAR-γ, the binding site does not tightly enclose the tri-dentate succinimide group and water is fully able to make all three hydrogen bonds with the protein in both the *rigid* and *flexible* simulations. While the water molecules cannot pack into the cavity to make these contacts from the same positions as the ligand, they are able to spread themselves out within the open cavity to make complementary hydrogen bond contacts with the protein. The distance between the nitrogen and oxygens in succinimide are 2.3 Å, a distance that is energetically prohibitive for two water molecules to have. In contrast, the water molecules that make the same contacts as the NH and carbonyl groups are 4.3 Å (HS0 and HS6) in the *rigid* system and 6.6 and 6.9 Å in the *flexible* system (HS2 with HS4 and HS19 respectively). Despite this, there is a reorganization of the protein in the *flexible* simulations such that the amide group of ASN-376 flips (88.1% of the simulation) so that the carbonyl is solvent exposed instead of the amino group (Table 6). This amide flipping also suggests that ligands that can donate or accept in that position will be complementary to the protein. The solvent exposure of the other protein functional groups that interact with the ligand remain relatively unchanged when comparing the *rigid* and *flexible* systems as does the number of hydrogen bonds (Table 7). Overall, the solvent reorganization energy of WEE1 is small compared to PPAR-γ with the *flexible* systems being more favorable 5.16 kcal/mol ($V_{cav,N}$) compared to 20.84 kcal/mol for PPAR-γ.

Table 6. The average number of water molecule neighbors (within 3.5 Å) of the functional groups of side chains that formed hydrogen bonds with the ligand in the cognate conformation of WEE1. This is a metric of how well solvated each functional group is in the *rigid* and *flexible* simulations.

	ASN376-ND2	ASN376-OD1	GLU377-O	CYS379-SG	CYS379-O	CYS379-N	**Total H_2O Neighbors**

[b] 4-(2-chlorophenyl)-8-[3-(dimethylamino)propoxy]-9-hydroxy-6-methylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione

Rigid	1.92	0	1.00	0.57	1.09	1.19	5.77
Flexible	0.19	1.55	1.01	0.67	1.04	1.04	5.50

Table 7. The average number of hydrogen bonds formed between water and the functional groups of the side chains that formed hydrogen bonds with the ligand in the cognate conformation of HMDH.

	ASN376-ND2	ASN376-OD1	GLU377-O	CYS379-SG	CYS379-O	CYS379-N	**Total H-bonds**
Rigid	0.92	0	0.96	0.45	0.97	0.93	4.23
Flexible	0.10	1.22	0.98	0.58	0.97	0.86	4.71

4.7.3 System 2: **HMDH (PDB id: 3CCW)**

In the prior sections, we outlined that there is a significant thermodynamic driving force for the protein to adopt conformations that are complementary to the water. Here, using HMDH as an example, we detail how, in the *flexible* simulations, the side chains proximal to the ligand-protein hydrogen bonds restructure when the ligand is not present. The left panel of Figure 16 shows the complex of ligand 4HI[c] with HMDH. The binding site encloses the ligand carboxylate and hydroxy group which form three hydrogen bonds with the protein. The middle and right panels show the solvent accessible volume in the binding site (blue scaffold) for the *rigid* and *flexible* simulations and hydration sites that hydrogen bond with the same functional groups as the ligand.

