Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein−Ligand Complexes

Yipin Lu, Renxiao Wang, Chao-Yie Yang, and Shaomeng Wang*

Departments of Internal Medicine, Pharmacology, and Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109

Received August 15, 2006

We have performed a comprehensive analysis of water molecules at the protein−ligand interfaces observed in 392 high-resolution crystal structures. There are a total of 1829 ligand-bound water molecules in these 392 complexes; 18% are surface water molecules, and 72% are interfacial water molecules. The number of ligand-bound water molecules in each complex structure ranges from 0 to 21 and has an average of 4.6. Of these interfacial water molecules, 76% are considered to be bridging water molecules, characterized by having polar interactions with both ligand and protein atoms. Among a number of factors that may influence the number of ligand-bound water molecules, the polar van der Waals (vdw) surface area of ligands has the highest Pearson linear correlation coefficient of 0.63. Our regression analysis predicted that one more ligand-bound water molecule is expected for every additional 24 Å^2 in the polar vdw surface area of the ligand. In contrast to the observation that the resolution is the primary factor influencing the number of water molecules in crystallographic models of proteins, we found that there is only a weak relationship between the number of ligand-bound water molecules and the resolution of the crystal structures. An analysis of the isotropic *B* factors of buried ligand-bound water molecules suggested that, when water molecules have fewer than two polar interactions with the protein−ligand complex, they are more mobile than protein atoms in the crystal structures; when they have more than three polar interactions, they are significantly less mobile than protein atoms.

INTRODUCTION

Protein−ligand interaction is fundamental in signal transduction in cells and is the basis for drug design. It is now commonly accepted that water molecules play an important role in protein−ligand binding. Water molecules can modify the shape and the flexibility of a ligand binding site in a protein, improving the complementarity between the protein and the ligand. Water molecules can act as hydrogen-bonding bridges between the protein and the ligand, establishing a network of hydrogen bonds which stabilize the protein−ligand interaction.[1−6] For example, there are few strong direct protein−ligand interactions between small peptides to oligopeptide binding protein (OppA). Rather, water molecules act as flexible adapters which bridge the protein−ligand interactions and shield charges on the buried ligand.[7,8] Ligand-bound water molecules can also govern ligand specificity and affinity, as can be seen in the crystal structures of the L-arabinose-binding protein with its ligand.[9] In drug design, attempts have been made to incorporate water molecules at the binding sites into ligands with a goal to improve binding affinities, as in the design of a class of HIV protease inhibitors.[10] Interestingly, however, the displacement of interface water molecules does not always lead to an increase in the binding affinity of the ligand.[11]

A number of recent molecular modeling studies have demonstrated the importance of interfacial water molecules. For example, when the empirical scoring function HINT[12] was used to estimate the free energy of binding in HIV-1 protease−ligand complexes, the correlation between HINT scores and the experimentally determined binding constants was significantly improved when one important interface bridging water molecule was included,[12] and in computational docking, it was shown that the inclusion of critical water molecules can lead to significant improvement in docking results.[13] Two recent computational studies have also clearly demonstrated that some but not all crystallographic water molecules in the ligand binding sites of proteins make significant contributions to the protein−ligand binding free energies.[14,15] Taken together, these studies[12−17] have clearly shown that certain interfacial water molecules play a significant role for protein−ligand binding and should be considered in drug design.

In this paper, we describe a comprehensive analysis of water molecules at ligand binding sites observed in high-resolution X-ray crystal structures of protein−ligand complexes. The goals of this work were to determine the number of water molecules bound to the ligand in different protein−ligand complexes and the factors that may influence the water binding. We also explored the propensities of different types of ligand atoms and protein residues to be hydrated at the protein−ligand interfaces. As we were preparing this manuscript, a study on the hydration of the protein−protein interface was published.[18] We have therefore compared the properties of water molecules at the protein−ligand interface with those at the protein−protein interface.

* Corresponding author tel.: (734) 615-0362; fax: (734) 647-9647; e-mail: shaomeng@umich.edu.

10.1021/ci6003527 CCC: $37.00
Published on Web 02/01/2007

Table 1. The Definitions and Codes for Different Types of Ligand Atoms

code	definition	code	definition
C.3	carbon sp^3	O.2	oxygen sp^2
C.2	carbon sp^2	O.co2	oxygen negatively charged, in carboxylate and phosphate groups
C.ar	carbon aromatic	S.3	sulfur sp^3
C.cat	carbocation (C+), used only in a guadinium group	S.2	sulfur sp^2
N.3	nitrogen sp^3	S.O	sulfoxide sulfur
N.2	nitrogen sp^2	S.O2	sulfone sulfur
N.ar	nitrogen aromatic	P.3	phosphorus sp^3
N.am	nitrogen amide	F	fluorine
N.4	nitrogen positively charged	Cl	chlorine
O.3	oxygen sp^3		

METHODS

A total of 392 protein−ligand complexes' structures were obtained from the PDBbind database,[20] a collection of protein−small molecule ligand complexes with experimentally measured binding affinity data and crystal structures,[21] and used in the present study. We selected from PDBbind the crystal structures that contain the atomic coordinates for water oxygen atoms and that were determined at room temperature. Since the accurate pictures of hydration in protein crystals are derived from high-resolution studies and the reliability of the X-ray-determined water molecules decreases as the resolution decreases, only crystal structures with a resolution ≤ 2.0 Å were considered in this study. Care should be taken when using the crystallographic structures because some of the apparent water sites may actually be disordered protein atom positions, other solvent molecules, or noise and phase errors.[22,23] The original crystal structure data for each complex was obtained from the Protein Data Bank (PDB),[21] and the PDB codes for these 392 crystal structures are provided in the Supporting Information. These 392 structures cover a wide range of protein−ligand binding sites. The number of non-hydrogen ligand atoms ranges from 6 to 68. The binding sites of the proteins exhibit different shapes, including both deep, buried pockets and large, open binding sites. In this study, we were interested in water molecules that are directly involved in the protein−ligand interaction, and accordingly, the binding site was defined here as including all protein residues located within 8 Å of a ligand atom. A water molecule was considered to have a polar interaction with a protein/ligand atom if the distance between the oxygen atom of the water and the polar atom (N, O, or S) of the protein or ligand was less than 3.60 Å.

