Robust Classification of "Relevant" Water Molecules in Putative Protein Binding Sites

Alessio Amadasi,[†] J. Andrew Surface,[§] Francesca Spyrakis,[†,¶] Pietro Cozzini,[‡,¶] Andrea Mozzarelli,*[,†,¶] and Glen E. Kellogg*[,§]

Department of Biochemistry and Molecular Biology, University of Parma, Via G.P. Usberti 23/A, 43100 Parma, Italy, Department of General and Inorganic Chemistry, University of Parma, Via G.P. Usberti 17/A, 43100, Parma, Italy, National Institute for Biostructures and Biosystems, Viale Medaglie d'Oro 305, 00136 Roma, Italy, and Department of Medicinal Chemistry and Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond, Virginia, 23298-0540

Received August 16, 2007

A statistically validated protocol to identify "relevant" water molecules in protein binding sites using HINT score and a geometric descriptor termed Rank is described. In training, conservation/nonconservation was modeled for 86% of the waters. For the test set, 87% of waters were correctly classified (92% when crystallographic resolution was $\leq$2.0 Å). Conserved waters make at least two hydrogen bonds with protein and gain 0.6−2.0 kcal mol^{-1} more binding energy than nonconserved waters.

Introduction

Protein–ligand docking and structure-based drug design exploit structural knowledge of a protein target to predict the binding properties of new compounds. Success in these endeavors requires simultaneously solving several difficult problems, among which evaluating and understanding the role of water molecules in the binding site are particularly relevant.[1–8] Klebe has shown that in about two-thirds of protein–ligand complexes at least one water molecule is involved in the binding.[9] Wang confirmed the importance of interfacial water molecules in 392 crystal structures.[10] All binding site waters may affect binding events. Strongly bound or "conserved" waters, i.e., those that are consistently observed in several crystallographic structures of the same protein, are not easily displaced by ligands and thus affect binding by modifying the shape of the protein surface recognized by the ligand or by mediating that interaction with hydrogen bonds. At the other end of the scale, waters only weakly interacting with the protein can contribute to the hydrophobic effect as they are displaced from nonpolar regions, or occupy voids at the interface, thus possibly contributing entropic terms to the overall energetics. While schemes to predict water conservation/displacement between uncomplexed and complexed structures[11,12] or among complexed structures with different ligands have been reported,[13] not all are validated over a diverse, well-characterized data set and none have been widely accepted.

However, water molecules are not universally conserved; i.e., any water can be displaced by a ligand designed for that purpose, such as displacement of the bound water 301 of HIV-1 protease[14,15] with the cyclic urea class of inhibitors. This water, seen in all uncomplexed and most structures complexed with ligands, was specifically targeted for displacement by a ligand feature that mimicked its hydrogen bonding pattern. This released an ordered water with a favorable gain in entropy and increased binding affinity.[16] Inhibition specificity of all aspartic protease ligands (including HIV-1 protease inhibitors) is largely controlled by displacement of a structural catalytic water at the active site, with the interesting exception of a β-secretase complex where that water bridges protein and ligand.[17] In effect, if a ligand chemical group is able to compensate for the (enthalpic) loss of hydrogen bonds between a water and protein by formation of new ligand−protein hydrogen bonds, it should be able to displace that water.

We suggest the classification "relevant" water molecules, i.e., water molecules endowed with structural and energetic features such that they are generally conserved but may be displaced through design/synthesis of ligands incorporating polar groups that reproduce that water's hydrogen bonds. These waters should be explicitly considered in docking and other molecular modeling experiments such as structure-based drug design because their "rational" retention/displacement may be desirable depending on the situation.[16,18] While a study describing automatic handling of important water molecules during GOLD[19] docking runs has recently appeared,[20] robust methods for their identification were not described. This is the primary aim of the present report.