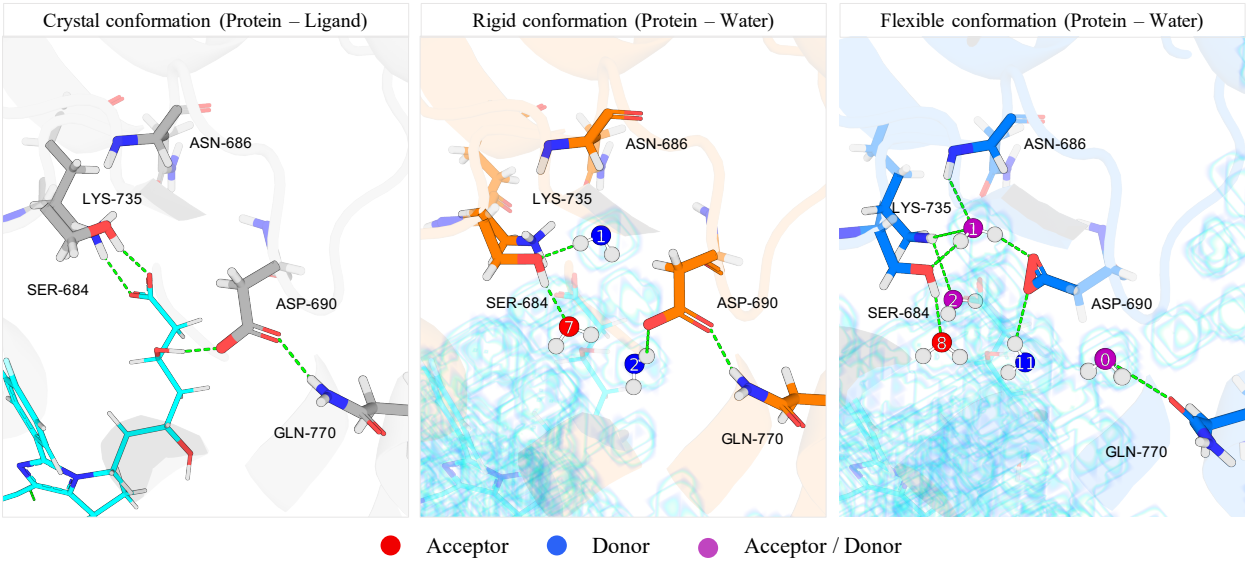

Figure 16. The binding cavity of HMDH. (Left) – Ligand 4HI and the hydrogen bonds that it forms with the protein, (Middle) – The solvent accessible surface (scaffolded) and SSTMap hydration sites that interact comparably to the

[c] (3R,5R)-7-[4-(benzylcarbamoyl)-2-(4-fluorophenyl)-5-(1-methylethyl)-1H-imidazol-1-yl]-3,5-dihydroxyheptanoic acid

ligand in the *rigid* simulation. (Right) The same results for the *flexible* simulation with the time averaged structure of the protein.

The protein functional groups that are hydrogen bonded with ligand 4HI are significantly better hydrated in the *flexible* simulations than in the *rigid* forming approximate 2.5 more hydrogen bonds with water (Table 8) and having more than 2 additional water neighbors (Table 9).

In order to make these hydrogen bonds, the carboxylate of ASP-690 rotates to interact with HS1 in the *flexible* simulations (Figure 16). This makes HS1, which is energetically unfavorable compared to bulk in the *rigid* simulations, energetically *favorable* in the *flexible* simulations lowering the energy by more than 2 kcal/mol. In the *rigid* system, HS1 is frustrated and only forms 1.34 hydrogen bonds with its surroundings and in the *flexible* system forms a full complement (3.52) hydrogen bonds with its surroundings (Table 10). Importantly, the carboxylate of ASP-690 forms a hydrogen bond with the ligand in the cognate structure, however, when the protein side chains are free to move in the MD simulations, the carboxylate moves to a position that is too distal to form a hydrogen bond with the ligand. We investigated this and found in our production trajectories what proportion of the frames were in the up position (Figure 16, right) and in the down position (Figure 16, middle). In all 20,000 simulation frames, the carboxylate was in the up position. In the *rigid* simulations and cognate structure, the amino group of GLN-770 forms a hydrogen bond with the carboxylate of ASP-690. When this carboxylate repositions to interact with HS1 in the *flexible* simulations, it is no longer available to accept a hydrogen bond from GLN-770 and the rotamer state of GLN-770 flips in 99.95% of the simulation frames (10 out of 20,000 frames were in a position comparable to the cognate structure). In the *rigid* structure, the side chain of GLN-770 is not available to form h-bonds in the binding cavity. Its amino group has no water neighbors, is not exposed to the binding pocket, and the carbonyl is flipped away from the pocket. In the *rigid* structure, however, the carbonyl flips to be exposed to the binding pocket and forms a hydrogen bond with HS0 (Table 11 and Table S2).

These structural rearrangements are important when considering the complementarity of potential ligands to the protein surface. The *rigid* simulations were restrained from moving significantly from the cognate structure of the protein which is complementary to the ligand. However, when the system is flexible, the position of the h-bond acceptors of ASP-690 have moved and a new hydrogen bonding acceptor from GLN-770 is revealed in the binding pocket. This makes the cognate ligand not complementary to the binding site generated in the *flexible* simulations. Importantly, the configurations of the protein that are complementary to the cognate ligand are never sampled in the 80 ns MD simulations however there may be ligands that are complementary to the structure revealed in the flexible simulations.

Table 8. The average number of hydrogen bonds formed between water and the functional groups of the side chains that formed hydrogen bonds with the ligand in the cognate conformation of HMDH (PDB id: 3CCW).