***B* Factors.** The *B* factors reported in the PDB entries reflect the thermal motion of each atom according to the relationship $B = 8\pi\bar{U}^2$, where $\bar{U}$ is the mean displacement of the atom. The Cartesian coordinates of an atom define its most probable position, and the *B* factor describes the magnitude of the oscillation of the atom around this position. The *B* factor is therefore a useful indicator of the mobility of water molecules observed in a crystal structure. *B* factors, however, vary, not only because of actual physical mobility but also because of the refinement strategies used.[24,25] In order to make the *B* factors from different crystal structures comparable, those within each complex structure were normalized to have a distribution of zero mean and unit variance. This normalization was carried out using the equation

$$B_{\text{norm}} = (B - \bar{B})/\text{sd}(B)$$

where $\bar{B}$ and sd(*B*) are, respectively, the mean value and the standard deviation of the distribution of *B* factors within each crystal structure. This normalization permits the comparison of the mobility of water molecules from different crystal structures that have been solved using different refinement protocols. Another advantage of normalization of the *B* factor is that it allows comparison of the mobility of water molecules with that of protein atoms. A water molecule with average mobility of the protein atoms will have a B_{norm} value of 0. A positive B_{norm} value indicates that this water molecule has above-average mobility compared to protein atoms, while a negative B_{norm} value indicates below-average mobility.

Calculation of Solvent-Accessible Surface Area and van der Waals Surface Area. A program was written to calculate the solvent-accessible surface area (SASA) and the van der Waals (vdw) surface area (VSA) of ligand, protein, and water molecules in a complex. The SASA of a molecule, which measures its accessibility to the external bulk aqueous environment, was defined by rolling a probe of a given size around the vdw surface and following the coordinates of the center of the probe.[26] The radius of the rolling probe was set to 1.4 Å. A modified Voronoi tessellation method[27−29] was used to calculate surface areas. This method calculates the contact surfaces following projection of the polyhedra edges on the extended sphere radius (van der Waals radii plus solvent atom radii, in this study, 1.4 Å). The solvent-accessible surface area is obtained by subtraction of the atomic contact surfaces from the sphere surface. The polar SASA (or VSA) of the ligand was calculated as the sum of SASA (or VSA) for all polar ligand atoms in the complex. The percentage of surface area that is accessible to the solvent (SASA%) is calculated as $\text{SASA\%} = \text{SASA}/4\pi(R_{\text{vdw}} + R_{\text{p}})^2$, where R_{vdw} is the VDW radius of the atom and R_{p} is the radius of the probe (1.4 Å).

Propensity for Hydration at the Protein−Ligand Interface. Each interface water molecule was assigned to the closest surface-exposed non-hydrogen ligand atom and closest surface-exposed non-hydrogen protein atom, which are treated as the primary hydrated ligand atom and protein atom for each interface water molecule, respectively. An atom-typing code similar to SYBYL atom typing was used for the ligand atoms (Table 1).

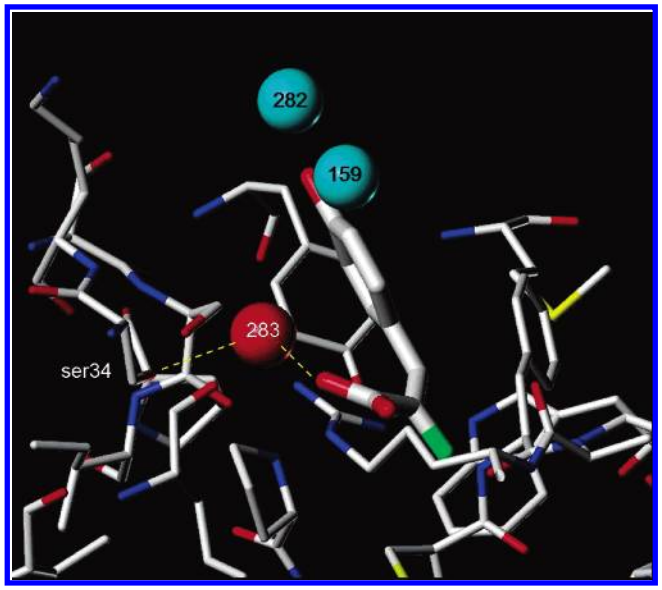

Figure 1. Binding site (thin sticks) of 4-oxalocrotonate tautomerase homologue (PDB code: 1gyy) with its ligand 2-fluoro-3-(4-hydroxyphenyl)-2E-propenoate (thick sticks) and crystallographically determined water molecules (spheres). The surface water molecules are shown in cyan, and interface water molecules are shown in red. The interface water molecule HOH 283 forms hydrogen bonds (yellow lines) with SER34 in the active site and the carboxylate group in the ligand.

The hydration propensity for each type of ligand atom at the protein−ligand interface was calculated as

$$\text{propensity of X-type ligand atom} = (n_X/n_{Total})/(m_X/m_{Total})$$

While X is the type of the ligand atom, n_X is the number of the primary hydrated ligand atoms with type X, and n_{Total} is the total number of all types of primary hydrated ligand atoms; m_X is the total number of the X-type ligand atoms, and m_{Total} is the total number of all types of ligand atoms. The propensity of each type of protein residue to be hydrated primarily by interface water molecules was calculated similarly.

Statistical Analysis. A statistical analysis was performed using SAS statistical software version 8.2 (SAS Institute, Cary, NC). An ordinary least-squares method was used to fit the regression model predicting the number of ligand-bound water molecules. One-way analysis of variance (ANOVA) was used to test the equality of the means of normalized *B* factors.