Results and Discussion

Our protocol relies on two main tools: the HINT free energy scoring model,[21] validated for a wide variety of biomolecular systems,[3,22–26] and the Rank algorithm that calculates the number and geometric quality of potential hydrogen bonds for each water molecule (to non-water atoms) in a protein structure.[27] HINT gives an estimate of the global interaction strength between each water molecule and its surrounding (protein) atoms, with respect to the relative chemical properties of donors and acceptors as well as their state of charge and accessibility, by evaluating the hydrophobic−polar properties of the environment surrounding that water molecule. Rank evaluates possible donor and acceptor matches for each water, yielding values from 0 for waters that do not form any hydrogen bonds with non-water molecules to about 6 for waters forming four quality hydrogen bonds with excellent bond length and angle geometry. Both tools have strengths and weaknesses regarding our goal of predicting which waters in an active site will be relevant with respect to ligand design for that site. Thus, we applied a pseudo-Bayesian statistical analysis to integrate the information provided by Rank and HINT score. First, we developed a statistical model on a training set of 13 proteins (with 125

* To whom correspondence should be addressed. For A.M.: phone, +39 0521 905138; fax, +39 0521 905151; e-mail, andrea.mozzarelli@unipr.it. For G.E.K.: phone, +01-804-828-6452; fax, +01-804-827-3664; e-mail, glen.kellogg@vcu.edu.

[†] Department of Biochemistry and Molecular Biology, University of Parma.

[§] Virginia Commonwealth University.

[¶] National Institute for Biostructures and Biosystems.

[‡] Department of General and Inorganic Chemistry, University of Parma.

10.1021/jm701023h CCC: $40.75
Published on Web 01/31/2008

Table 1. Protein Crystallographic Structures in the Absence and Presence of Ligands, in the Training Set

protein	PDB uncomplexed/PDB complexed (resolution, Å)	no. waters[a]	excluded waters[b,c]
carboxypeptidase A	5cpa (1.95)/6cpa (2.00), 7cpa (2.00)	11	W314,[b] W574,[b] W313[b]
concanavalin A	2ctv (1.95)/5cna (2.00)	6	
endothiapepsin	4ape (2.10)/1ent (1.90), 1epp (1.90)	21	
periplasmic glucose/ galactose receptor	1gcg (1.90)/2gbp (1.90), 2hph (1.33)	11	
HIV-1 protease	1g6l (1.90)/4phv (2.10), 1hxw (1.80)	10	W270[b]
lipid binding protein	1lib (1.70)/1lid (1.60), 1lie (1.60)	5	
major urinary protein I	1i04 (2.00)/1i05 (2.00), 1i06 (1.90)	2	
penicillopepsin	3app (1.80)/1 ppm (1.70), 1ppk (1.80)	17	W111[c]
phosphodiesterase 4B	1f0j (1.77)/1xlx (2.19), 1xm6 (1.92)	16	W173,[b] W741[b]
retinoic acid binding protein II	1xca (2.30)/1cbs (1.80), 2cbs (2.10)	7	
β-secretase	1w50 (1.75)/1tqf (1.80), 2irz (1.80)	16	
trypsin	1tpo (1.70)/1tnh (1.80), 1tnl (1.90)	4	
thrombin	1jou (1.80)/1a4w (1.80), 2c8w (1.96)	6	

[a] Count of waters located in the binding pocket. [b] Excluded after GRID analysis suggested water may be anomalous. [c] Excluded because of minimal conformational change in that region.

Table 2. Protein Crystallographic Structures in the Absence and Presence of Ligands, in the Test Set

protein	PDB uncomplexed/PDB complexed (resolution, Å)	no. waters[a]	excluded waters[b,c]
acetylcholinesterase	1ea5 (1.80)/2ack (2.40), 2c5g (1.95)	11	W624
cholesterol oxidase	3cox (1.80)/1coy (1.80)	13	
cyclophilin A	1ist (1.90)/1bck (1.80), 1cwf (1.86)	7	W126[c]
dihydrofolate reductase	1ai9 (1.85)/1aoe (1.60), 1ia1 (1.70)	4	
FKBP12	1fkk (2.20)/1fkl (1.70), 1j4 h (1.80)	11	
neuraminidase	2ht5 (2.40)/2ht8 (2.40), 2htq (2.20)	5	W25[c]
retinol binding protein II	1opa (1.90)/1opb (1.90)	4	
ribonuclease A	1fs3 (1.40)/1u1b (2.00), 1afk (1.70)	13	W60[c]
thymidine kinase	1e2 h (1.90)/1e2l (2.40), 1ki2 (2.20)	7	W14,[c] W46,[c] W92[b]

[a] Count of waters located in the binding pocket. [b] Excluded after GRID analysis suggested water may be anomalous. [c] Excluded because of minimal conformational change in that region.

discrete water molecules) with different structural and functional properties (Table 1). We then tested this model on an independent set of 9 proteins (Table 2) with 68 waters. Each protein in the training and test sets had well-characterized crystallographic structures for the unliganded and liganded forms. Proteins where specific waters are known to play a role in mediating the interaction between protein and ligand were preferred. Other selection criteria for creating the data sets were applied, including consideration of crystallographic uncertainties with respect to resolution[28–30] and validation of water positions with GRID[31] (see Experimental Section).