	LYS735-NZ	SER684-OG	ASP690-OD1,2	**Total H-bonds**
Rigid	0.29	1.88	1.11	3.28
Flexible	0.92	1.84	3.08	5.84

Table 9. The average number of water molecule neighbors (within 3.5 Å) of the functional groups of side chains that formed hydrogen bonds with the ligand in the cognate conformation of HMDH (PDB id: 3CCW). This is a metric of how well solvated each functional group is.

	LYS735-NZ	SER684-OG	ASP690-OD1,2	**Total H_2O Neighbors**
Rigid	1.04	2.88	1.75	5.67
Flexible	1.90	2.72	4.13	8.75

Table 10. The protein–water interaction energy (E_{sw}), water–water interaction energy (E_{ww}), total energy (E_{tot}), and the average number of protein – water hydrogen bonds of hydration site 1 in each system.

	E_{sw}	E_{ww}	$\boldsymbol{E_{tot}}$	**Protein – Water H-bonds**
Rigid	-5.45	-5.40	-10.85	1.34
Flexible	-11.69	-1.23	-12.92	3.52

Table 11. The average number of water molecule neighbors (within 3.5 Å) and hydrogen bonds formed between water and GLN770 of HMDH (PDB id: 3CCW).

	H_2O Neighbors			**H-bonds**		
	GLN770-NE2	GLN770-OE1	**Total**	GLN770-NE2	GLN770-OE1	**Total**
Rigid	0.00	1.00	1.00	0.00	0.98	0.98
Flexible	1.00	2.22	3.22	0.91	1.80	2.72

5 Discussion

The concept of ligand-protein complementarity is a fundamental premise of structure-based drug discovery and essentially states that tightly binding ligands make complementary contacts with the protein: donating or accepting hydrogen bonds and making hydrophobic contacts where appropriate. Correspondingly, the concept of induced fit[56] suggests that protein binding sites conform their shapes to the ligands in response to the presence of the ligand. Here, we'd like to discuss the concept of *protein-water complementarity*. Our results suggest that when a protein is in its ligand bound conformations, it cannot make optimal hydrogen bond interactions with water and the solvation thermodynamics provides a considerable driving force to restructure the protein. In every simulated system, when the *rigid* restraints on the side chains were absent, the protein binding sites restructured to form a greater number of hydrogen bonds with water, have more favorable protein-water interactions, and have a lower overall free energy of solvation - effectively adopting binding site conformations that are *complementary* to the water molecules in the binding site.

When a protein is unbound, these results suggest that there is a significant solvation free energetic cost to forming binding site conformations that are *complementary* to small molecule chemical matter. The average cost is substantial and was estimated here to be approximately 35.69 kcal/mol with the estimations ranging from system to system with a low of 3.60 kcal/mol for NOS1 (PDB id: 1QW6[57]) and a high of 42.16 kcal/mol for THRB (PDB id: 1YPE[58]) (Figure 13). Consistent with the idea of induced fit, this cost may prevent the apo-protein from sampling conformations that are complementary to tightly binding ligands. In the simulations of HMDH and PPAR-γ, the un-liganded simulations rarely or never sampled the cognate conformations that were complementary to the bound ligands for PDB ids 3CCW and 2GTK, respectively.

We believe the results have significant implications in both the search for cryptic pockets and with regards to the proper use of solvation mapping methods in molecular recognition. We will briefly discuss each of these in the following subsections:

Implications for Cryptic Pocket Identification.

Due to the inherent flexibility of proteins binding cavities change shape and the positions of hydrogen bond and hydrophobic interaction sites vary with these structural fluctuations. This is exciting for drug discovery programs as differently shaped cavities will correspondingly bind different chemical matter with potential improvements to ADMET properties, selectivity, and patentability over known compounds. For these reasons, significant efforts, both computational and experimental, have been dedicated to identifying alternate druggable conformations of binding pockets than those already known from experimental structures.

Experimentally, studies using methods such as NMR[59–61], Cryo-EM[62–65] and room temperature crystallography[66–68] have focused on revealing alternate conformations of proteins and identifying such "cryptic" pockets. Computationally methods have focused on simulation sampling based approaches that improve the exploration of protein conformational free energy surfaces using enhanced sampling methods such as metadynamics[69–72], umbrella sampling[73–75] and replica exchange molecular dynamics (REMD)[76,77] or machine learning approaches such as CryptoSite[78] and incarnations of AlphaFold[79–82].