RESULTS AND DISCUSSION

Number of Water Molecules at the Ligand Binding Site. The 392 crystal structures in the data set contain a total of 92 022 water molecules. A water molecule is considered to be ligand-bound if it is within 3.60 Å of any of the ligand atoms. This 3.60 Å threshold will include water molecules involved in van der Waals interactions or hydrogen bonds to the ligand atoms.[30] There are a total of 1829 such ligand-bound water molecules, and they can be classified into two groups: the first is of water molecules that have direct contact with the ligand, but not with the protein (for example, HOH 282 and HOH 159 in Figure 1). Water molecules in this group are commonly found at the interface between the ligand and the solvent and are, therefore, defined as "surface water molecules". The second group consists of water molecules having direct contact with both the protein and the ligand (for example, HOH 283 in Figure 1). Such water molecules are located at the interface between the protein and the ligand; they mediate the protein−ligand interactions and are defined as "interfacial water molecules". Of these 1829 ligand-bound water molecules, 330 (18%) are surface water molecules and 1499 (72%) are interfacial water molecules. Table 2 shows the distribution of the number of ligand-bound water molecules in the 392 protein−ligand complexes. This number ranges from 0 to 21 and has an average of 4.6.

Table 2. The Distribution of the Number of Ligand-Bound Water Molecules in Each Complex Structure Based on the 392 Crystal Structures of Protein−Ligand Complexes

number of ligand-bound water molecules	count	percent (%)	cumulative count	cumulative percent (%)
0	21	5	21	5
1	43	11	64	16
2	57	15	121	31
3	47	12	168	43
4	56	14	224	57
5	35	9	259	66
6	38	10	297	76
7	22	6	319	81
8	19	5	338	86
9	16	4	354	90
10	10	3	364	93
11∼15	20	5	384	98
16∼21	8	2	392	100

Table 3. The Relationship between the Number of Ligand-Bound Water Molecules and Different Variables

variables	Pearson linear correlation coefficient
polar vdw surface area of the ligand	0.63
overall vdw surface area of the ligand	0.25
polar SASA of the ligand	0.45
overall SASA of the ligand	0.17
polar vdw surface area of the protein active site	0.43
number of polar ligand atoms	0.58
number of polar protein atoms at the binding site	0.39
resolution of crystal structure	−0.16

Since the number of the ligand-bound water molecules in a protein−ligand complex has such a wide range, it is instructive to examine the factors that may influence this number. The factors that have been examined in this study are listed in Table 3.

It has been reported[19] that the resolution is the primary factor influencing the number of water molecules included in crystallographic models of proteins. We found, however, that there is only a weak relationship between the number of ligand-bound water molecules (N_{ligHOH}) and the resolution (Table 3); both the number of ligand-bound water molecules ($r = -0.16$) and the number of bound water molecules per ligand atom ($r = -0.15$) show no significant correlation with the resolution.

Of all the factors examined, the polar vdw surface area of ligands has the highest Pearson linear correlation coefficient with N_{ligHOH} (0.63), followed by the number of polar ligand atoms (0.58), polar SASA of the ligand (0.45), polar vdw surface of the protein binding site (0.43), number of polar

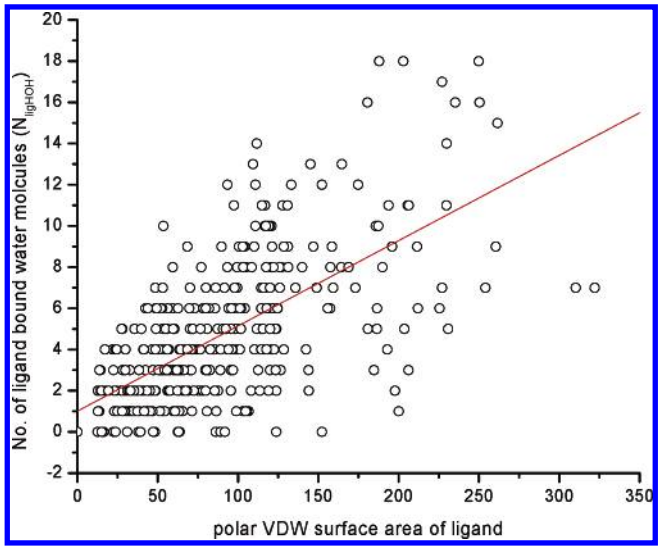

Figure 2. Plot of the polar vdw surface area of the ligand vs the number of ligand-bound water molecules. $R = 0.63$.

protein atoms at the binding site (0.39), overall vdw surface of the ligand (0.25), overall SASA of the ligand (0.17), and resolution (−0.16).

Although only a weak dependence was reported[18] between the number of water molecules in the protein crystal structures and the fraction of polar/apolar surface of the protein,[19] our analysis showed that the polar vdw surface area of ligands has the highest correlation (Figure 2 and Table 3). As can be seen in Figure 2, the number of water molecules binding to the ligand (N_{ligHOH}) depends strongly ($r = 0.63$) on the polar vdw surface area of the ligand (PVSAL). When an ordinary linear regression model was used, the following relationship between N_{ligHOH} and PVSAL was obtained:

$$N_{ligHOH} = 1.01 + 0.041\text{PVSAL}$$

Both the intercept (1.01) and slope (0.041) are statistically significantly different from 0.0 ($p < 0.0001$). According to this relationship, one more water molecule will be expected to bind to the ligand for every additional 24 Å^2 in PVSAL. The number of polar ligand atoms is proportional to PVSAL, and a similar correlation between N_{ligHOH} and the number of polar ligand atoms was obtained with $r = 0.58$. N_{ligHOH} also has a moderate correlation with the polar vdw surface area of the binding site ($r = 0.46$) and has a weaker correlation with the SASA than with the vdw surface area. Although N_{ligHOH} has a strong correlation with the polar vdw surface area of the ligand, it has only a modest correlation ($r = 0.33$) with the overall vdw surface area of the ligand. This suggests that a ligand with a large surface area but little or no polar surface would have very few ligand-bound water molecules observable by crystallography.

