A Large Scale Study of Hydration Environments Through Hydration Sites

Benedict W. J. Irwin,[†] Sinisa Vukovic,[†] Michael C. Payne,[†] and David J. Huggins[*,†,‡,¶]

[†]*Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, 19 J J Thomson Avenue, Cambridge, CB3 0HE, United Kingdom*

[‡]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

[¶]*Weill Cornell Medical College, Department of Physiology and Biophysics, 1300 York Avenue, New York, New York 10065, United States*

E-mail: djh210@cam.ac.uk

Abstract

Hydration sites are locations of interest to water and they can be used to classify the behaviour of water around chemical motifs commonly found on the surface of proteins. Inhomogeneous fluid solvation theory (IFST) is a method for calculating hydration free energy changes from molecular dynamics (MD) trajectories. In this paper hydration sites are identified from MD simulations of 380 diverse protein structures. The hydration free energies of the hydration sites are calculated using IFST and distributions of these free energy changes are analysed. The results show that for some hydration sites near features conventionally regarded as attractive to water, such as hydrogen bond donors, the water molecules are actually relatively weakly bound and are easily displaced. We also construct plots of the spatial density of hydration sites with high medium and low

hydration free energy changes which represent weakly and strongly bound hydration sites. It is found that these plots show consistent features around common polar amino acids for all of the proteins studied.

Introduction

Hydration plays a large role in the interaction of molecules with proteins.[1–5] The behaviour of water molecules at the site of interaction depends strongly on the local protein environment.[6–10] The concept of hydration sites (HS) has been used in many methods to describe the energetic properties of water molecules based on their local environment.[10–16] Recent methods based on HS include finding hydrophobic hotspots on the protein surface which can be used for ligandable hotspot prediction and the prediction of hydrophobic protein-protein interactions[17,18] . If a water molecule is easy to displace (i.e. highly displaceable) then the HS associated with that water molecule is relatively hydrophobic. If the HS is not displaceable, then it is hydrophilic. By analysing the distribution of hydration free energy across the protein surface, one can see which protein locations on average have the most displaceable water.

In the context of this work, the HS used are points of interest which water molecules occupy both on the surface of a protein and within pockets accessible to water molecules.[19] The HS are identified by 'clustering' water molecules across frames from a molecular dynamics (MD) simulation.[20] The free energy contribution from a water molecule at that site is then calculated using inhomogeneous fluid solvation methods derived from thermodynamics and statistical mechanics.[4,21**?** ,22] This work in particular uses inhomogeneous fluid solvation theory (IFST)[3,4,19,23–25] which calculates the free energy associated with removing water molecules from a HS on the surface of a protein by decomposing the free energy into energetic and entropic contributions. In the current work, this breakdown also goes further, decomposing the entropy into translational and orientational terms.[19,26] Those entropy terms can be broken down yet further into solute-water and water-water terms.[15,27] The predictive

properties of these higher order terms have not yet been established, but the total free energy composed of them is well understood and has shown to reproduce free energies calculated by free energy perturbation well.[19,26–28]

By calculating the free energy of hydration of each HS using IFST for a range of proteins, statistical information can be gleaned about the average nature of water in the local protein environment. Examples of these local environments are carboxyl oxygen groups and amide nitrogen groups. Regions above and below the plane of amide groups which contribute unfavourably to hydration free energies have been noticed in IFST calculations on small molecules.[28] The geometry of such motifs is similar in protein amide residues and may contribute to the presence of highly displaceable hydration sites. A notable class of such sites that are connected with displacability are the so called dehydrons.[29]

Dehydrons are biologically important structural motifs related to packing defects in the protein backbone which lead to solvent exposed hydrogen bonds.[29–31] They are highly sensitive to the local solvent environment through the changing Coulombic interactions associated with the displacement of water molecules and can be stabilised by bringing a hydrophobe toward to a backbone hydrogen bond which has fewer than average hydrophobic groups.[29,32] On the other hand, they act as sticky sites which may become hydrated leading to conformational changes in the protein.[32,33] The presence of dehydrons correlates with binding sites in proteins. They are important features in protein-protein complexation, protein-ligand interactions, high level structural arrangement and other biochemical phenomena and bioinformatical applications such as measuring proteomic complexity and selective inhibitor design.[29,31,34] These examples support and motivate the goal of understanding the specific hydration of protein binding sites. Although dehydrons are specific to hydrogen bonding near the protein backbone, one would expect to find both analogous and alternative hydration features near other protein side-chains. Amide groups in asparagine and glutamine should share similar features to the protein backbone, whereas carboxyl groups in aspartate and glutamate residues will have different hydration properties.

Similar studies of hydration sites have been made on a small set of proteins in the past. Grid cell theory was used by Gergiokas et al. on 17 proteins which had large variations in structure.[16] These authors comment that for water molecules near polar or hydrophobic regions of the protein, stability strongly depends on the coordination of the local hydration environment. Beuming et al. used WaterMap to conduct a distributional analysis on a larger set of 27 proteins.[14] These authors found that high energy hydration sites exist near backbone amide and other hydrophilic groups and concluded that hydration sites near the backbone of the protein are much less favourable than HS near polar side chains. They also found that the water molecules in the vicinity of secondary structures are less strongly bound. They also conclude that there is no direct relation between the energetics of a hydration site and the degree of burial of the HS within a protein. Both the strong dependence on coordination and the existence of high energy sites near the protein backbone agree with the work of Fernandez et al. on dehydron environments.[29,30,32] However, Fernandez et al. find that the degree of wrapping of hydrogen bonds near the protein backbone by hydrophobic groups strongly affects the stability of the protein, and the connection of this to the general 'degree of burial' is not clear.

Beuming et al.[14] found no correlation between the solvent accessible area and the hydration free energy of hydration sites. Whereas Gerogiokas et al.[16] state that the solvent accessible volume is large in binding sites. Gerogiokas et al. also state that there is no correlation between average water thermodynamic properties and the classification of a protein face as binding or non-binding.[16] However, Beuming et al.[14] state that regions with unstable HS indicate binding sites for drug-like molecules. This agrees with studies by Vukovic et al. that have reported success in identifying binding sites using HS for the bromodomains family of proteins.[17] Furthermore, there is a correlation between the net free energy of the most displaceable site on a protein as calculated using HS and an experimentally derived measure of 'ligandability'.[35] Irwin et al. extended this method to protein-protein interactions where HS with hydration free energies calculated using IFST revealed the location of attractive

patches on the $\gamma 2$ subunit of the $GABA_A$ receptor.[18] These patches were associated with hydrophobic amino acids and helped infer a docking pose for the two proteins. From this and other results using the hydration sites method we can conclude there is a correlation between the local environment of the solvated region and the free energy of water molecules around that region so that a detailed study of many hundreds of proteins will yield useful and interpretable results. To this end, this work applies IFST hydration site analysis to 380 proteins which gives substantially more data from which to not only understand HS in greater detail with statistical measures of their properties but also infer important properties that can be deduced from a knowledge of the HS. We attempt to classify the hydration free energy of sites according to their local environment by looking at distributions of the hydration free energy of HS around many different types of protein. From these distributions a notion of relative displacability of water molecules can be obtained which may be useful to medicinal chemists when designing ligands to displace certain targets. We provide overlaid residues and find the average 3-dimensional density of hydration sites around common residues along with their free-energetic density. We also investigate a proxy for burial as the number of hydrogen atoms close to a HS, and the impact this has on HS stability.

Methods

Proteins Simulated

380 proteins were simulated using molecular dynamics. The water molecules in the simulation were clustered using the same algorithm as Vukovic et al.[17,18,35] The resulting hydration sites were analysed using the HS method also used by Vukovic et al.[17,18,35] This gave a total of 357467 HS each of which has associated free energy, enthalpy and entropy values. The largest group of proteins in the dataset belonged to the bromodomain family covering 34 of the 61 known types of bromodomain.[36] A list of types of proteins studied is given in table 1. Some Protein Data Bank (PDB) entries had multiple structures and the analysis was performed

Table 1: Types of proteins included in this study which are sampled from a diverse range of families. Entries have been roughly categorised by their biological function. A full table of PDB tags is included in the supplementary material.

Family	Types
Integrase	HIV
Transcriptase	HIVRT
Caspases	CASP1
Gyrases	DNA-B
Esterases	AChE-1,
Receptors	A2AR, AChR
Phosphodiesterase	PDE 4D, PDE 5A
Reductases	aldose, enoyl, HMG CoA
Kinases	cAbl, EGRF, tyrosine, urokinase, CDK2, P38
Proteases	CTSK, CTSS, HIV-1p, HVC serine, thrombin, factor Xa
Bromodomains	ASH1L, ATAD2, ATAD2B, BAZ2A, BAZ2B, BPTF, BRD1, BRD2(1)
Bromodomains	BRD2(2), BRD3(1), BRD3(2), BRD4(1), BRD4(2), BRD7, BRD9
Bromodomains	BRPF1, CECR2, CREBBP, EP300, KAT2A, KAT2B, PBRM1(1)
Bromodomains	PBRM1(2), PBRM1(3), PBRM1(4), PBRM1(5), PHIP(2), SMARCA2
Bromodomains	SMARCA4, SP100, TAF1(1), TAF1(2), TAF1L(2), TIF1α
Other Enzyme	ACE, APE-1, COX-2, PTP1B, KRAS, CA1, CA2, IMPDH
Other	serum albumin, 14-3-3, cell division ZipA, fungal Cyp51, MDM2
Other	neuraminidase, Methylcytosine dioxygenase

on these replicates. A full table of PDB tags, conformer numbers and protein names and abbreviations is included in the supplementary material for reference.