The results here describe a significant cost for un-ligated proteins to adopt conformations that are complementary to potential or known chemical matter. The magnitude of this cost suggests that there may be little to no overlap between the binding site configurational integrals of un-ligated proteins with those of ligated proteins. In two of the three systems that bind multi-dentate ligands (HMDH and PPAR-γ), we found that the conformations that are complementary to the cognate chemical matter were rarely if ever sampled in unbiased molecular dynamics simulations. Other computational approaches aimed at sampling the conformational space of un-ligated binding pockets may also rarely if ever sample conformations that can tightly bind chemical matter. Experimental methods suffer from the same problems. Un-ligated proteins will rarely be in conformations that are complementary to chemical matter and will therefore have a minimal, likely unresolvable, signal in NMR, X-Ray, or Cryo-EM experiments.

This suggests that computational techniques that explore conformations with chemical matter present (such as mixed solvent MD or high throughput fragment based experimental methods) may have a better chance of finding relevant binding pocket conformations than methods that explore the free energy landscape of un-ligated proteins. This is likely also true for computational methods developed to improve sampling in molecular dynamics simulations. These methods are generally designed to overcome free energy barriers and better sample regions about minima on free energy

surfaces. What is described here, however is a cost to generating these conformations and, on the free energy landscape, these conformations are not in the region of these minima.

We note that the lack of conformational overlap may only apply to a subset of proteins. In WEE1, which is characterized as a relatively unenclosed binding site, there seemed to be significant overlap in the apo and ligand bound conformations and conformations that bound the cognate ligand were explored.

Implications for the Use of Solvation Mapping methods in Drug Discovery

We also believe that the results presented here have an impact on how solvation mapping methods are best used. Most methods aimed at computationally estimating the thermodynamics of solvating binding sites rely upon an analysis of the solvation of *rigid* cognate ligand-bound protein structures for which the ligand has been removed.[4–6,9,11,12,14–20,41,83,84]

In particular, solvation mapping methods using IST have been widely used to estimate the solvent contribution to binding of ligands or the differences in binding affinity of congeneric pairs of ligands. Many of these applications have relied upon the supposition that the main contribution to binding is due to the displacement of water from the binding site and that the solvent contribution can be estimated by the difference in thermodynamic properties of displaced water in the binding site and in neat water to which the water has presumably been displaced.[4,41,45,85,86]. The maps used in this solvent displacement approximation are generally generated from simulations of *rigid* cavities. As the solvation of *rigid* binding cavities is significantly less favorable than the solvation of *flexible* cavities, the results here suggest that this yields a significant overestimation to the contribution of binding. Along with the thermodynamics, the structure of water often differs considerably between *flexible* and *rigid* cavities. Approaches that estimate the contribution of displacing water from hydration sites may also be flawed as the positions and thermodynamics of the hydration sites often changes between *rigid* and *flexible* cavities (e.g. see Figure 16, HMDH).

Acknowledgments: R35-GM144089, OpenEye tools[87]. Many thanks to Professor Emilio Gallichio @ Brooklyn College, Dan McKay from Ventus Therapeutics, and Professor Michael K. Gilson @ UCSD for insightful comments and discussion.

6 References

(1) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys J* **1997**, *72* (3), 1047–1069. https://doi.org/10.1016/S0006-3495(97)78756-3.

(2) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102* (18), 3531–3541. https://doi.org/10.1021/jp9723574.

(3) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102* (18), 3542–3550. https://doi.org/10.1021/jp972358w.

(4) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein–Ligand Binding. *Proceedings of the National Academy of Sciences* **2007**, *104* (3), 808–813. https://doi.org/10.1073/pnas.0610202104.

(5) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]Uril. *J Chem Phys* **2012**, *137* (4), 044101. https://doi.org/10.1063/1.4733951.