Water Molecules at the Protein−Ligand Interfaces. The polar interactions that each ligand-bound water molecule has with the protein−ligand complex were analyzed. On average, each interfacial water molecule has three polar interactions with the complex, while each surface water molecule has one polar interaction with the complex (Figure 3). It was found that in the protein−protein interfaces each water molecule has, on average, only one polar interaction with the complex,[18] similar to the surface waters in protein−ligand

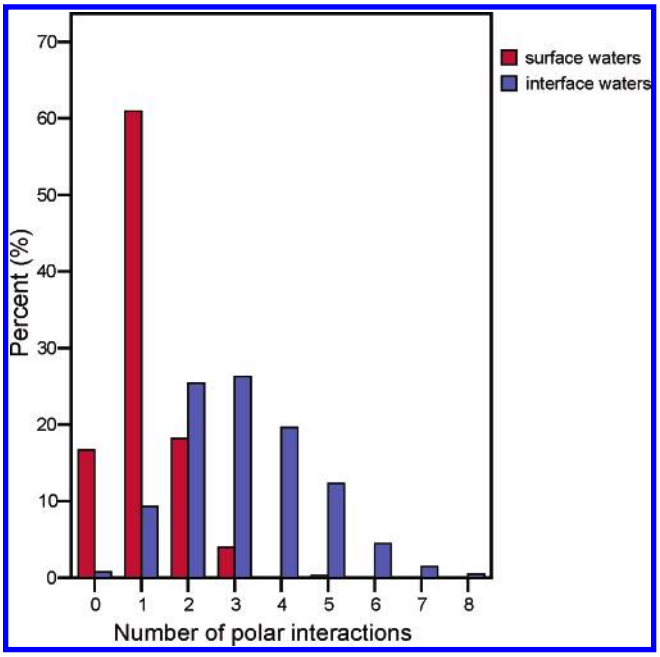

Figure 3. Distribution of the number of polar interactions with the complex for surface waters (red) and interface waters (blue) at the ligand-binding sites.

complexes. The fact that the water molecules at the protein−ligand interface have more polar interactions than those at the protein−protein interfaces may explain the discrepancy that resolution has a significant influence on the number of observed protein−protein interface waters[18] but only a weak influence on the number of ligand-bound water molecules. Since each ligand-bound water, on average, has three polar interactions with the complex, this will limit the motion of the water through directional hydrogen bonds. In contrast, each water at the protein−protein interface on average has only one polar interaction with the complex,[18] and such water molecules may not be ordered sufficiently to be resolved in electron density maps, particularly those determined at low resolution.[31]

Interestingly, interfacial water molecules make more polar interactions with the protein than with the ligand. Of the total polar interactions between the interfacial waters and the protein−ligand complex, 67% are with protein atoms and 33% are with ligand atoms. Protein−ligand interfacial water molecules have a mean SASA of 10.4 ± 12.2 Å^2, much smaller than that for surface water molecules (38.8 ± 14.6 Å^2). Interfacial water molecules have a mean B factor of 26.2 ± 12.9 and B_{norm} of 0.44 ± 1.16, significantly lower than the B factor (38.5 ± 15.0 Å^2) and B_{norm} (1.60 ± 1.08) for surface water molecules.

Of the 1499 interfacial water molecules, only 11 (0.7%) are in a hydrophobic environment where water molecules have no polar interactions with either protein atoms or ligand atoms. A previous study found that, for waters buried in monomeric proteins, the percentage of water molecules that make no polar contact to protein atoms is 3%.[32] Thus, our analysis suggests that waters located in a hydrophobic environment are even less frequently observed at the protein−ligand interface than in the protein interior. Interestingly, the B_{norm} for these 11 interface waters with no polar contacts is 1.70 ± 1.06, higher than the B_{norm} for all of the other interface water molecules (0.42 ± 1.15), indicating that

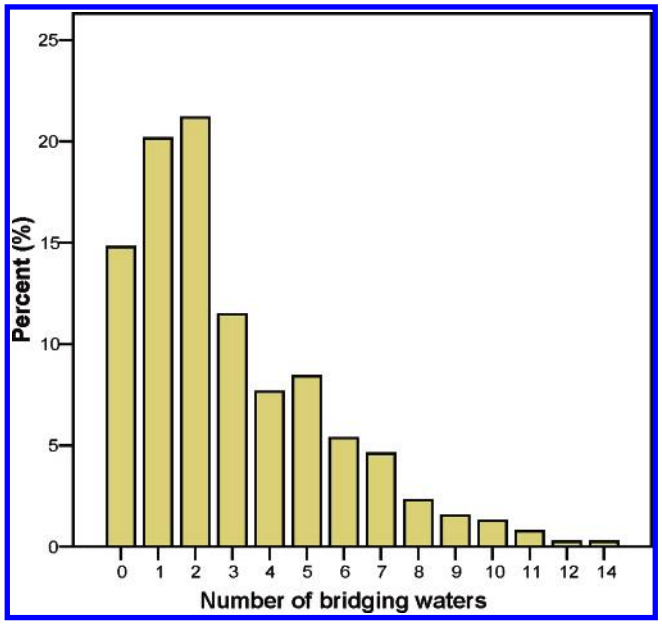

Figure 4. Distribution of the number of waters that bridge the protein−ligand polar interactions in 392 crystal structures of protein−ligand complexes.