These uncomplexed and complexed structures of the same protein were superimposed to investigate the role of water molecules on ligand binding. The analysis focused on waters within 4 Å of the protein and ligand in the complex. For calibration all solvent molecules were first manually classified as relevant or nonrelevant (Figure 1) by careful examination of the overlapped structures. Relevant waters were (a) "conserved" in all the structures, i.e., the distance between its location in the unliganded and liganded structures is ≤1.2 Å,[12,22] or (b) displaced only by polar groups able to replace all or nearly all of their hydrogen bonds. Note that waters displaced by polar groups without forming substitute hydrogen bonds with the protein were considered the same as being displaced simply as a consequence of steric factors. Nonrelevant waters were those displaced sterically, located in external, highly solvent exposed regions of the binding site or simply missing in some structures. Rank and HINT score values relative to the interaction with the uncomplexed protein were then calculated for all water molecules in the training and test sets.

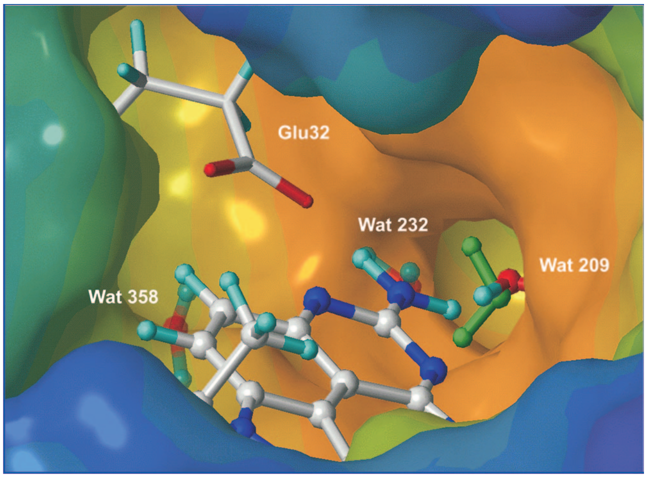

Figure 1. Different roles for water molecules found in the active site of dihydrofolate reductase (PDB code 1AI9) after the binding of the inhibitor 1,3-diamino-7-(1-ethylpropyl)-7*H*-pyrrolo[3,2-*f*]quinazoline (GW345)[35] (rendered in balls and sticks, PDB code 1AOE). Sterically (wat 358) and functionally (wat 232) displaced waters are displayed with transparent rendering. Wat 209 is conserved (green) in the complex structure. Some of the protein surface has been removed to show Glu32.

The mathematical system used to describe water behavior, depending on Rank and Score, was built heuristically on the training set. The dependences of water conservation on Ranks and HINT scores (parts a and b of Figure 2) were analyzed using the following nonlinear polynomial regressions that describe the relationships between Rank (eq 1) or HINT score (eq 2) and the percentage of conserved waters:

$$P_{\mathrm{R}} = -2.446R^2 + 33.746R - 4.107 \qquad (r^2 = 0.96) \qquad (1)$$

$$P_H = (-1.000 \times 10^{-4})H^2 + 0.187H + 21.621 \qquad (r^2 = 0.97) \qquad (2)$$

where R is the Rank value and H is the HINT score. P_R should be considered the percent probability for conservation based on Rank and P_H the probability based on HINT score.

Next, within the training set, the distributions of P_R and P_H for the relevant waters were compared (through subtraction) to those distributions for all waters. The dashed lines of Figure 2c (blue for P_R and red for P_H) represent the results of this subtraction after scaling with respect to the highest absolute value obtained (i.e., corresponding to P_R of 30%). This distribution difference (a sort of background subtraction) indicates how much a given P_R or P_H is diagnostic for a relevant water compared to a random water, i.e., the confidence (or weight) that should be accorded to conservation probabilities calculated from the Rank or HINT score polynomials of eqs 1 and 2. Differences near zero indicate that it is not possible to distinguish between relevant and nonrelevant waters at that P_R or P_H, while positive or negative differences are indicative of conservation or nonconservation, respectively. Thus, the highest absolute difference value corresponds to the probability in which we can be most confident. Weightings for P_R (solid blue line in Figure 2c) and P_H (solid red line in Figure 2c) were calculated on the basis of these results, using nonlinear regressions (see Supporting Information). Consider a case with P_R of 30% and P_H of 70%: the weightings in Figure 2c show that P_R is much more reliable in indicating nonconservation than the P_H is in indicating conservation; i.e., the water should be predicted as nonrelevant.