(6) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST. *J Comput Chem* **2016**, *37* (21), 2029–2037. https://doi.org/10.1002/jcc.24417.
(7) Percus, J. K.; Frisch, H. L.; Lebowitz, J. L. The Equilibrium Theory of Classical Fluids. *by HL Frisch and JL Lebowitz, Benjamin, New York. (1964)* **1964**, No. p II-33.
(8) He, P.; Sarkar, S.; Gallicchio, E.; Kurtzman, T.; Wickstrom, L. Role of Displacing Confined Solvent in the Conformational Equilibrium of β-Cyclodextrin. *J. Phys. Chem. B* **2019**, *123* (40), 8378–8386. https://doi.org/10.1021/acs.jpcb.9b07028.
(9) Haider, K.; Cruz, A.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **2018**, *14* (1), 418–425. https://doi.org/10.1021/acs.jctc.7b00592.
(10) Haider, K.; Wickstrom, L.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Enthalpic Breakdown of Water Structure on Protein Active-Site Surfaces. *The Journal of Physical Chemistry B* **2016**, *120* (34), 8743–8756. https://doi.org/10.1021/acs.jpcb.6b01094.
(11) Kovalenko, A.; Hirata, F. Self-Consistent Description of a Metal–Water Interface by the Kohn–Sham Density Functional Theory and the Three-Dimensional Reference Interaction Site Model. *The Journal of Chemical Physics* **1999**, *110* (20), 10095–10112. https://doi.org/10.1063/1.478883.
(12) Imai, T.; Kovalenko, A.; Hirata, F. Solvation Thermodynamics of Protein Studied by the 3D-RISM Theory. *Chemical Physics Letters* **2004**, *395* (1–3), 1–6. https://doi.org/10.1016/j.cplett.2004.06.140.
(13) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6* (3), 607–624. https://doi.org/10.1021/ct900460m.
(14) SZMap. SZMAP, 1.0.0; OpenEye Scientific Software Inc.: Santa Fe, NM, USA, 2011.There is no corresponding record for this reference.
(15) Bayden, A. S.; Moustakas, D. T.; Joseph-McCarthy, D.; Lamb, M. L. Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *J. Chem. Inf. Model.* **2015**, *55* (8), 1552–1565. https://doi.org/10.1021/ci500746d.
(16) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47* (2), 279–294. https://doi.org/10.1021/ci600253e.
(17) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer: New York, NY, 2012; pp 393–404. https://doi.org/10.1007/978-1-61779-465-0_24.
(18) Velez-Vega, C.; McKay, D. J. J.; Aravamuthan, V.; Pearlstein, R.; Duca, J. S. Time-Averaged Distributions of Solute and Solvent Motions: Exploring Proton Wires of GFP and PfM2DH. *J. Chem. Inf. Model.* **2014**, *54* (12), 3344–3361. https://doi.org/10.1021/ci500571h.
(19) Velez-Vega, C.; McKay, D. J. J.; Kurtzman, T.; Aravamuthan, V.; Pearlstein, R. A.; Duca, J. S. Estimation of Solvation Entropy and Enthalpy via Analysis of Water Oxygen–Hydrogen Correlations. *J Chem Theory Comput* **2015**, *11* (11), 5090–5102. https://doi.org/10.1021/acs.jctc.5b00439.