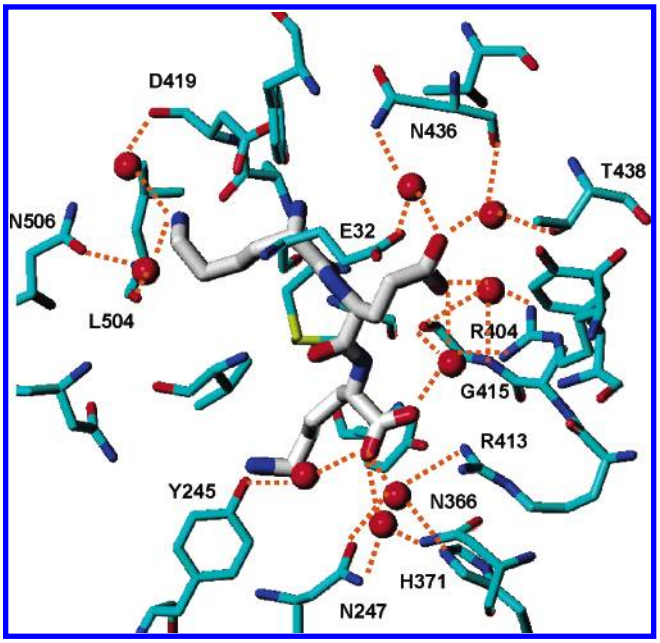

Figure 5. Bridging water molecules (red spheres) at the binding site of the periplasmic oligopeptide binding protein (OppA, PDB code: 1jeu) (thin capped bond) with its ligand KEK (thick capped bond). The hydrogen bonds formed between the interface water molecules and the complex are shown as orange dashed lines.

interfacial water molecules that make no polar contacts with either protein or ligand atoms are more mobile than those water molecules with polar interactions.

Interfacial waters can bridge protein−ligand interaction by having polar interactions with both the protein and the ligand. We found that such bridging water molecules account for 77% of the interfacial water molecules in our data set. In comparison, the corresponding fraction of such bridging waters at the protein−protein interfaces was much lower, only about 30%.[18] Figure 4 shows the distribution of the number of bridging water molecules observed in each complex structure. The number ranges from zero to 14, and on average, three bridging water molecules are found in each protein−ligand complex. Over 85% of the protein−ligand complex structures have one or more bridging water molecules present at the interface of the protein and the ligand. Complexes that have eight or more bridging water molecules (6.4% of the data set) usually have large polar ligands, such as a tripeptide or a nucleotide. One example is periplasmic OppA, a protein found in Gram-negative bacteria, which acts as a receptor for peptide transport across the cell membrane and is a potential target for antibacterial drug design. Figure 5 shows the binding of the tripeptide Lys−Glu−Lys (KEK) to OppA (PDB code: 1jeu), which has nine water molecules.[33] The ligand is completely buried in the protein interior but has few direct interactions with the protein. Instead, most of the hydrogen bonds and electrostatic interactions involving KEK are established through the nine water molecules, which serve as bridges in the interaction between the ligand KEK and OppA protein residues. These nine bridging water molecules are all well-defined and have low *B* factors (13.59 ± 5.6, $B_{norm} = -0.32 \pm 0.4$).

Mobility of the Interfacial Water Molecules. The *B* factor is a good indicator of the positional uncertainty of each water molecule observed in the crystal structures. The value of the *B* factor, however, also depends on the refinement strategies used. Consequently, as explained in the Methods section, the normalized *B* factor was used to make the *B* factor values from different crystal structures comparable. We also focused our analysis on interfacial water molecules with a SASA percentage $\leq 5\%$, which are essentially buried water molecules. Overall, there are 759 buried water molecules in the data set. Table 4 shows the overall relationship between the number of polar interactions and normalized *B* factors. As can be seen from this table, as the number of polar interactions increases, the water molecules become more stable and more localized. The water molecules that only have one or two polar interactions have a mean normalized *B* factor of 0.96 ± 1.37 and 0.41 ± 1.24, respectively, indicating that, on average, they are more mobile than the protein atoms. As the number of polar interactions increases, the mean normalized *B* factor decreases rapidly, reaching −0.03 for waters with three polar interactions and −0.24 for waters with four polar interactions. This indicates that interfacial water molecules with three or more polar interactions have below-average mobility compared with the protein atoms on average. When the number of polar interactions of the waters is larger than four, however, the mean normalized *B* factor becomes stable (Figure 6). The one-way ANOVA test indicates that there is no significant difference in the mean normalized *B* factor values for groups of water molecules with between four and seven polar interactions ($p = 0.53$). The reason for this is

Table 4. Frequency, Mean, and Standard Deviation of the Normalized *B* Factor (B_{norm}) for Buried Interface Water Molecules, Grouped by the Number of Polar Interactions with the Complex

number of polar interactions	frequency	percent (%)	mean of the B_{norm}	standard deviation of the B_{norm}
1	19	2.5	0.96	1.37
2	96	13	0.41	1.24
3	186	25	−0.03	0.81
4	204	27	−0.24	0.80
5	159	21	−0.12	0.95
6	65	8.7	−0.13	0.73
7	22	2.9	−0.18	0.88

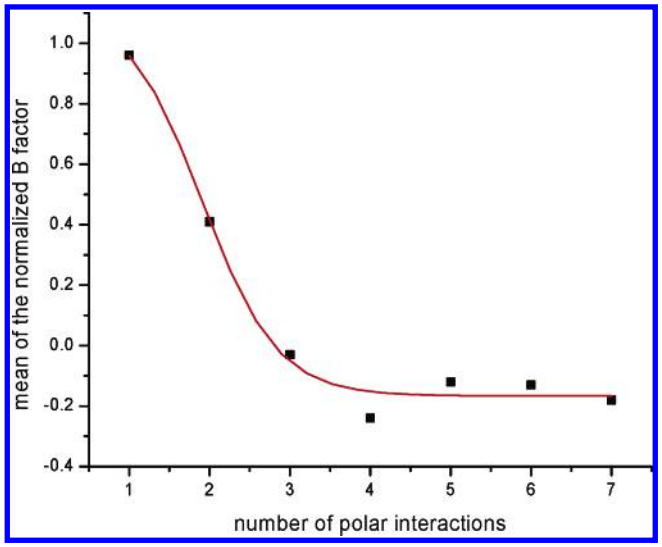

Figure 6. Plot of the mean of the normalized *B* factor vs the number of polar interactions between water molecules and the complex.

that water molecules can form hydrogen bonds with polar groups at the protein−ligand interface, thus becoming immobilized. In general, a single water molecule can form at most four hydrogen bonds, and even when there are more than four polar groups present in the immediate environment, one water molecule can only have at most four hydrogen-bond interactions. This explains why the water molecules with more than four polar interactions show a stability behavior similar to that of water molecules that have only four polar interactions.