Simulation Protocol

The simulation protocol was the same as that used by Vukovic et al.[17,35] PDB files were obtained from the Protein Data Bank.[37] Structures were prepared by removing co-solvents, adding hydrogen atoms and protonation states which were set using the Schrödinger PrepWizard[38] as performed by Vukovic et al.[17,35] Hydrogen bonds were also optimised using PrepWizard.[38] Proteins were simulated using NAMD2 in a Na^+, Cl^- neutralised periodic water box with 10 Å buffer from the edge using CHARMM27 atom charges protein atom types and TIP4P-2005 parameters for water molecules.[39–42] Harmonic restraints of strength 1.0 kcal/mol/Å^2 were used on all protein heavy atoms and systems underwent NPT equili-

bration at 300K for 1 ns followed by 3.6 ns of NPT data collection simulation. To generate HS, 1000 evenly spaced snapshots were used from the data collection runs. HS could thus be generated in any region where water was found across the MD simulations which includes the external surface of the protein and pockets and crevices deeper into the protein. An HS radius of 1.2 Å was used for clustering. Then the binding affinities were calculated using IFST from the Solvaware package.[17]

IFST

IFST in this work estimates the change in hydration free energy by measuring the change in solvation energy from the MD simulation and approximating the change in solvation entropy by truncating a series of entropy and mutual information integrals.[3,4,19,23–25] The implementation of IFST used in this work calculates the full first order entropy change and uses a fixed form for the second order term which has shown agreement with experimental results and other free energy methods.[19,26] Specifically, IFST has been compared directly to free energy perturbation methods for the CHARMM27 forcefield as used in this work and shows good agreement for Lennard-Jones systems,[27] for which the entropy terms could be accurately truncated to first order, the solvation of small molecules[25,26,28,43] and the desolvation of water molecules in binding pockets.[19,20] The IFST and HS clustering methods used in this work that are derived from MD snapshots will naturally have convergence problems if the protein moves excessively across the simulation. For this reason the protein is harmonically restrained which has worked well in previous studies.[17,35]

Statistics of Water Sites by Chemical Group

Histograms were made of the hydration free energy for all HS. The HS were divided into classes based on the local protein environment including the number of potential hydrogen bonds that could be made and/or the number of nearby non-polar heavy atoms (carbon and sulphur) that are near the hydration site. The definition of a hydrogen bond and heavy atom

in this context is purely based on a distance cut-off of 3.2 Å from the relevant O or N atom for hydrogen bonds and 4.5 Å for nearby heavy atoms (which could be of any type C,N,O,S).

The general distance characteristics of HS have been investigated with respect to neighbouring atoms. Figure 1 shows that hydration sites with a hydrogen bonding atom as their nearest heavy atom neighbour are likely to be closer to that nearest heavy atom neighbour, which is to be expected. There are very few HS near O or N atoms with a distance of more than 4 Å. The distribution appears to be roughly comprised of two sub peaks. These sub peaks likely belong to charged and non-charged interactions in order of increasing distance.

The 4.5 Å cut-off for heavy atoms is defined to keep the entire peak of the C and S atom distribution in figure 1. It can be seen that the range of distances under the hydrogen bonding atom curve is much wider than the curve for carbon and sulphur. There is a long left tail on this second distribution, likely because of slightly charged carbons and sites which are forced close to carbon atoms by unfavourable packing effects.

Defining Local Chemical Environments

The HS were initially classified into 19 classes of local chemical environment after which 6 of these classes were considered to either have too few data points or be too similar to another class. The final 13 classes are shown in table 2. The 6 classes that did not make final selection were, methionine sulphur, indole nitrogen, neutral histidine donor nitrogen, positive histidine nitrogen, arginine chain nitrogen and oxygen atoms on the terminus of the protein model, these classes were merged into the most similar class, which were cysteine sulphur, amide nitrogen, amide nitrogen, positive lysine nitrogen, positive arginine nitrogen and carboxylate oxygen, respectively. The classes roughly cover all 'interesting' types of C, N, O and S atoms which are likely to show chemical differences.

There were also HS which had more than one hydrogen bonding atom or heavy atom (or both), which created a more complex environment. These were not split up into exhaustive pairs of the classes in table 2 otherwise the data per class in each of the combinations would

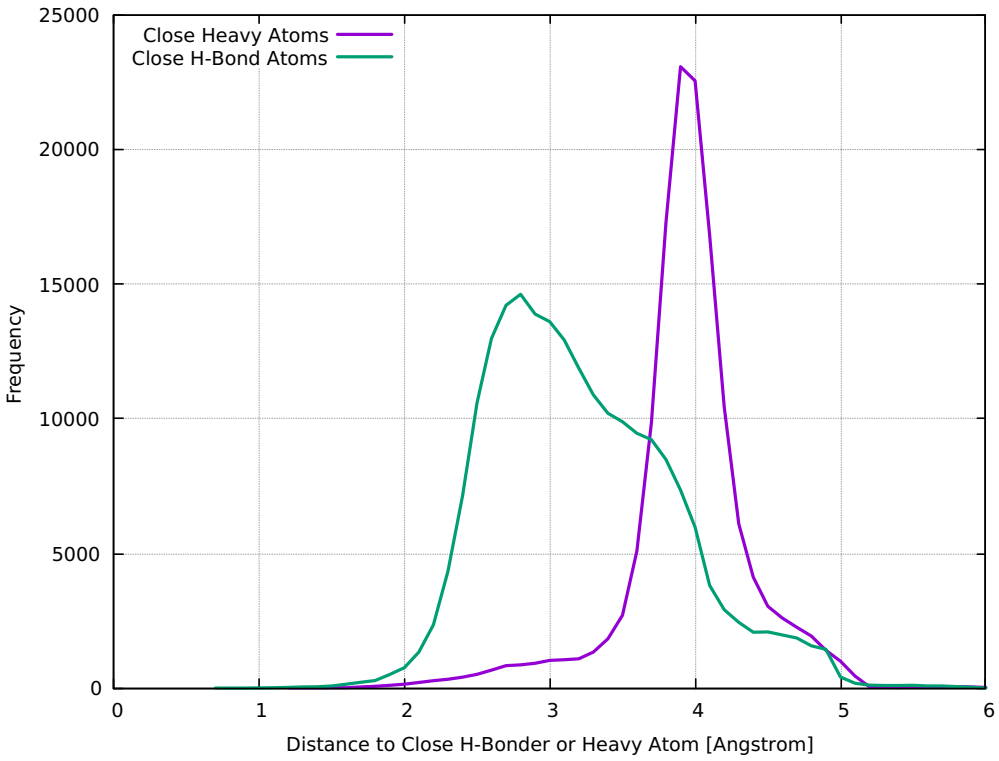

Figure 1: A comparison of the average nearest neighbour distance for hydration sites to their nearest O or N in the case of hydrogen bonding, or C or S in the case of heavy atoms. The HS near hydrogen bonders are generally closer to their nearest heavy atom.

have been too few. An additional labelling of sites is given in table 3, which covers all of the hydration sites in the data set. It can be seen from the count column in table 3 that sites with many hydrogen bonds are increasingly rare. This may be due to multiplication of probabilities, but evidence will be shown later that indicates these sites have a high probability of being unstable. There are also many HS with no heavy atoms or hydrogen bonds.

Table 2: The final set of classes used to define chemical environments around hydration sites. Sites with either 1 hydrogen bonding atom within range (for nitrogen and oxygen) or 1 heavy atom in range (for carbon and sulphur) could be classified in this way.

Class	Definition
Amide Oxygen	HS $<$ 2.4 Å of an O within amide part of Asn, Gln
Amide Nitrogen	HS $<$ 2.4 Å of an N within amide part of Asn, Gln
Hydroxyl Oxygen	HS $<$ 2.4 Å of Ser, Thr, Tyr OH group
Carboxylate Oxygen	HS $<$ 2.4 Å of Asp, Glu =O group
Non-polar Aliphatic Carbon	HS $<$ 3.2 Å of aliphatic C
Non-polar Aromatic Carbon	HS $<$ 3.2 Å of aromatic C (His, Phe, Tyr, Trp)
Sulphur	HS $<$ 3.2 Å of S on Cys, Met
Neutral Histidine Acceptor Nitrogen	HS $<$ 2.4 Å of N on Neutral His
Lysine Nitrogen	HS $<$ 2.4 Å of N on Lys
Arginine Nitrogen	HS $<$ 2.4 Å of N on Arg
Backbone Carbon	HS $<$ 3.2 Å of C on any backbone
Backbone Oxygen	HS $<$ 2.4 Å of O on any backbone
Backbone Nitrogen	HS $<$ 2.4 Å of N on any backbone

Table 3: Partitioning of the hydration sites into classes based on the number of hydrogen bonding atoms (O,N) within 3.2 Å denoted #HB and the number of heavy atoms (O,N,S,C) within 4.5 Å denoted #CA. The number of data points in each class is shown.

Count	#HB	HB Type	#CA	CA Type
71849	0	-	0	-
47463	0	-	1	Aliphatic C
5187	0	-	1	Aromatic C
951	0	-	1	Sulpur
32582	0	-	1	Not C,S
41737	0	-	2	Any
44716	0	-	>2	Any
2627	1	Amide Nitrogen	≥ 1	Any Type
5839	1	Amide Oxygen	≥ 1	Any Type
5521	1	Arginine Nitrogen	≥ 1	Any Type
4349	1	Backbone Nitrogen	≥ 1	Any Type
28943	1	Backbone Oxygen	≥ 1	Any Type
7541	1	Hydroxyl Oxygen	≥ 1	Any Type
9036	1	Lysine Nitrogen	≥ 1	Any Type
1298	1	Neutral Histidine Acceptor N	≥ 1	Any Type
27441	1	Carboxylate Oxygen	≥ 1	Any Type
49	1	Other	≥ 1	Any Type
16077	2	Any N,O	≥ 2	Any Type
3498	3	Any N,O	≥ 3	Any Type
763	>3	Any N,O	≥ 4	Any Type
357467	Total			

Hydration Site Densities for Local Environments

In order to characterise the average locations of HS around amino acids all instances of each residue were collected and aligned and the HS from these alignments were binned into density plots. The residues were aligned using PyMol.[44] Hydration sites within 3.5 Å of polar atoms were used to contribute to the density of different chemical environments. The density maps were created using the VolMap tool in the VMD software package.[45] Three density clusters were generated for each amino acid type: Dehydron/Strong hydrophobe HS with free energy greater than A kcal/mol (coloured yellow on density plots), midrange clusters with free energy from B to C kcal/mol (coloured purple), and relatively strongly binding sites with free energy less than D kcal/mol (coloured red) where the corresponding values for A, B, C and D for each amino acid were derived from the distributions of HS hydration free energies presented in the following section and are shown in table 4.