(20) López, E. D.; Arcon, J. P.; Gauto, D. F.; Petruk, A. A.; Modenutti, C. P.; Dumas, V. G.; Marti, M. A.; Turjanski, A. G. WATCLUST: A Tool for Improving the Design of Drugs Based on Protein-Water Interactions. *Bioinformatics* **2015**, *31* (22), 3697–3699. https://doi.org/10.1093/bioinformatics/btv411.
(21) Amber20.Pdf. https://ambermd.org/doc12/Amber20.pdf (accessed 2024-05-15).
(22) Amber22.Pdf. https://ambermd.org/doc12/Amber22.pdf (accessed 2024-05-22).
(23) Mysinger, M. M.; Carchia, M.; Irwin, John. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594. https://doi.org/10.1021/jm300687e.
(24) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(25) Schrödinger Release 2024-2: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2024; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2024.
(26) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J Comput Aided Mol Des* **2013**, *27* (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.
(27) Schrödinger Release 2024-2: Maestro, Schrödinger, LLC, New York, NY, 2024.
(28) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(29) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871. https://doi.org/10.1021/jz501780a.
(30) Boothroyd, S.; Behara, P. K.; Madin, O. C.; Hahn, D. F.; Jang, H.; Gapsys, V.; Wagner, J. R.; Horton, J. T.; Dotson, D. L.; Thompson, M. W.; Maat, J.; Gokey, T.; Wang, L.-P.; Cole, D. J.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Shirts, M. R.; Mobley, D. L. Development and Benchmarking of Open Force Field 2.0.0: The Sage Small Molecule Force Field. *J. Chem. Theory Comput.* **2023**, *19* (11), 3251–3275. https://doi.org/10.1021/acs.jctc.3c00039.
(31) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics* **1980**, *72* (4), 2384–2393. https://doi.org/10.1063/1.439486.
(32) Grest, G. S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys Rev A Gen Phys* **1986**, *33* (5), 3628–3631. https://doi.org/10.1103/physreva.33.3628.
(33) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-Methylamide. *Biopolymers* **1992**, *32* (5), 523–535. https://doi.org/10.1002/bip.360320508.
(34) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(35) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Computer Physics Communications* **2013**, *184* (2), 374–380. https://doi.org/10.1016/j.cpc.2012.09.022.
(36) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.
(37) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.
(38) Chen, L.; Cruz, A.; Roe, D. R.; Simmonett, A. C.; Wickstrom, L.; Deng, N.; Kurtzman, T. Thermodynamic Decomposition of Solvation Free Energies with Particle Mesh Ewald and Long-Range Lennard-Jones Interactions in Grid Inhomogeneous Solvation Theory. *J Chem Theory Comput* **2021**, *17* (5), 2714–2724. https://doi.org/10.1021/acs.jctc.0c01185.
(39) Huggins, D. J. Application of Inhomogeneous Fluid Solvation Theory to Model the Distribution and Thermodynamics of Water Molecules around Biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14* (43), 15106–15117. https://doi.org/10.1039/C2CP42631E.
(40) Huggins, D. J. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophysical Journal* **2015**, *108* (4), 928–936. https://doi.org/10.1016/j.bpj.2014.12.035.
(41) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10* (7), 2769–2780. https://doi.org/10.1021/ct401110x.
(42) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817–2831. https://doi.org/10.1021/ja0771033.
(43) Luzar, A.; Chandler, D. Hydrogen-Bond Kinetics in Liquid Water. *Nature* **1996**, *379* (6560), 55–57. https://doi.org/10.1038/379055a0.
(44) Gilson, M. K.; Kurtzman, T. Free Energy Density of a Fluid and Its Role in Solvation and Binding. *J. Chem. Theory Comput.* **2024**, *20* (7), 2871–2887. https://doi.org/10.1021/acs.jctc.3c01173.
(45) Uehara, S.; Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **2016**, *21* (11), 1604. https://doi.org/10.3390/molecules21111604.
(46) Ndubaku, C.; Varfolomeev, E.; Wang, L.; Zobel, K.; Lau, K.; Elliott, L. O.; Maurer, B.; Fedorova, A. V.; Dynek, J. N.; Koehler, M.; Hymowitz, S. G.; Tsui, V.; Deshayes, K.; Fairbrother, W. J.; Flygare, J. A.; Vucic, D. Antagonism of C-IAP and XIAP Proteins Is Required for Efficient Induction of Cell Death by Small-Molecule IAP Antagonists. *ACS Chem. Biol.* **2009**, *4* (7), 557–566. https://doi.org/10.1021/cb900083m.
(47) Yang, X.; Hu, Y.; Yin, D. H.; Turner, M. A.; Wang, M.; Borchardt, R. T.; Howell, P. L.; Kuczera, K.; Schowen, R. L. Catalytic Strategy of S-Adenosyl-l-Homocysteine Hydrolase: Transition-State Stabilization and the Avoidance of Abortive Reactions. *Biochemistry* **2003**, *42* (7), 1900–1909. https://doi.org/10.1021/bi0262350.
(48) Oyama, T.; Toyota, K.; Waku, T.; Hirakawa, Y.; Nagasawa, N.; Kasuga, J.; Hashimoto, Y.; Miyachi, H.; Morikawa, K. Adaptability and Selectivity of Human Peroxisome