The Propensity of Hydration at the Protein−Ligand Interface. We next analyzed the hydration propensities for each ligand atom type by interfacial water molecules (Figure 7). There are in total 17 ligand atom types (Table 1) occurring more than 15 times in these 392 crystal structures. The order, from highest to lowest, of hydration propensities of these at the protein−ligand interfaces is

N.4, O.co2 > O.2, O.3, N.3 > N.ar, N.2, N.am, F, Cl, S.3 > C.2, C.ar, C.3 > P.3, C.cat, S.O2

Positive-charged nitrogen atoms (N.4) and negative-charged oxygen atoms (O.co2) have the highest propensity for hydration at the interface and are followed by two types of neutral oxygen atoms and N.3 atoms. The other three types of nitrogen atoms, N.ar, N.2, and N.am, have hydration propensities lower than that of N.3. In N.ar and N.am, the nitrogen lone pair is involved in the orbital overlap with the π bond. Such orbital overlap reduces the ability of nitrogen, the hydrogen-bond donor, to interact with the water molecules. Interestingly, two halogen atoms, chloride and fluorine, have similar hydration propensities, although debate continues[33] as to whether the fluorine atom engages in hydrogen bonds. As expected, the propensities of carbon and phosphorus are among the lowest, consistent with their hydrophobic nature.

For proteins, 39% of the interfacial waters' primary protein hydration sites are on the protein backbone and 61% on the protein side chains. The hydration propensities of each amino acid side chain at the protein−ligand interfaces were calculated (Figure 8). The propensities vary from 1.59 for arginine to 0.02 for proline. The four charged side chains, glutamic acid, arginine, aspartic acid, and lysine, exhibit the highest hydration propensity at the interfaces. Asparagine and glutamine have a slightly lower propensity but are comparable to their parent acids. Tyrosine, serine, and threonine are three residues that have hydroxyl groups at their side chains. We found that tyrosine in the protein−ligand interface has a hydration propensity of 1.20, significantly higher than the propensity of serine (0.82) and close to the propensities of histidine and asparagine. Threonine has one more methyl group than serine, which gives it a

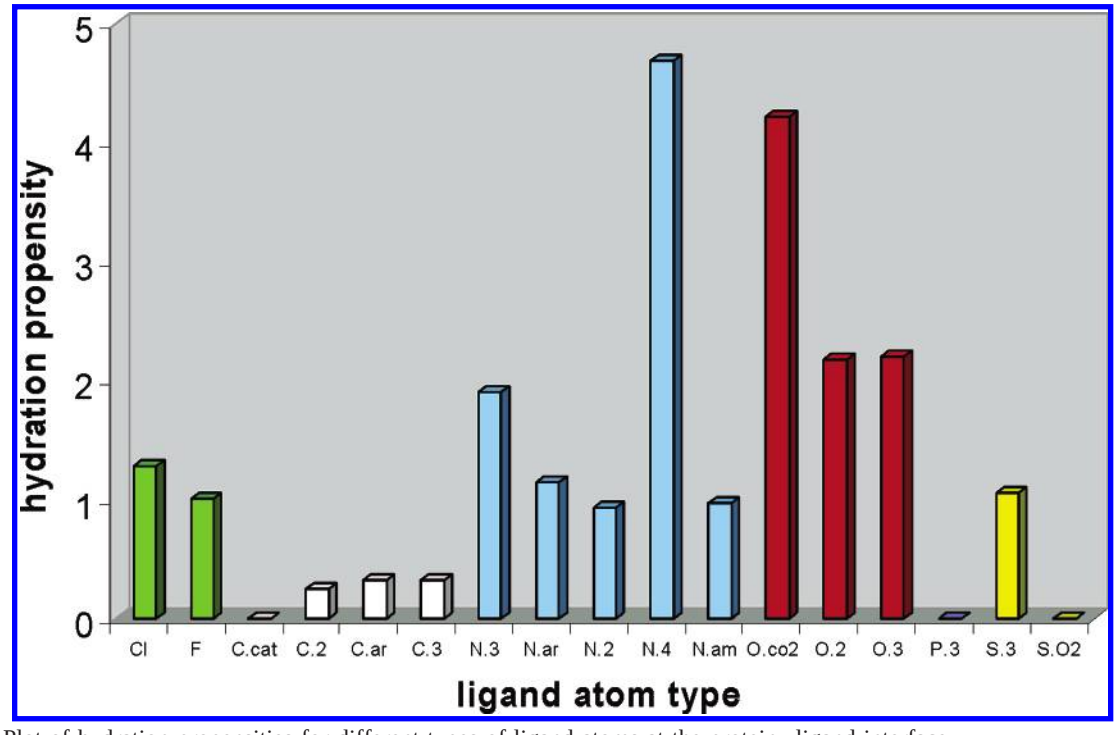

Figure 7. Plot of hydration propensities for different types of ligand atoms at the protein−ligand interface.