Results & Discussion

Figure 2 shows the log density of hydration sites as a function of the distance from the hydration site to the nearest O or N atom. Blue implies a low density and yellow a high density. This plot shows a few distinct features. Firstly, there is a spike of very displaceable sites with low hydration free energies (between -2 and 0 kcal/mol). These sites appear at all distances above 1 Å. Some of them are very close to their nearest O or N atom (< 2.2 Å) when compared to expected hydrogen bond distances. The rest of the sites are broadly scattered across a wide region which is dense for energies ranging from 0 to -20 kcal/mol. Sites with very negative free energy changes lower than -20.0 make up 0.1% of the dataset. It appears that sites with a positive hydration free energy (0.44% of all sites) do not appear at distances less than 2.5 Å from the nearest O or N atom. This plot indicates that this nearest heavy atom distance can be used to filter unusually displaceable sites to some extent.

Histograms of the distribution of HS hydration free energies with respect to different

Table 4: The free energy thresholds for labelling sites as strong hydrophobes, mid-level free energy sites in the broad part of the distribution or highly bound sites in the left tail of the free energy distribution. The inequalities used depend on the distributions of free energy for each class of HS.

Colour: Type/Amino Acid	Yellow Hydrophobe ΔG Range	Purple Mid ΔG Range	Red High ΔG Range
LYS NH_3	$\Delta G > -1.0$	$-6.0 < \Delta G < -3.0$	$\Delta G < -8.0$
ASN NH_2	$\Delta G > -1.0$	$-4.0 < \Delta G < -1.5$	$\Delta G < -8.0$
GLN NH_2	$\Delta G > -1.0$	$-4.0 < \Delta G < -1.5$	$\Delta G < -5.0$
ASP COOH	$\Delta G > -3.0$	$-9.0 < \Delta G < -4.0$	$\Delta G < -10.0$
GLU COOH	$\Delta G > -3.0$	$-9.0 < \Delta G < -4.0$	$\Delta G < -10.0$
SER OH	$\Delta G > -1.0$	$-4.5 < \Delta G < -2.0$	$\Delta G < -5.5$
THR OH	$\Delta G > -1.0$	$-4.5 < \Delta G < -2.0$	$\Delta G < -5.5$
TYR OH	$\Delta G > -1.0$	$-4.5 < \Delta G < -2.0$	$\Delta G < -5.5$
Backbone N	$\Delta G > -2.5$	$-6.5 < \Delta G < -3.5$	$\Delta G < -7.5$

chemical environments were created. In the following plots, all of the sites near different types of nitrogen or oxygen atoms were considered. These sites are expected to show the greatest activity according to their local environment as the water will be strongly affected by larger charges. The histograms for nitrogen type environments presented in figure 3 would suggest that HS near amide nitrogens are in general more displaceable than for other types. This shows a strong and unexpected contrast to the backbone nitrogen curve which is likely due to the average solvent exposure of amide side chains. The broad part of the amide nitrogen distribution is around 2 kcal/mol more positive than the other types of nitrogen. A large population of strong hydrophobes with free energies $-1 < \Delta G < 0$ kcal/mol can also be seen. This effect is somewhat less in sites near neutral histidine nitrogens but is still relatively pronounced and all classes appear to have some members in this range of energies. Backbone nitrogen sites show the deepest dip in free energy values between -3 and -1 kcal/mol. Based on the height of the peak at the broadest part of its distribution, backbone nitrogens are the least displaceable of the nitrogen types. All of the distributions have a sharp decline as the hydration free energy becomes very negative.

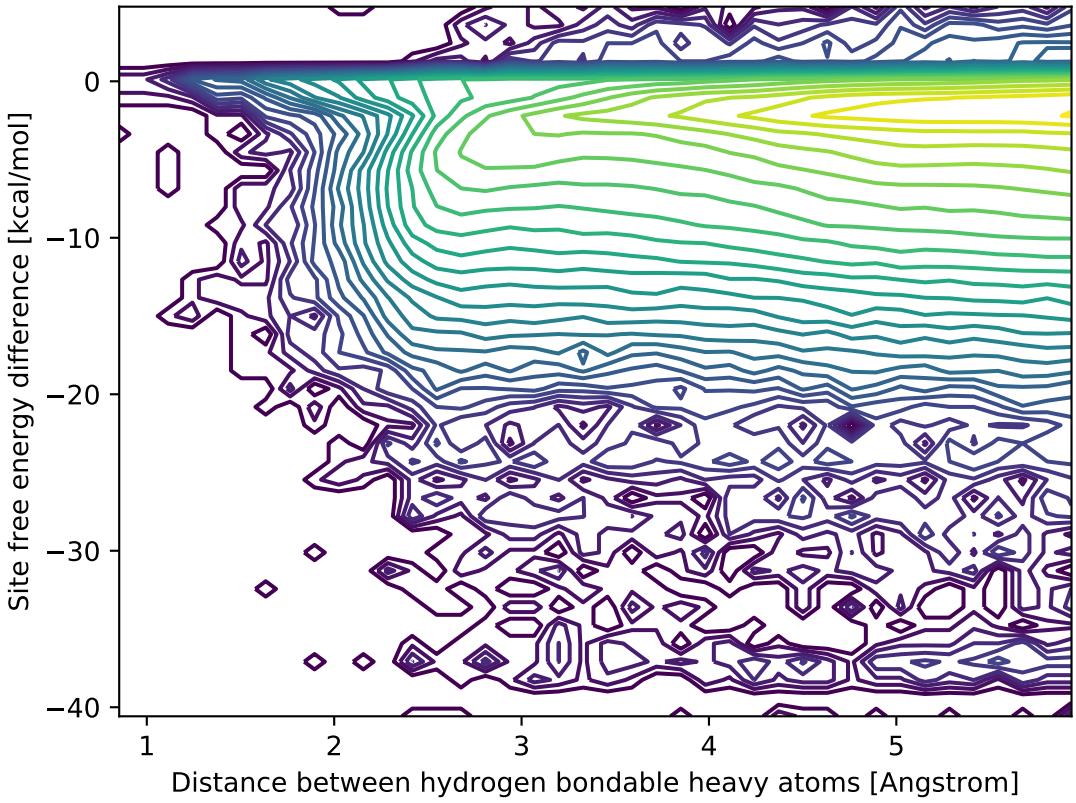

Figure 2: The hydration free energy of sites as a function of distance from their nearest nitrogen or oxygen atom. A heavily contoured feature can be seen for low free energy differences (0-1 kcal/mol) i.e. very displacable sites which are noticeably closer to their respective heavy atom than the majority of sites (less than 2 Å). There are very few sites with a positive free energy difference or highly negative free energy difference less than -20 kcal/mol. Yellow indicates high density, blue indicates low density.

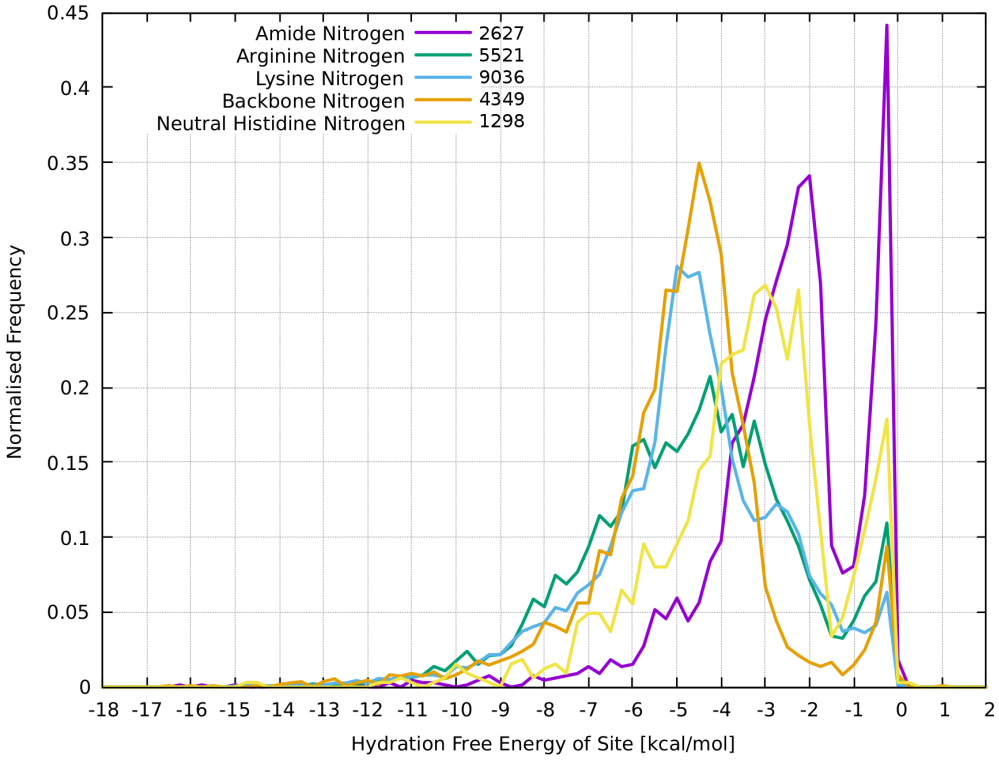

Figure 3: Histogram of the hydration free energy values for the five main types of nitrogen. Amide and histidine nitrogens can be seen to provide quite different environments for water. The number of hydration sites in each category is displayed in the legend.