Proliferator-Activated Receptor (PPAR) Pan Agonists Revealed from Crystal Structures. *Acta Crystallogr D Biol Crystallogr* **2009**, *65* (Pt 8), 786–795. https://doi.org/10.1107/S0907444909015935.
(49) Watermeyer, J. M.; Kröger, W. L.; O'Neill, H. G.; Sewell, B. T.; Sturrock, E. D. Probing the Basis of Domain-Dependent Inhibition Using Novel Ketone Inhibitors of Angiotensin-Converting Enzyme. *Biochemistry* **2008**, *47* (22), 5942–5950. https://doi.org/10.1021/bi8002605.
(50) Wendt, M. D.; Geyer, A.; McClellan, W. J.; Rockway, T. W.; Weitzberg, M.; Zhao, X.; Mantei, R.; Stewart, K.; Nienaber, V.; Klinghofer, V.; Giranda, V. L. Interaction with the S1β-Pocket of Urokinase: 8-Heterocycle Substituted and 6,8-Disubstituted 2-Naphthamidine Urokinase Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2004**, *14* (12), 3063–3068. https://doi.org/10.1016/j.bmcl.2004.04.030.
(51) Sarver, R. W.; Bills, E.; Bolton, G.; Bratton, L. D.; Caspers, N. L.; Dunbar, J. B.; Harris, M. S.; Hutchings, R. H.; Kennedy, R. M.; Larsen, S. D.; Pavlovsky, A.; Pfefferkorn, J. A.; Bainbridge, G. Thermodynamic and Structure Guided Design of Statin Based Inhibitors of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase. *J. Med. Chem.* **2008**, *51* (13), 3804–3813. https://doi.org/10.1021/jm7015057.
(52) Wang, J. L.; Limburg, D.; Graneto, M. J.; Springer, J.; Hamper, J. R. B.; Liao, S.; Pawlitz, J. L.; Kurumbail, R. G.; Maziasz, T.; Talley, J. J.; Kiefer, J. R.; Carter, J. The Novel Benzopyran Class of Selective Cyclooxygenase-2 Inhibitors. Part 2: The Second Clinical Candidate Having a Shorter and Favorable Human Half-Life. *Bioorganic & Medicinal Chemistry Letters* **2010**, *20* (23), 7159–7163. https://doi.org/10.1016/j.bmcl.2010.07.054.
(53) Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *The Journal of Chemical Physics* **1945**, *13* (11), 507–532. https://doi.org/10.1063/1.1723985.
(54) Kuhn, B.; Hilpert, H.; Benz, J.; Binggeli, A.; Grether, U.; Humm, R.; Märki, H. P.; Meyer, M.; Mohr, P. Structure-Based Design of Indole Propionic Acids as Novel PPARα/γ Co-Agonists. *Bioorganic & Medicinal Chemistry Letters* **2006**, *16* (15), 4016–4020. https://doi.org/10.1016/j.bmcl.2006.05.007.
(55) Smaill, J. B.; Lee, H. H.; Palmer, B. D.; Thompson, A. M.; Squire, C. J.; Baker, E. N.; Booth, R. J.; Kraker, A.; Hook, K.; Denny, W. A. Synthesis and Structure-Activity Relationships of Soluble 8-Substituted 4-(2-Chlorophenyl)-9-Hydroxypyrrolo[3,4-c]Carbazole-1,3(2H,6H)-Diones as Inhibitors of the Wee1 and Chk1 Checkpoint Kinases. *Bioorg Med Chem Lett* **2008**, *18* (3), 929–933. https://doi.org/10.1016/j.bmcl.2007.12.046.
(56) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis*. *Proc Natl Acad Sci U S A* **1958**, *44* (2), 98–104.
(57) Fedorov, R.; Hartmann, E.; Ghosh, D. K.; Schlichting, I. Structural Basis for the Specificity of the Nitric-Oxide Synthase Inhibitors W1400 and N ω-Propyl-l-Arg for the Inducible and Neuronal Isoforms *. *Journal of Biological Chemistry* **2003**, *278* (46), 45818–45825. https://doi.org/10.1074/jbc.M306030200.
(58) Fokkens, J.; Klebe, G. A Simple Protocol To Estimate Differences in Protein Binding Affinity for Enantiomers without Prior Resolution of Racemates. *Angewandte Chemie International Edition* **2006**, *45* (6), 985–989. https://doi.org/10.1002/anie.200502302.
(59) Baldwin, A. J.; Kay, L. E. NMR Spectroscopy Brings Invisible Protein States into Focus. *Nat Chem Biol* **2009**, *5* (11), 808–814. https://doi.org/10.1038/nchembio.238.