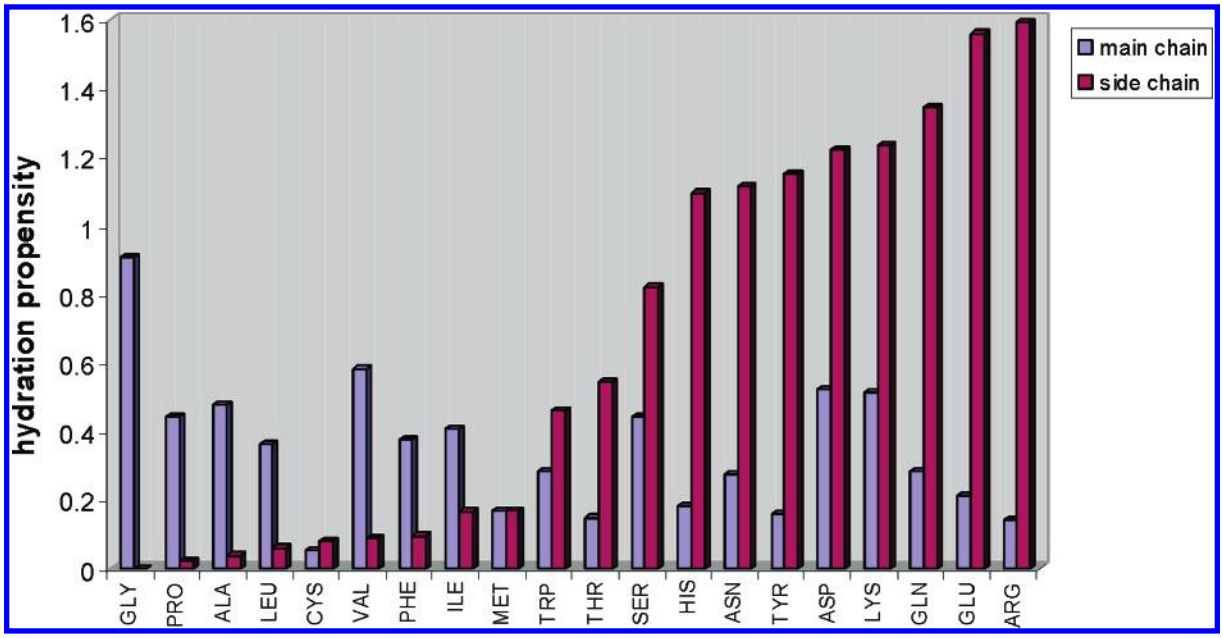

Figure 8. Plot of hydration propensities for amino acid main chains (blue) and side chains (brown) at the protein−ligand interface.

little more nonpolar character. Consistent with this, the hydration propensity of the interface threonine side chain (0.54) is smaller than the one of serine (0.82). Cysteine is often classified as polar because of the presence of the S in the sulfhydryl group SH, which can behave as either a hydrogen-bond donor or hydrogen-bond acceptor. However, the hydration propensity of the interface cysteine side chain is only 0.07, which is significantly lower than that of other polar side chains and, instead, is close to those of hydrophobic side chains (Figure 8). As expected, the hydration propensities of hydrophobic side chains are very low, all below 0.2 except in the case of tryptophan. Although tryptophan has an aromatic ring structure which was thought to be very water-insoluble and so very hydrophobic, an interfacial tryptophan side chain has a hydration propensity of 0.45, which is significantly higher than those of other hydrophobic side chains and close to the propensity of threonine (0.54). This is probably due to the presence of the indole nitrogen, which affords some polar character to the tryptophan, allowing it to have hydrogen-bond interactions with the waters.

Among the water−protein main-chain interactions, 72% of them are with main-chain carbonyl oxygens and 26% of them are with main-chain NH groups. The hydration propensity values for the main chains of different amino acids were calculated (Figure 8), and glycine was found to have the highest main-chain hydration propensity (0.91), probably because of the associated accessibility of the main chain. The propensities of other amino acid main chains are all below 0.6.

SUMMARY

We have performed a comprehensive analysis of water molecules at the protein−ligand interfaces observed in 392 high-resolution protein−ligand complexes' structures obtained from the PDBbind database. We found that there are a total of 1829 ligand-bound water molecules in these 392 complexes, with 18% as surface water molecules and 72% as interfacial water molecules. The number of ligand-bound water molecules in each complex structure ranges from 0 to 21 and has an average of 4.6. Of these interfacial water molecules, 76% are considered as bridging water molecules, characterized by having polar interactions with both ligand and protein atoms.

Among all of the factors analyzed that may influence the number of ligand-bound water molecules, the polar vdw surface area of ligands has the highest Pearson linear correlation coefficient, at 0.63. Our regression analysis predicted that one more ligand-bound water molecule is expected for every additional 24 Å^2 in the polar vdw surface area of the ligand. In contrast to the observation that the resolution is the primary factor influencing the number of water molecules in crystallographic models of proteins, we found that there is only a weak relationship between the number of ligand-bound water molecules and the resolution of the crystal structures.

Analysis of the isotropic *B* factors of ligand-bound water molecules suggested that, when water molecules have fewer than two polar interactions with the protein−ligand complex, they are more mobile than protein atoms in the crystal structures, and when they have more than three polar interactions, they are significantly less mobile than protein atoms.

Our analysis of the hydration propensities for each ligand atom type also showed that charged atoms have the highest propensities to be hydrated by ligand-bound water molecules in the crystal structures, while neutral and hydrophobic atoms have the least hydration propensities. For protein residue side chains, arginine has the highest hydration propensity, while proline has the lowest. For protein main-chain atoms, 72% of the water−protein main-chain interactions are with the carbonyl oxygens, while 26% are with NH groups. Glycine has the highest main-chain hydration propensity among all the residues.

ACKNOWLEDGMENT

We thank Dr. George W. A. Milne for his critical reading of the manuscript and his editorial assistance.

Supporting Information Available: Information on the crystal structures used in this study in Table S1. This information is available free of charge via the Internet at http://pubs.acs.org

REFERENCES AND NOTES

(1) Mancera, R. L. De novo Ligand Design with Explicit Water Molecules: An Application to Bacterial Neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 479−499.

(2) Wilson, I. A.; Fremont, D. H. Structural Analysis of MHC Class I Molecules with Bound Peptide Antigens. *Semin. Immunol*. **1993**, *5*, 75−80.