In the following figures the local density of hydration sites within the three classes, strongly bound (red), moderately bound (purple) and weakly bound (yellow) as described in table 4 are shown relative to atoms from the superimposed residues. The coloured atoms correspond to parts of the residue which were closely aligned across all residues to that type. In these figures grey atoms are parts of the amino acid which cannot be aligned due to the inherent flexibility of the residue. These atoms take a range of positions and the grey atoms only represent one of the possible positions the atom could be found in for structural continuity. The coloured atoms are always aligned well due to the rigidity and geometry of the residue and for these atoms we can expect the HS density around that fixed feature to offer structural detail. Figure 4 shows the density of HS around the NH_3 group in lysine residues from two angles. The strong, moderate and weakly binding patches all have distinct locations. The weakly binding HS are concentrated around the hydrophobic carbon. Figure 5 shows the same information for nitrogen atoms in the asparagine residues. In this case of the weakly bound HS are found straddling the two hydrogen atoms and are relatively close to the nitrogen atom unable to form a good hydrogen bond. The weakly binding patch on glutamine is in a similar position to that of asparagine whose hydration site densities are shown in the supplementary material, for glutamine, the strong and moderate site patches overlap to some degree.

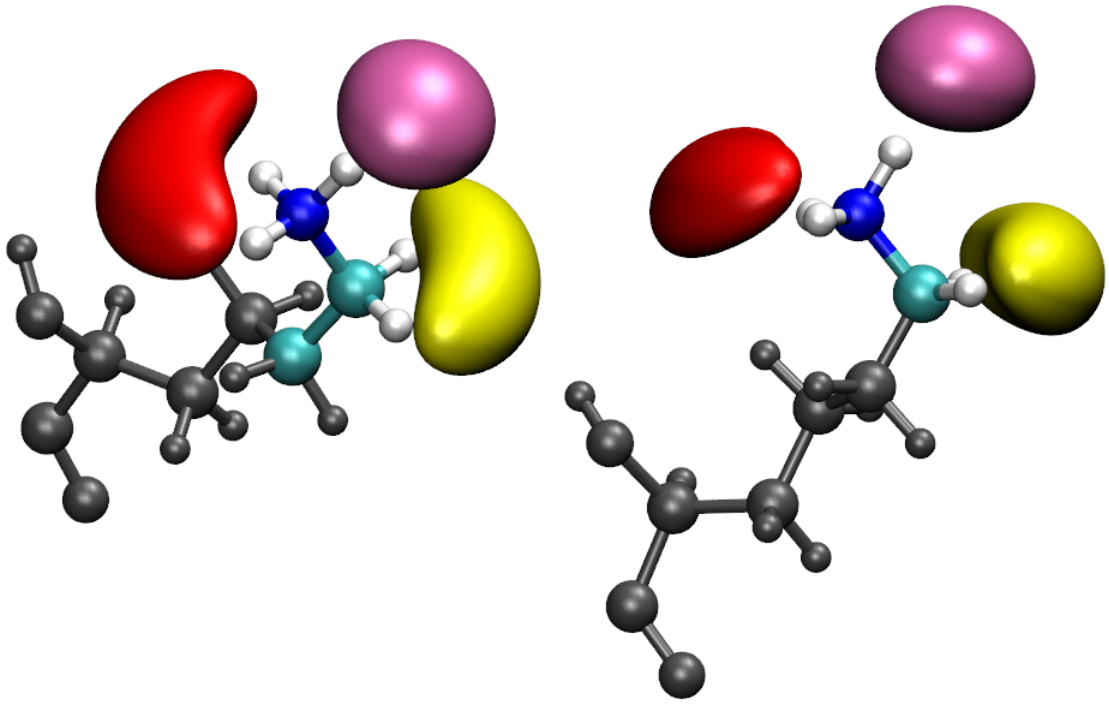

Figure 4: Lysine: For all Hydration sites within range, 5.6% are in the hydrophobe category (yellow), 27% are in the moderate category (purple) and 9.8% are in the strongly bound category (red) with other sites falling between these.

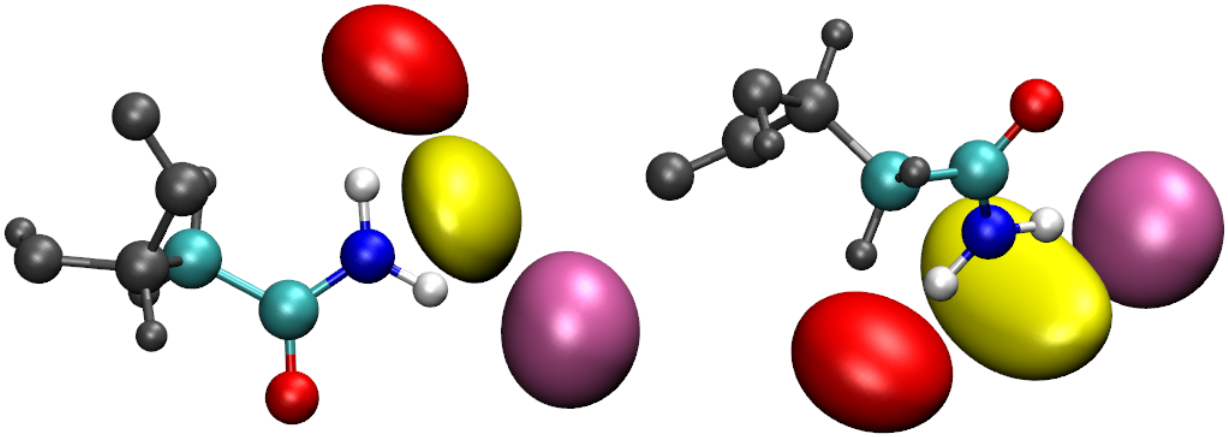

Figure 5: Asparagine: For all Hydration sites within range, 8.6% are in the hydrophobe category (yellow), 48% are in the moderate category (purple) and 17% are in the strongly bound category (red) with other sites falling between these.

Chemical Environment of Backbone Nitrogen

The backbone nitrogen environment has two environments; one for the carbon atom shown in figure 6 and one for the nitrogen atom (shown in the supplementary material), both shown in three different viewing angles. The hydration environment near the oxygen atom was not distinct, with all three classes of site overlapping directly opposite the oxygen.

Oxygen

Figure 7 shows that, in general, sites around the carboxylate groups (in aspartic and glutamic acid) are much harder to displace and show a broader distribution of hydration free energies. The hydration free energies in the range from -3 to -14 kcal/mol would suggest that often two separate hydrogen bonds are made. There is a very low chance of finding carboxylate sites with hydration free energies between -3 and 0 kcal/mol especially compared to the amide oxygen environment which appears to be the most displaceable common oxygen environment. Figure 7 also shows that the behaviour of backbone and hydroxyl oxygen sites is similar. There exist more very displaceable hydroxyl sites, but sites which are in the broad part of the distribution are marginally more stable than backbone oxygen sites.

Figure 8 show the HS density for glutamate residues. The grey atoms have been removed in this figure for clarity. In glutamate the three classes of HS overlap to some degree with a broader range of strongly binding sites. A figure of aspartate residues is shown in the supplementary material for comparison. In aspartate residues, the strong and moderately binding HS are found in distinct locations. In both cases, the strongly binding sites lie in the same plane, which is in plane with the oxygen atoms. In both cases the weakly binding HS straddle this plane.

Figure 9 shows the hydroxyl oxygen based residue on serine. Similar plots for threonine and tyrosine are located in the supplementary material for comparison. In the case of serine, all classes of HS density overlap to some degree. The strongly binding sites take a broad range of positions. In all three of the hydroxyl bearing residues the weakly binding HS are

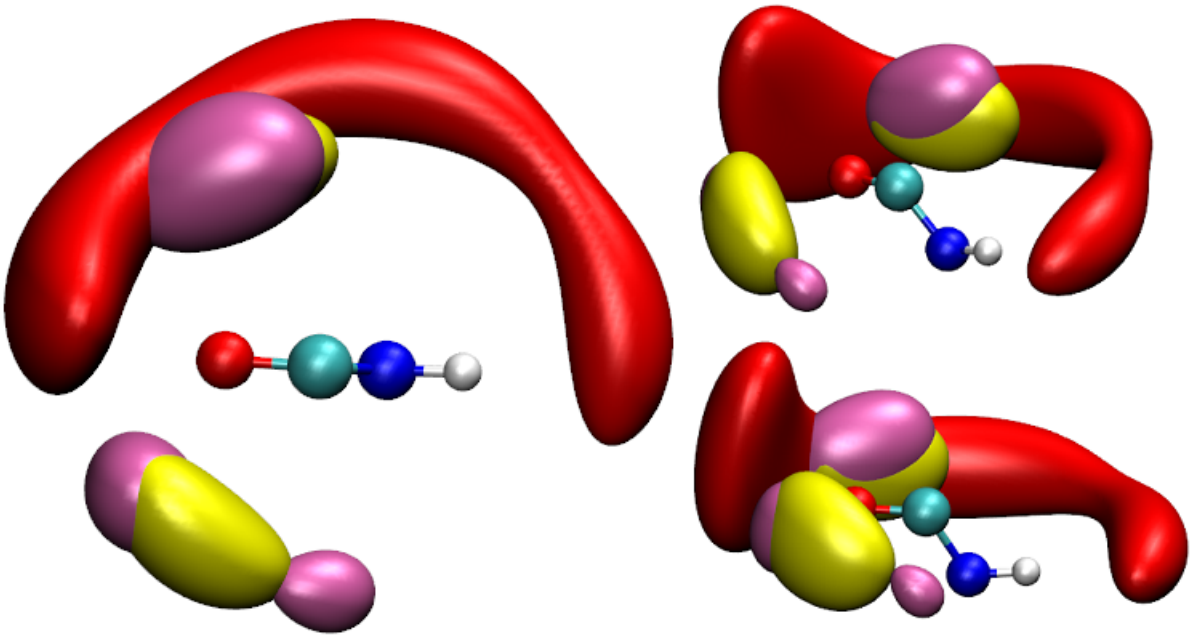

Figure 6: Backbone Nitrogen Carbon Environment: For all Hydration sites within range, 42% are in the hydrophobe category (yellow), 26% are in the moderate category (purple) and 12% are in the strongly bound category (red) with other sites falling between these.

seen to be concentrated directly opposite the oxygen atom along the oxygen carbon bond, whereas the strong and moderate sites appear to follow the possible directions of the hydroxyl group.