(60) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein Structure Determination from NMR Chemical Shifts. *Proceedings of the National Academy of Sciences* **2007**, *104* (23), 9615–9620. https://doi.org/10.1073/pnas.0610313104.
(61) Sekhar, A.; Kay, L. E. NMR Paves the Way for Atomic Level Descriptions of Sparsely Populated, Transiently Formed Biomolecular Conformers. *Proc Natl Acad Sci U S A* **2013**, *110* (32), 12867–12874. https://doi.org/10.1073/pnas.1305688110.
(62) Peplow, M. Cryo-Electron Microscopy Reaches Resolution Milestone. *ACS Cent. Sci.* **2020**, *6* (8), 1274–1277. https://doi.org/10.1021/acscentsci.0c01048.
(63) Sun, C.; Zhu, H.; Clark, S.; Gouaux, E. Cryo-EM Structures Reveal Native GABAA Receptor Assemblies and Pharmacology. *Nature* **2023**, *622* (7981), 195–201. https://doi.org/10.1038/s41586-023-06556-w.
(64) Glaeser, R. M. How Good Can Cryo-EM Become? *Nat Methods* **2016**, *13* (1), 28–32. https://doi.org/10.1038/nmeth.3695.
(65) Glaeser, R. M. How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits? *Annu. Rev. Biophys.* **2019**, *48* (1), 45–61. https://doi.org/10.1146/annurev-biophys-070317-032828.
(66) Fenwick, R. B.; van den Bedem, H.; Fraser, J. S.; Wright, P. E. Integrated Description of Protein Dynamics from Room-Temperature X-Ray Crystallography and NMR. *Proceedings of the National Academy of Sciences* **2014**, *111* (4), E445–E454. https://doi.org/10.1073/pnas.1323440111.
(67) Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing Protein Conformational Ensembles Using Room-Temperature X-Ray Crystallography. *Proceedings of the National Academy of Sciences* **2011**, *108* (39), 16247–16252. https://doi.org/10.1073/pnas.1111325108.
(68) Skaist Mehlman, T.; Biel, J. T.; Azeem, S. M.; Nelson, E. R.; Hossain, S.; Dunnett, L.; Paterson, N. G.; Douangamath, A.; Talon, R.; Axford, D.; Orins, H.; von Delft, F.; Keedy, D. A. Room-Temperature Crystallography Reveals Altered Binding of Small-Molecule Fragments to PTP1B. *eLife* **2023**, *12*, e84632. https://doi.org/10.7554/eLife.84632.
(69) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Computational Molecular Science* **2011**, *1* (5), 826–843. https://doi.org/10.1002/wcms.31.
(70) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat Rev Phys* **2020**, *2* (4), 200–212. https://doi.org/10.1038/s42254-020-0153-0.
(71) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a P38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139* (13), 4780–4788. https://doi.org/10.1021/jacs.6b12950.
(72) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111* (23), 230602. https://doi.org/10.1103/PhysRevLett.111.230602.
(73) Govind Kumar, V.; Polasa, A.; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Binding Affinity Estimation from Restrained Umbrella Sampling Simulations. *Nat Comput Sci* **2023**, *3* (1), 59–70. https://doi.org/10.1038/s43588-022-00389-9.
(74) You, W.; Tang, Z.; Chang, C. A. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J. Chem. Theory Comput.* **2019**, *15* (4), 2433–2443. https://doi.org/10.1021/acs.jctc.8b01142.
(75) Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J. Chem. Theory Comput.* **2007**, *3* (4), 1231–1235. https://doi.org/10.1021/ct700032n.

(76) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Current Opinion in Structural Biology* **2011**, *21* (2), 150–160. https://doi.org/10.1016/j.sbi.2011.01.011.
(77) Qi, R.; Wei, G.; Ma, B.; Nussinov, R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Methods Mol Biol* **2018**, *1777*, 101–119. https://doi.org/10.1007/978-1-4939-7811-3_5.
(78) Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Woldeyes, R. A.; Schneidman-Duhovny, D.; Demerdash, O. N.; Mitchell, J. C.; Wells, J. A.; Fraser, J. S.; Sali, A. CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol* **2016**, *428* (4), 709–719. https://doi.org/10.1016/j.jmb.2016.01.029.
(79) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(80) Meller, A.; Bhakat, S.; Solieva, S.; Bowman, G. R. Accelerating Cryptic Pocket Discovery Using AlphaFold. *J. Chem. Theory Comput.* **2023**, *19* (14), 4355–4363. https://doi.org/10.1021/acs.jctc.2c01189.
(81) del Alamo, D.; Sala, D.; Mchaourab, H. S.; Meiler, J. Sampling Alternative Conformational States of Transporters and Receptors with AlphaFold2. *eLife 11*, e75751. https://doi.org/10.7554/eLife.75751.
(82) Stein, R. A.; Mchaourab, H. S. SPEACH_AF: Sampling Protein Ensembles and Conformational Heterogeneity with Alphafold2. *PLoS Comput Biol* **2022**, *18* (8), e1010483. https://doi.org/10.1371/journal.pcbi.1010483.
(83) Nguyen, C. N.; Kurtzman, T.; Gilson, M. K. Spatial Decomposition of Translational Water–Water Correlation Entropy in Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12* (1), 414–429. https://doi.org/10.1021/acs.jctc.5b00939.
(84) Nguyen, C.; Yamazaki, T.; Kovalenko, A.; Case, D. A.; Gilson, M. K.; Kurtzman, T.; Luchko, T. A Molecular Reconstruction Approach to Site-Based 3D-RISM and Comparison to GIST Hydration Thermodynamic Maps in an Enzyme Active Site. *PLOS ONE* **2019**, *14* (7), e0219473. https://doi.org/10.1371/journal.pone.0219473.
(85) Huang, N.; Shoichet, B. K. Exploiting Ordered Waters in Molecular Docking. *J. Med. Chem.* **2008**, *51* (16), 4862–4865. https://doi.org/10.1021/jm8006239.
(86) Higgs, C.; Beuming, T.; Sherman, W. Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Med. Chem. Lett.* **2010**, *1* (4), 160–164. https://doi.org/10.1021/ml100008s.
(87) QUACPAC 2.2.3.3. OpenEye, Cadence Molecular Sciences, Santa Fe, NM. Http://Www.Eyesopen.Com.