Because some probability regions are sparsely populated, this subtraction of distributions methodology generates results inconsistent with chemical meaning in the extreme regions of the graphs (90% < P_R < 30% and 80% < P_H < 40%). Waters with very high or very low Rank (hydrogen bonds) and/or HINT score (binding free energies) are unsurprisingly not very frequent. The distribution model (Figure 2c) in those regions illustrates the low probability of having a randomly taken (and relevant) water with these values but does not indicate a lower degree of safety of P_R and P_H in those regions; i.e., there are no reasons to believe that a water with four hydrogen bonds to the protein would have a lower probability of being conserved than a water with three hydrogen bonds. Chemically, the probability is at least the same. Thus, the weight assigned in our model to P_R and P_H falling in these regions is the same as at the closest maximum point in the graphs. Another poorly sampled region, with similar issues, is present in the Rank curve for P_R ranging from 75 to 85 (Figure 2c). This arises because the Rank algorithm, designed to evaluate the geometric features of a water molecule's potential hydrogen bonds, disallows nonrealistic bond angles, thus generating some underrepresented regions within the range of Rank values.[27] Because of the low statistical and chemical meaning of this underrepresented region, P_R values between 75 and 85 were excluded from the regressions performed to obtain the weightings.

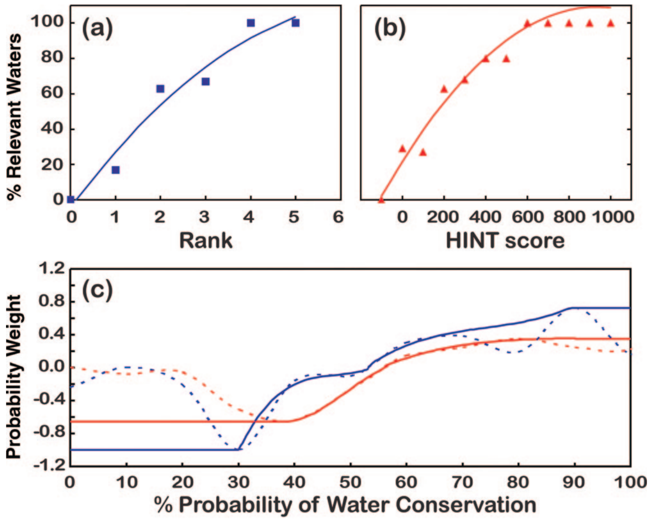

Figure 2. Fraction of relevant water molecules vs (a) Rank and (b) HINT score. (c) Probability distributions for weighting of P_R (blue) and P_H (red), after background subtraction and scaling. Solid lines are math functions derived from nonlinear regressions used to determine weights. Dashed lines are smoothed raw data.

The training set results were then integrated in the following weighted probability equation:

$$P_A = \frac{P_R(|W_R| + 1)^2 + P_H(|W_H| + 1)^2}{(|W_R| + 1)^2 + (|W_H| + 1)^2} \qquad (3)$$

where P_A is the overall probability of the entire system and W_R and W_H are the weights of Rank and HINT score probabilities, respectively, as shown in Figure 2c. The weighting coefficients are squared to further differentiate the weights in exponential space. If P_A is 50% or greater, the water is considered relevant. On the entire training set the model is able to correctly predict the ultimate role for 108 (86%) of the 125 water molecules (with P_R and P_S alone, the success rates are 81% and 78%, respectively). By analysis of the results in terms of crystallographic data quality, the success rate is 81% for the 37 waters in proteins where one or more of the examined crystallographic structures were of >2.0 Å resolution (endothiapepsin, HIV-1 protease, retinoic acid binding protein II; see Table 1) and is 89% for the 88 waters in proteins with all structures of ≤2.0 Å resolution.