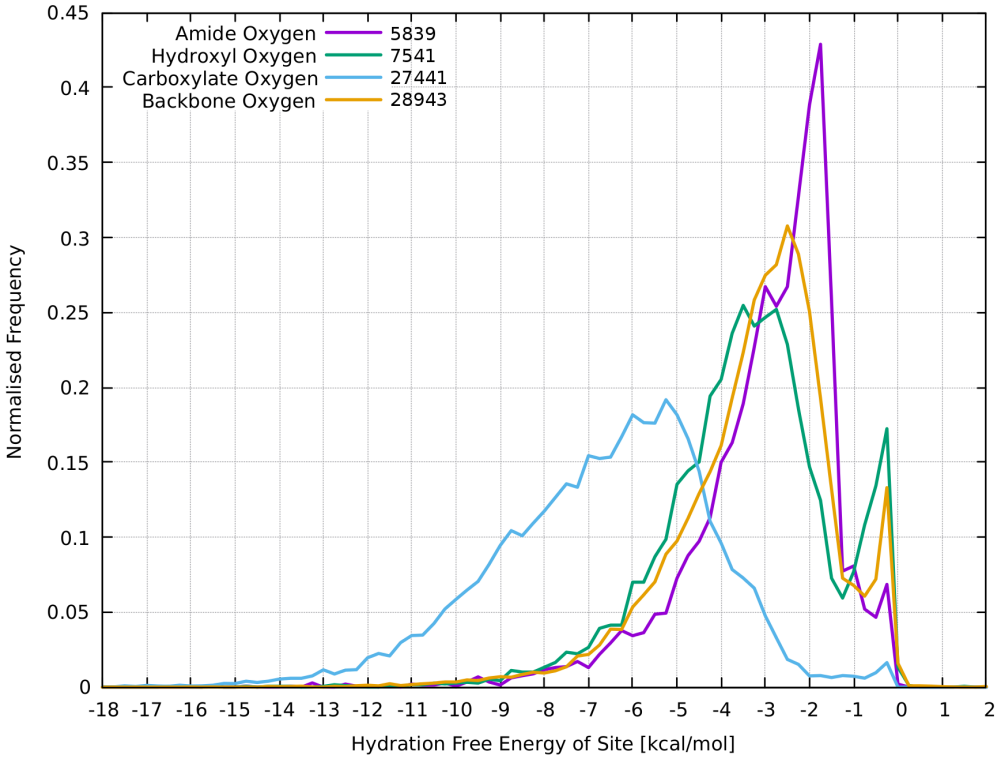

Figure 7: Histogram of the hydration free energy values for the 4 main types of oxygen. Carboxylate oxygens can be seen to be quite different environments for water around proteins. The number of hydration sites in each category is displayed in the legend.

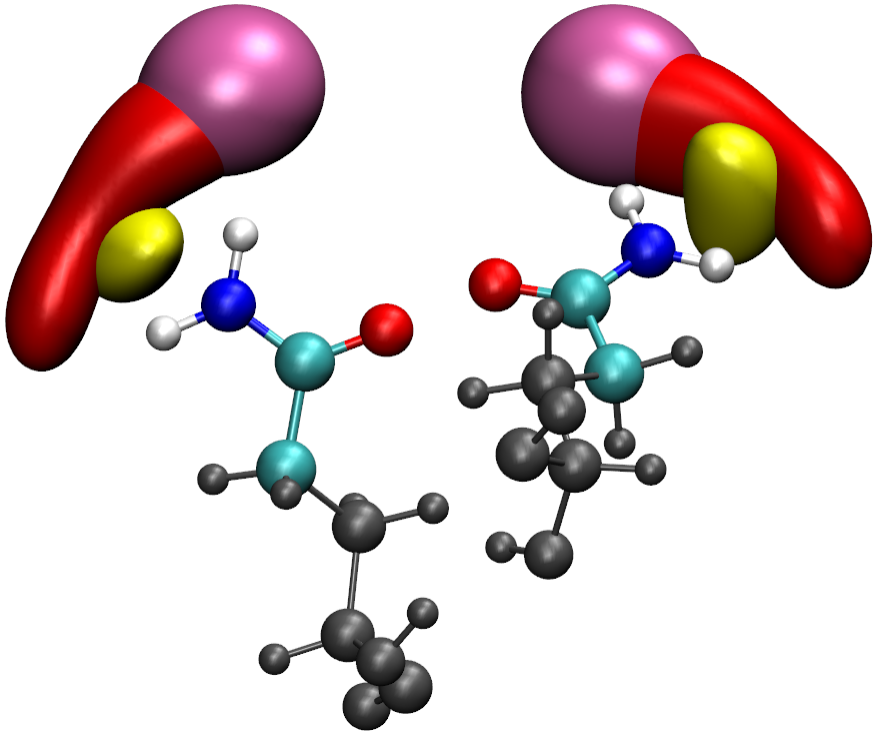

Figure 8: Glutamate: For all Hydration sites within range, 31% are in the hydrophobe category (yellow), 44% are in the moderate category (purple) and 11% are in the strongly bound category (red) with other sites falling between these.

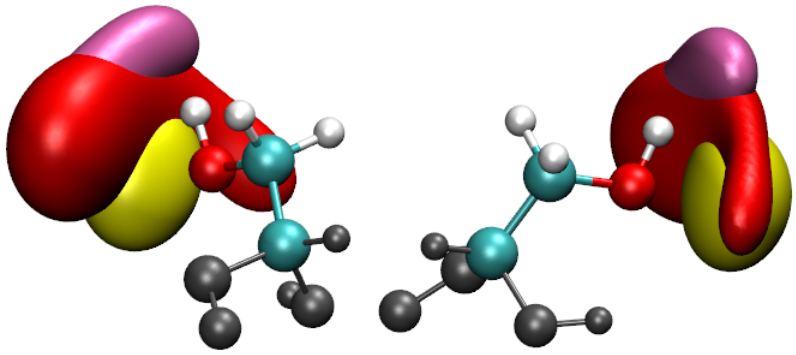

Figure 9: Serine: For all Hydration sites within range, 8.6% are in the hydrophobe category (yellow), 33% are in the moderate category (purple) and 17% are in the strongly bound category (red) with other sites falling between these.

Number of Heavy Atoms

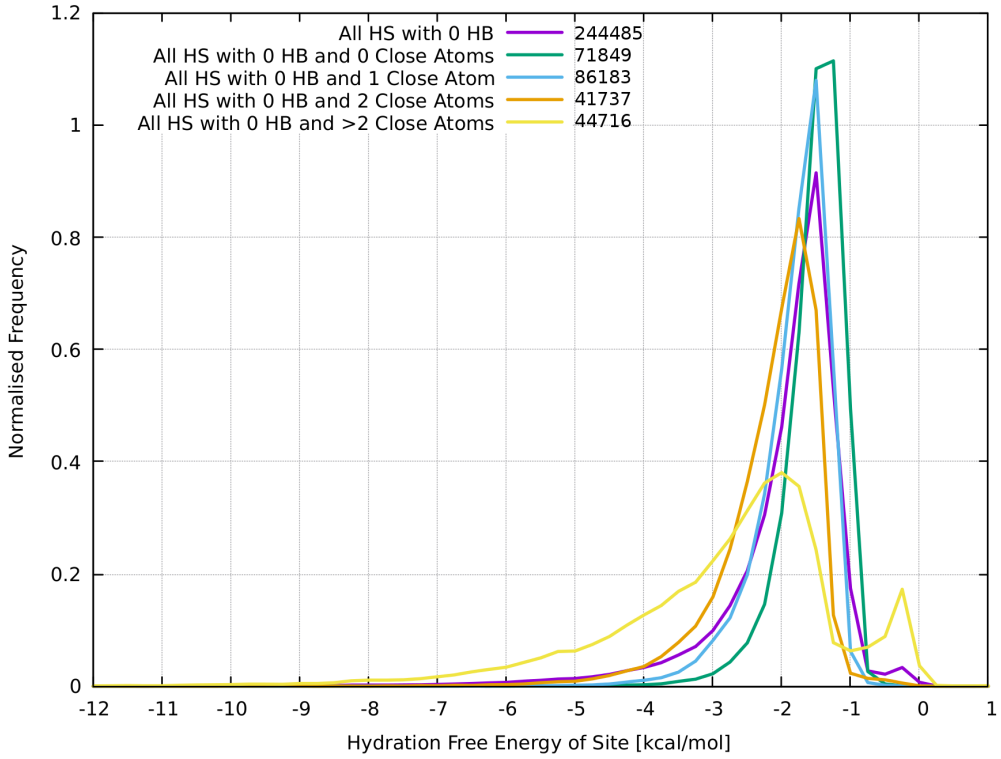

Figure 10: Histogram of the hydration free energy values for the sites with no hydrogen bonders in range, but varying numbers of nearby heavy atoms. The number of hydration sites in each category is displayed in the legend.