(3) Rejto, P. A.; Verkhivker, G. M. Mean Field Analysis of FKBP12 Complexes with FK506 and Rapamycin: Implications for a Role of Crystallographic Water Molecules in Molecular Recoginition and Specificity. *Proteins* **1997**, *28*, 313−324.

(4) Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, M. J.; Air, G. M.; Luo, M. Novel Aromatic Inhibitors of Influenza Virus Neuraminidase Make Selective Interactions with Conserved Residues and Water Molecules in the Active Site. *J. Mol. Biol.* **1999**, *293*, 1107−1119.

(5) Palomer, A.; Perez, J. J.; Navea, S.; Llorens, O.; Pascual, J.; Garcia, L.; Mauleon, D. Modeling Cyclooxygenase Inhibition. Implication of Active Site Hydration on the Selectivity of Ketoprofen Analogues. *J. Med. Chem.* **2000**, *43*, 2280−2284.

(6) Huang, K.; Lu, W.; Anderson, S.; Laskowski, M.; James, M. N. G. Water Molecules Participate in Proteinase−Inhibitor Interactions: Crystal Structure of Leu18, Ala18, and Gly18 Variants of Turkey Ovomucoid Inhibitor Third Domain Complexed with *Streptomyces griseus* Proteinase B. *Protein Sci*. **1995**, *4*, 1985−1997.

(7) Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R. H. Crystallographic and Calorimetric Analysis of Peptide Binding to OppA Protein. *J. Mol. Biol.* **1999**, *291*, 393−415.

(8) Tame, J. R.; Murshudov, G. N.; Dodson, E. J.; Neil, T. K.; Dodson, G. G.; Higgins, C. F.; Wilkinson, A. J. The Structural Basis of Sequence-Independent Peptide Binding by OppA Protein. *Science* **1994**, *264*, 1578−1581.

(9) Quiocho, F. A.; Wilson, D. K.; Vyas, N. K. Substrate Specificity and Affinity of a Protein Modulated by Bound Water Molecules. *Nature* **1989**, *340*, 404−407.

(10) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science* **1994**, *263*, 380−384.

(11) Mikol, V.; Papageorgiou, C.; Borer, X. The Role of Water Molecules in the Structure-Based Design of (5-Hydroxynorvaline)-2-cyclosporin: Synthesis, Biological Activity, and Crystallographic Analysis with Cyclophilin A. *J. Med. Chem.* **1995**, *38*, 3361−3367.

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

(13) Rarey, M.; Kramer, B.; Lengaure, T. The Particle Concept: Placing Discrete Water Molecules during Protein−Ligand Docking Predictions. *Proteins* **1999**, *34*, 17−28.

(14) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.

(15) Lu, Y.; Yang, C.-Y.; Wang, S. Binding Free Energy Contributions of Interfacial Waters in HIV-1 Protease/Inhibitor Complexes. *J. Am. Chem. Soc.* **2006**, *128*, 11830−11839.

(16) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 1. Multiple Hydrogen-Bonding Features of Water Molecules in Mediating Protein−Ligand Interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500−512.

(17) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 2. Influence of Local Site Surface Shape on Water Binding. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 513−520.

(18) Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of Protein−Protein Interfaces. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 36−45.

(19) Carugo, O.; Bordo, D. How Many Water Molecules Can Be Detected by Protein Crystallography? *Acta Crystallogr., Sect. D* **1999**, *55*, 479−483.

(20) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.

(21) 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 Res.* **2000**, *28*, 235−242.

(22) Dixon, M. M.; Matthews, B. W. Is γ-chymoTrypsin a Tetrapeptide Acyl-enzyme Adduct of a Chymotrypsin? *Biochemistry* **1989**, *28*, 7033−7038.

(23) Ghosh, D.; O'Donnell, S.; Furey, W., Jr.; Robbins, A. H.; Stout, C. D. Iron−Sulfur Clusters and Protein Structure of Azotobacter Ferredoxin at 2.0 Å Resolution. *J. Mol. Biol.* **1982**, *158*, 73−109.

(24) Tronrude, D. E. Knowledge-Based *B*-Factor Restraints for the Refinement of Proteins. *J. Appl. Crystallogr.* **1996**, *29*, 100−104.

(25) Frauenfelder, H.; Petsko, G. A. Structural Dynamics of Liganded Myoglobin. *Biophys. J.* **1980**, *32*, 465−478.

(26) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.

(27) Voronoi, G. Nouveles Applications des Parameters Continues a la Theorie des Formes Quadratique. *J. Reine Angew. Math.* **1908**, *134*, 198−287.

(28) McConkey, B. J.; Sobolev, V.; Edelman, M. Quantification of Protein Surfaces, Volumes and Atom−Atom Contacts Using a Constrained Voronoi Procedure. *Bioinformatics* **2002**, *18*, 1365−1373.

(29) Preparata, F. P.; Shamos, M. I. *Computational Geometry. An Introduction*; Springer: New York, 1985.

(30) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting Conserved Water-Mediated and Polar Ligand Interaction in Proteins using a K-Nearest-Neighbors Genetic Algorithm. *J. Mol. Biol.* **1997**, *265*, 445−464.

(31) Levitt, M.; Park, B. H. Water: Now You See It, Now You Don't. *Structure* **1993**, *1*, 223−226.

(32) Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. Buried Waters and Internal Cavities in Monomeric Proteins. *Protein Sci.* **1994**, *3*, 1224−1235.

(33) Tame, J. R. H.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E. The Role of Water in Sequence Independent Ligand Binding by an Oligopeptide Transporter Protein. *Nat. Struct. Biol.* **1996**, *3*, 998−1001.

(34) Biffinger, J. C.; Kim, H. W.; DiMagno, S. G. The Polar Hydrophobicity of Fluorinated Compounds. *ChemBioChem* **2004**, *5*, 622−627.

(35) Dunitz J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **1994**, *264*, 670−671.

(36) 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.

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

CI6003527