Figure 10 shows the change in the distribution of sites with no hydrogen bondable atoms nearby as the number of heavy atoms changes. The average behaviour for all these sites is a sharp distribution with a long but quickly decaying left tail. These sites are fairly displaceable with energies only ranging from -1 to -4 kcal/mol. The peak of the average distribution is closer to 0 than any of the oxygen or nitrogen distributions in figures 7 and 3. This implies that having a single hydrogen bondable atom nearby improves the stability of a water molecule in the site. Increasing the number of heavy atoms appears to strengthen the binding of water at a site. This can be seen in a leftward shift in the peak of the distributions in figure 10, along with a leftward broadening leading to very heavy tails and possibilities of sites with hydration free energies up to -7 kcal/mol in the case of more than two heavy

atoms. This energy is comparable to the distributions for amide, hydroxyl and backbone oxygen sites and amide nitrogen sites. However, larger numbers of heavy atoms can also destabilise the site, creating very weakly binding water. This feature is present as a small sharp peak near 0 kcal/mol on the distribution with more than two heavy atoms only. It should be noted that the scale on figure 10 is different to figures 3 and 7 and care should be taken when comparing the plots.

Carbon and Sulphur

Figure 11 shows that there is little difference in the free energy distribution of sites around different types of carbon and sulphur. This suggests that water will behave in a similar way in these sites, and carbon need not be partitioned into aliphatic and aromatic types when there are no potential hydrogen bonders nearby. The shape of the distribution is similar to that in figure 10 as this is essentially a further splitting of that dataset. There was not enough data to discern a difference between sulphur atoms on methionine and cysteine amino acids. All of these sites should be considered easy to displace.

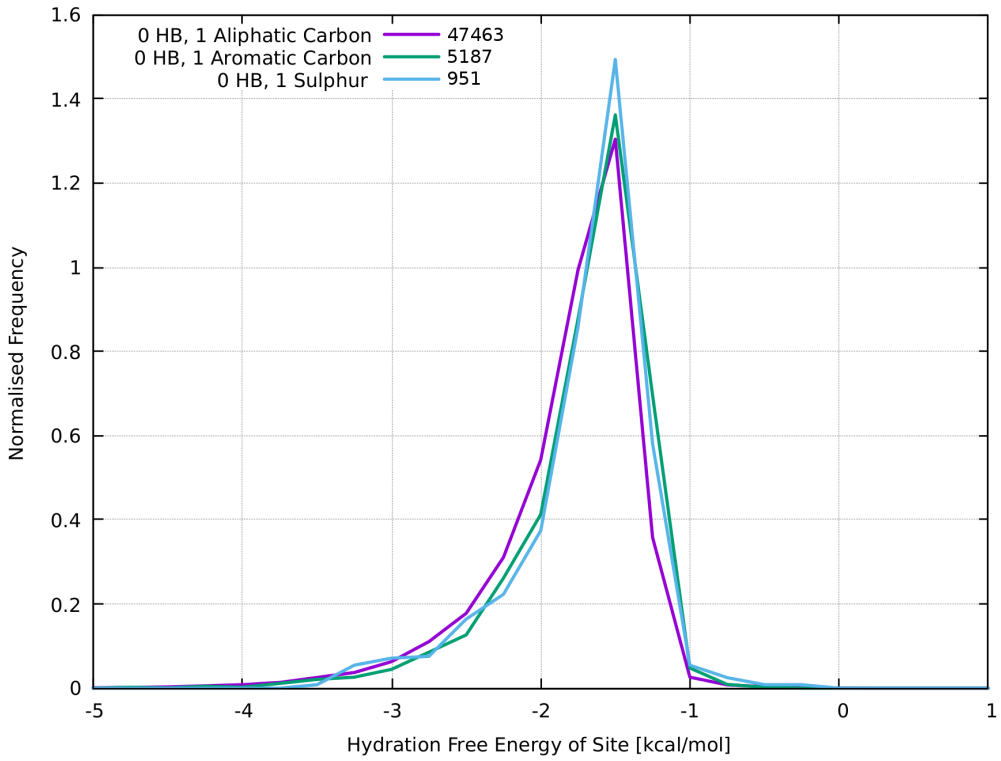

Figure 11: Histogram of the hydration free energy values for the sites near a single carbon or sulphur with no hydrogen bonders in range. There is little difference in the behaviour of water around these atoms. The number of hydration sites in each category is displayed in the legend.

Number of Hydrogen Bonding Atoms

Figure 12 shows that increasing the number of hydrogen bondable atoms around hydration sites leads to a rapid and extreme broadening of the distribution of hydration free energies. This plot includes both oxygen and nitrogen like sites and also includes mixed classes from table 3. Transitioning from zero to one hydrogen bond doubles the maximum expected free energy of the site from around -5 kcal/mol to -10 kcal/mol. Increasing from one to two hydrogen bonding atoms again adds -5 kcal/mol allowing sites with large -15 kcal/mol scores. However, the strongest effect when increasing the number of hydrogen bonding atoms is a very pronounced sharpening of distribution around strong hydrophobes with free energies between -1 and 0 kcal/mol. This implies that sites with many, possibly charged neighbours are unstable for water molecules. It is conceivable that these additional important degrees of

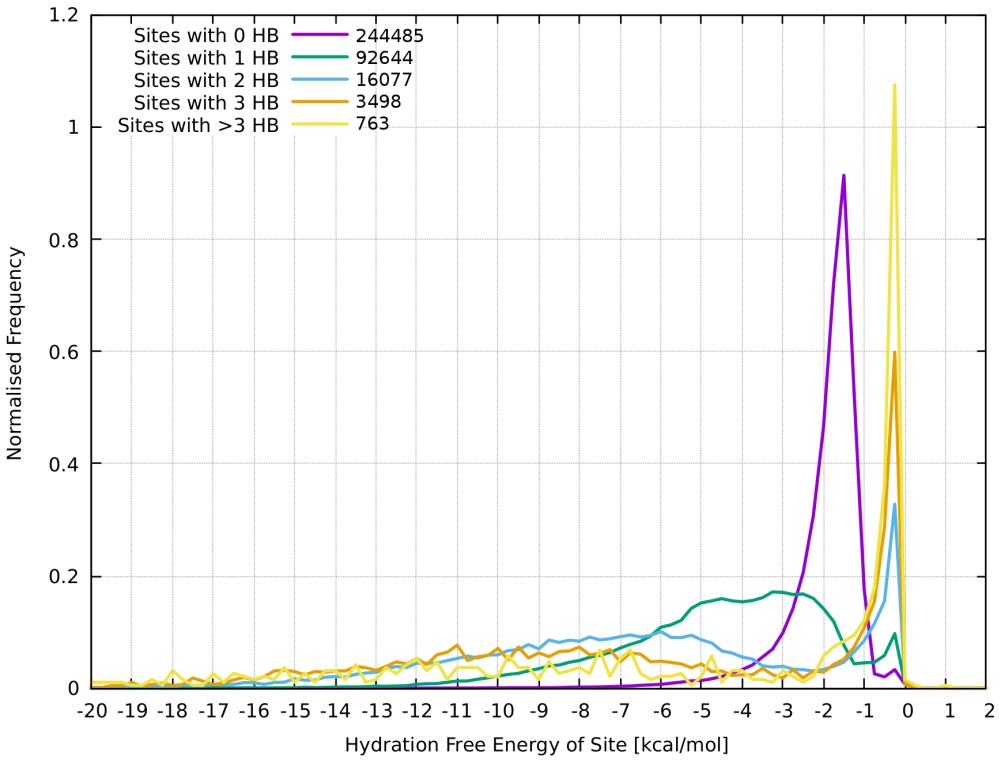

Figure 12: Histogram of the hydration free energy values for the sites near varying numbers of hydrogen bondable atoms. Adding more hydrogen bonders allows much strong binding of the water, but can also lead to destabilisation. The number of hydration sites in each category is displayed in the legend.

freedom do lead to unpredictable behaviour. There could be a dynamic effect in which water is drawn to the site from afar, and once reaching the site quickly becomes expelled. Such dynamic effects would not be well captured by these distributions. Due to a lack of data counts for sites with many hydrogen bonding atoms (as seen from table 3) the distributions are noisy. However there is a strong flattening of density for higher distributions at very negative free energies. There is a prominent dip around -3 kcal/mol such that HS with one hydrogen bonder are fairly common, but sites with two hydrogen bonders are fairly rare.

In order to investigate this destabilising effect further plots have been made with respect to an additional parameter: the number of hydrogen atoms within 4.0 Å of the site. This parameter will again correlate with the degree of protein embedding, with large numbers of hydrogen atoms corresponding to large numbers of neighbouring residues found deeper

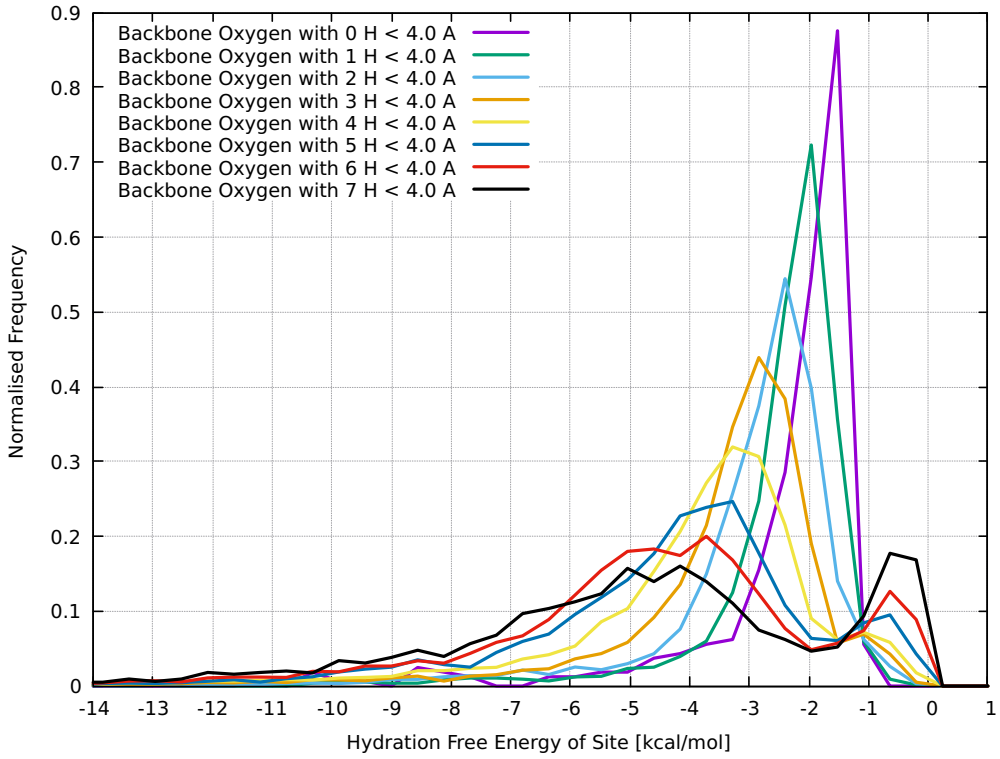

Figure 13: Histograms of the hydration free energy for sites near a backbone oxygen with varying numbers of hydrogen atoms within 4.0 Å. The curves show broadening with increased numbers of hydrogen.

within the protein.

Number of Hydrogen Atoms

Figures 13 and 14 show the behaviour of the hydration site free energy distribution as a function of the number of hydrogen atoms within 4.0 Å. There are additional plots for other chemical environments in the supplementary material. All of the plots broadly show the same behaviour. As the number of hydrogen atoms nearby is increased, the sites have a broader distribution of hydration free energies in which the peak of the distribution shifts leftwards and the left tail becomes increasingly heavy. Each additional hydrogen atom appears to shift the peak by approximately 0.5 kcal/mol. The backbone oxygen site distributions in figure 13 are very smooth as there were more data points in this class. The plots start from

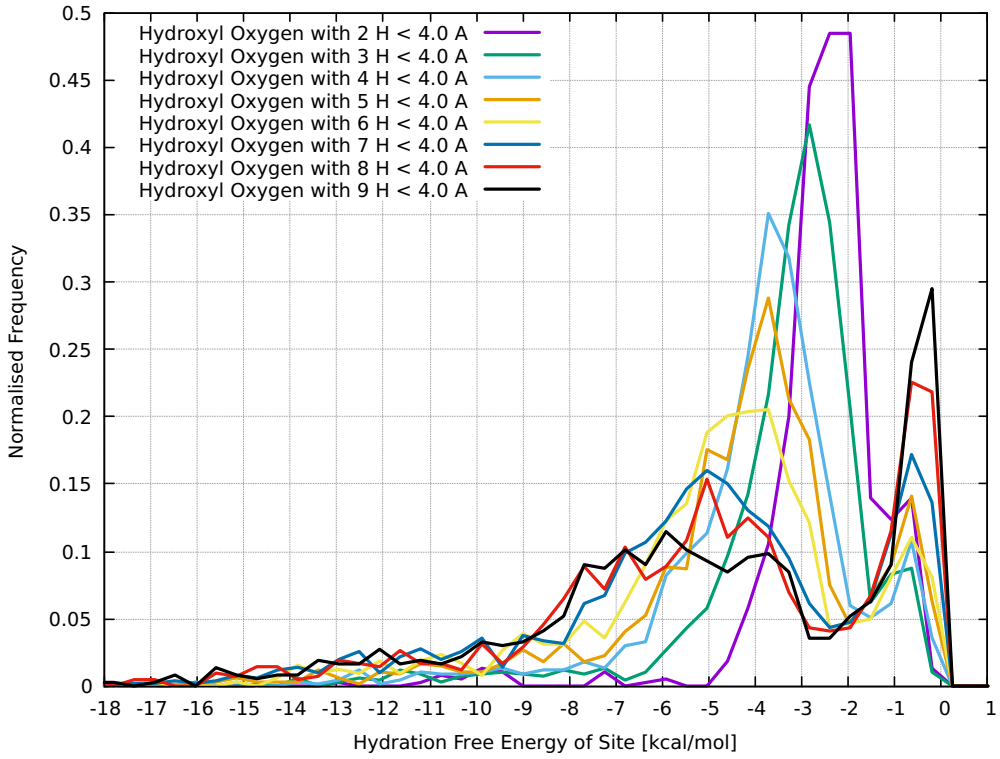

Figure 14: Histograms of the hydration free energy for sites near a hydroxyl oxygen with varying numbers of hydrogen atoms within 4.0 Å. The curves show broadening with increased numbers of hydrogen.

distributions for 0 hydrogen atoms up to 7, not all data sets start from 0 hydrogen atoms as there was very little data for either hydroxyl oxygen or arginine nitrogen like sites with 0 or 1 hydrogen atom. This is purely because the OH group or NH_2 groups always have hydrogen atoms nearby, whereas the backbone oxygen has a double bond. Once again these distributions all show an increase in extreme displacability with more neighbouring atoms, this effect appears reduced in the case of backbone oxygen in figure 13 but highly pronounced for hydroxyl sites in figure 14. This comparison is not entirely fair as the data for 9 hydrogen atom neighbours is missing and the trend continues for higher number of neighbours which is not shown here. For each of the plots there is a region between -1 and -3 kcal/mol which becomes decreasingly occupied for a larger number of neighbours.

Conclusions

The distribution of HS hydration free energies obtained from IFST show numerous insights into the behaviour of water at the surface of a protein depending on the local environment. In general, hydrogen bonding atoms create stability for local water molecules but having too many heavy atoms nearby causes either extreme stability or instability.

Although the proteins and hydration sites used in this work only represent a sample of potential hydration sites, the large number of proteins used should represent commonly seen sites well compared to previous studies. Bromodomains make up a large proportion of the dataset as shown in the tables in appendix A, and the distributions may have some bias towards their characteristics. This bias should be mitigated by the semi-random distribution of atom types around each residue on a fine scale. In conclusion, it is likely that these distributions could be used as a useful prior when inspecting hydration environments in unknown proteins.

Evidence of strong hydrophobes was seen throughout the analysis of the distributions. These appeared as large and sharp peaks at weakly binding values of hydration free energy. Some of these strong hydrophobes could be compared to the dehydrons of Fernandez et al. particularly those near the backbone nitrogen groups, which unexpectedly displayed significantly different behaviour to general amide nitrogen groups.

In previous studies of the statistics of HS, Beuming et al. concluded that hydration sites near the backbone of the protein are much less favourable than HS near polar side chains.[14] This could well relate to the presence of dehydrons due to unfavourable packing of hydrophobic groups near the protein backbone as described by Fernandez et al.[29,30] In this work we see that HS associated with backbone oxygen environments behaved no differently from hydroxyl and amide oxygens on polar side groups, however the behaviour of carboxylate oxygens was very different (figure 7). These carboxylate oxygens bound water around twice as strongly as other types, and based on the large frequency of occurrence across all proteins (27441 HS) which is similar to the occurrence of backbone oxygen groups (28943 HS), one

could indeed conclude that for oxygen environments the backbone HS are less favourable on average. For the less frequently occurring nitrogen based HS, the backbone nitrogen group free energy distributions are actually very similar to arginine and lysine environments (figure 3). These are all the strongest binding nitrogen hydration sites, and HS associated with histidine and amide residues are in general easier to displace. The data on nitrogen hydration sites would disagree with the statement that backbone hydration sites are much less favourable. In conclusion, it depends strongly on the local environment, but figures 3 and 7 are useful guides to separate these average behaviours based on the closest residue. The strength of binding near backbone sites versus solvent exposed residues will also relate to the average degree of burial of the HS. Sites near the protein backbone would be more buried on average for globular proteins of sizeable volume. If the density of hydrogen atoms near the residue was high, figures 13 and 14 would indicate that there is an increased chance of a site becoming either highly unstable or strongly bound. For increased degree of burial the chance of instability greatly exceeds the chance of strong binding as highlighted by figure 12, although such sites became progressively rarer with interactions involving more atoms. Beuming et al. also concluded that there is no direct relation between the energetics of a hydration site and the degree of burial the HS has within a protein.[14] We would consider that burial has the propensity to *either* extremely stabilise *or* extremely destabilise a hydration site with no middle ground and that there is a signal in the energetics for this phenomenon.

Density plots were made for three classes of hydration sites in each residue environment. These classes were strongly and weakly bound HS and the moderately bound HS for appropriate choices of the binding energies which depend on the residue. It was seen that hydroxyl groups on serine, threonine and tyrosine residues all showed the same patterning of weak/mid/strong binding hydration sites. This consisted of a weakly binding zone directly opposite the oxygen atom, a strongly binding patch at an optimal angle to the hydrogen from that oxygen and a mid range patch binding with an overextended angle. These patches are clear in figure 9 and can be seen in addition plots for threonine and tyrosine in the

supplementary material.

Hydration sites near asparagine nitrogens (figure 5) show displaceable water close to the nitrogen and between the hydrogen atoms which are presumably sites that straddle the two hydrogen atoms compromising a formation of a hydrogen bond with any of them. Similar behaviour is seen for glutamine resides (shown in the supplementary material). In both cases the strongest binding patch was opposite the neighbouring oxygen atom, which may have occurred from a mild distribution to any water networks made in this region. Glutamate (figure 8) shows strong hydrophobes above and below the carbon-oxygen plane. This mimics the dehydron like effect described in backbone nitrogen regions, however this effect is distinct to the carboxyl local environment. This behaviour is also observed in aspartate resides (shown in the supplementary material). The backbone nitrogen and carbon environments were also studied. The carbon environment shows this splitting for strong hydrophobes which could be the dehydron like sites. Even the midrange energies show some degree of splitting in this environment, whereas the strongest binding sites are those that avail a direct hydrogen bond with the oxygen and hydrogen, and possibly sweep in an angle.

For the backbone nitrogen environment, the hydration free energy distribution around the carbon atom was shown in figure 6. The hydrophobe and mid-range classes both have split distributions about the carbonyl group. These locations are the likely places to find dehydrons. The strongly bound HS wrap the backbone environment in a band from the oxygen round to the hydrogen passing on the outside of the dehydron pocket. This distribution could indicate a network of strongly co-bound water molecules, or a flexible range of locations for strongly bound water to be found. The hydration environment around the nitrogen atom in the backbone nitrogen environment is quite different (see supplementary material). The weakly bound hydrophobes have a distinct location that does not overlap other classes. This is also a likely position for dehydrons to be found and represents a angle unsuitable for bonding with oxygen, nitrogen or hydrogen atoms. Most hydration sites with moderate binding strengths are positioned either opposite the oxygen or the hydrogen pre-

sumably forming hydrogen bonds. The strongly bound region, which may again indicate a frozen network of more than one water molecule, favours the hydrogen and nitrogen.

Future Applications

Hydration sites calculated with IFST have been useful in finding large hydrophobic patches whose displacement correlates strongly with ligandability.[17,35] These first few applications have relied on the binding of either ligands or proteins whose interactions are dominated by the hydrophobic effect. For electrostatic interactions, some HS with relatively large hydration free energies could be displaced to form very favourable Coulombic connections with a ligand.[20] The algorithm of Vukovic et al. could be updated to not only find the most displacable cluster in terms of hydration free energy, but also the cluster which when displaced would offer the strongest binding. The data collected in this study may help create such an adaptation. Figures 3 and 7 tell us for example that a backbone nitrogen HS with a hydration free energy of -3 kcal/mol is relatively rare. Although this is harder to displace than an amide nitrogen HS with hydration free energy -2 kcal/mol, the amide nitrogen HS is much more common, and relative to the majority of backbone nitrogen HS, -3 kcal/mol is very displaceable, and this HS could be a good target, especially if a strong electrostatic connection could be made upon displacement. This inference can be repeated for carboxyl and amide oxygen HS; -2 kcal/mol is very rare for the carboxyl oxygen and very common for the amide oxygen.

Vukovic et al. have already presented a combinatoric searching strategy applied to the *absolute* hydration free energy of HS to find the most displaceable cluster containing a set number of HS. One could conceive an adaptation to the combinatoric search algorithm of Vukovic et al. which finds the cluster with the highest *relative* displaceablility.

Supporting Information Available

The following files are available free of charge.

- Filename: Supplementary Material containing additional plots and figures.

Acknowledgement

Benedict W. J. Irwin acknowledges the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science for funding under grant number EP/L015552/1. Work in David J. Huggins' lab was supported by the Medical Research Council (MRC) under Grant No. ML/L007266/1. This work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service, provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council. This work used ARCHER the UK national supercomputer as part of the eCSE03-03 grant. The authors acknowledge Arno Proeme from the EPCC support team for assistance porting the Solvaware code for IFST calculations to ARCHER.

References

(1) Wang, H.; Ben-Naim, A. A Possible Involvement of Solvent-Induced Interactions in Drug Design. *J. Med. Chem.* **1996**, *39*, 1531–1539.

(2) Ladbury, J. E. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and its Potential Application to Drug Design. *Chem. Biol.* **1996**, *3*, 973 – 980.

(3) Li, Z.; Lazaridis, T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *J. Am. Chem. Soc.* **2003**, *125*, 6636–6637.

(4) Li, Z.; Lazaridis, T. Thermodynamics of Buried Water Clusters at a Protein-Ligand Binding Interface. *J. Phys. Chem. B* **2006**, *110*, 1464–1475.

(5) Huang, N.; Shoichet, B. K. Exploiting Ordered Waters in Molecular Docking. *J. Med. Chem.* **2008**, *51*, 4862–4865.

(6) Olano, L. R.; Rick, S. W. Hydration Free Energies and Entropies for Water in Protein Interiors. *J. Am. Chem. Soc.* **2004**, *126*, 7991–8000.

(7) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of Water Molecules in Protein Binding Sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.

(8) Yu, H.; Rick, S. W. Free Energy, Entropy, and Enthalpy of a Water Molecule in Various Protein Environments. *J. Phys. Chem. B* **2010**, *114*, 11552–11560.

(9) England, J. L.; Pande, V. S. Charge, Hydrophobicity, and Confined Water: Putting Past Simulations into a Simple Theoretical Framework. *Biochem. Cell Biol.* **2010**, *88*, 359–369.

(10) 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*, 2769–2780.

(11) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci.* **2007**, *104*, 808–813.

(12) 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*, 2817–2831.

(13) 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*, 160–164.

(14) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic Analysis of Water Molecules at the Surface of Proteins and Applications to Binding Site Prediction and Characterization. *Proteins* **2011**, *80*, 871–883.

(15) 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*, 414–429.

(16) Gerogiokas, G.; Southey, M. W. Y.; Mazanetz, M. P.; Heifetz, A.; Bodkin, M.; Law, R. J.; Henchman, R. H.; Michel, J. Assessment of Hydration Thermodynamics at Protein Interfaces with Grid Cell Theory. *J. Phys. Chem. B* **2016**, *120*, 10442–10452.

(17) Vukovic, S.; Brennan, P. E.; Huggins, D. J. Exploring the Role of Water in Molecular Recognition: Predicting Protein Ligandability Using a Combinatorial Search of Surface Hydration Sites. *J. Phys. Condens. Matter* **2016**, *28*, 344007.

(18) Irwin, B. W. J.; Chau, P. L.; Vukovic, S.; ElGamacy, M.; Payne, M. C. Prediction of GABARAP Interaction with the GABA type A Receptor. *Proteins* **2018**, *86*, 1251–1264.

(19) Huggins, D. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophys. J.* **2015**, *108*, 928–936.

(20) Haider, K.; Huggins, D. J. Combining solvent thermodynamic profiles with functionality maps of the Hsp90 binding site to predict the displacement of water molecules. *J. Chem. Inf. Model.* **2013**, *53*, 2571–2586.

(21) Raman, P. E.; MacKerell, A. D. Rapid Estimation of Hydration Thermodynamics of Macromolecular Regions. *J. Chem. Phys.* **2013**, *139*, 055105.

(22) Huggins, D. J.; Marsh, M.; Payne, M. C. Thermodynamic Properties of Water Molecules at a Protein-Protein Interaction Surface. *J. of Chem. Theory Comp.* **2011**, *7*, 3514–3522.

(23) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531–3541.

(24) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. of Phys. Chem. B* **1998**, *102*, 3542–3550.

(25) Lazaridis, T. Solvent Reorganization Energy and Entropy in Hydrophobic Hydration. *J. Phys. Chem. B* **2000**, *104*, 4964–4979.

(26) Huggins, D. J. Estimating Translational and Orientational Entropies Using the k - Nearest Neighbors Algorithm. *J. Chem. Theory Comput.* **2014**, *10*, 3617–3625.

(27) Irwin, B. W. J.; Huggins, D. J. On the Accuracy of One-and Two-Particle Solvation Entropies. *J. Chem. Phys.* **2017**, *146*, 194111.

(28) Huggins, D. J.; Payne, M. C. Assessing the Accuracy of Inhomogeneous Fluid Solvation Theory in Predicting Hydration Free Energies of Simple Solutes. *J. Phys. Chem. B* **2013**, *117*, 8232–8244.

(29) Fernandez, A.; Scott, R. Dehydron: A Structurally Encoded Signal for Protein Interaction. *Biophys. J.* **2003**, *85*, 1914–1928.

(30) Fernandez, A.; Crespo, A. Protein Wrapping: a Molecular Marker for Association, Aggregation and Drug Design. *Chem. Soc. Rev.* **2008**, *37*, 2373–2382.

(31) Zhang, X.; Crespo, A.; Fernandez, A. Turning Promiscuous Kinase Inhibitors into Safer Drugs. *Trends Biotechnol.* **2008**, *26*, 295 – 301.

(32) Fernandez, A. Incomplete Protein Packing as a Selectivity Filter in Drug Design. *Structure* **2005**, *13*, 1829 – 1836.

(33) Crespo, A.; Fernandez, A. Kinase Packing Defects as Drug Targets. *Drug Discov. Today* **2007**, *12*, 917 – 923.

(34) Fraser, C. M.; Fernandez, A.; Scott, L. R. Dehydron Analysis: Quantifying the Effect of Hydrophobic Groups on the Strength and Stability of Hydrogen Bonds. *Adv. Exp. Med. Biol.* **2010**, *680*, 473–479.

(35) Vukovic, S.; Huggins, D. J. Quantitative Metrics for Drug-Target Ligandability. *Drug Discovery Today* **2018**, *23*, 1258–1266.

(36) Panagis, F.; Sarah, P.; Maria, M.; Tracy, K.; Jean-Philippe, L.; Dalia, B.-L.; Ildiko, F.; Rudolf, V.; Susanne, M.; et al., P. T. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012**, *149*, 214–231.

(37) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(38) 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*, 221–234.

(39) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(40) MacKerell A. D., et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(41) Mackerell Jr., A. D.; Feig, M.; Brooks III, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. of Comp. Chem. 25*, 1400–1415.

(42) Abascal, J. L.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(43) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 149901.

(44) Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8. **2015**,

(45) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. 1996.