This model gives insight into the characteristics of a conserved water molecule. P_R values higher than 60% begin to be diagnostic for water conservation (Rank ≥ 2.3); i.e., a water should form at least two geometrically suitable hydrogen bonds with the protein. In the application of HINT score to waters, P_H values higher than 80% (HINT score ≥ 400) are strongly predictive for water conservation while P_H lower than 40% (HINT score ≤ 100) are indicative for nonconservation. We have reported that about 515 HINT score units correspond to a $\Delta\Delta G$ of -1 kcal mol^{-1};[23,24] thus, for waters interacting with protein, the free energy difference between those with high probability of conservation and those with high probability of nonconservation ranges between 0.6 and about 2.0 kcal mol^{-1}. This is consistent with a number of previous reports on water energetics.[5,22,32]

The X-ray crystallographic B factor has previously been invoked as a parameter in water displacement prediction[11,12] because it indicates the magnitude of oscillation of an atom around its crystallographic position due to temperature, disorder, or other factors, i.e., a water with low B would be presumed to be more conserved as opposed to a water with high B (and more uncertainty in electron density). However, B factors may vary between structures depending on refinement strategies,[10,29] and relating B factors to a strength of interaction may be misleading.[30] We applied the strategy described above using the crystallographic B factors for the training set waters, as reported in the PDB entries and as normalized with the method of Wang.[10] While the B factor derived models were predictive, they were so to a lesser extent than the Rank or HINT score models (66% successful prediction rate), and three parameter

Table 3. Protein Crystallographic Structures in the Presence of Ligands, in the Complex Test Set

protein	PDB complexed (resolution, Å)	no. waters[a]
aldose reductase	2ikg (1.43), 2ikh (1.55), 2iki (1.47)	3
carbonic anhydrase	1a42 (2.25), 1bn1 (2.10), 1bn4 (2.10)	2
estrogen receptor α	3ert (1.90), 1xpc (1.60), 1xp6 (1.70)	3
factor Xa	1ezq (2.20), 1f0r (2.10), 1f0s (2.10)	4
matrix metalloproteinase-3	1caq (1.80), 1g49 (1.90), 1hfs (1.70)	8
scytalone dehydratase	3std (1.65), 4std (2.15), 5std (1.95)	2

[a] Count of waters located in the binding pocket.

models (Rank, HINT score, and B) were less efficient than the two parameter model of eq 3. In our hands, B is apparently adding more noise than information to the model.

The predictive model of eq 3, using only Rank and HINT score, was then applied to the test set with results that were superior to those obtained on the training set: 59 (87%) of the 68 test set waters were correctly predicted. Again, a correlation was found between results and crystallographic data quality: 24/30 (80%) of the waters from structures with >2.0 Å resolution and 35/38 (92%) of the waters from proteins with all structures having ≤2.0 Å resolution were correctly predicted. These results were not dependent on the split between training and test data sets; models built from other training sets extracted from the pool gave essentially the same predictive accuracy when tested with the remaining pool data. It is worth noting that Rank and HINT score are not highly correlated with respect to each other.[22] The former is purely a geometric index, while the latter encodes chemical information in terms of the actual strengths and biomolecular interactions surrounding each water molecule. These two metrics are clearly providing complementary information to the overall model. In some cases the relevance of a water is not apparent from the uncomplexed protein structure. Such is the case with water 301 of HIV-1 protease. There is little indication from the unliganded crystal structure that this water would have significance; it has only fair hydrogen-bond interactions with the protein and a relatively high B factor.[34] Consequently, our model failed to predict the conservation of water 301 from the uncomplexed structure.

However, as recently noted by Essex,[13] predicting which waters may be easily displaced (or not) in different liganded structures is as important as making these predictions on the free protein. HIV-1 protease water 301 is clearly relevant when examined within the complexed structures. For many drug target proteins only liganded structures are available; being able to predict relevant waters in a protein site already occupied by a substrate or druggable lead would allow more informed choices in targeting waters for displacement via chemical modification of the ligand scaffold or for database search queries. Our insight into the more complex problem of predicting the fate of water molecules between a free and liganded protein, as above, should allow us to predict the role of water molecules in complexed structures. Thus, an additional small test set (6 proteins and 22 waters, Table 3), focusing only on different liganded structures of the same protein, was evaluated by using Rank and HINT score calculated with respect to protein and existing ligand. Our model failed to predict only two of these waters, a success rate of 91%. Although not in this test set, water 301 of HIV-1 protease is predicted to be conserved with >90% probability in the two complexed structures.

Conclusions

Even if often excluded from molecular modeling experiments, water molecules have been shown to play an important role in protein–ligand recognition. In this work we developed a protocol combining in a statistically robust model the HINT score and Rank values, two computational metrics previously shown to be of value for describing water behavior in protein structures.[22] The aim, to identify relevant water molecules that should be considered in protein–ligand docking simulations and structure-based drug design efforts, was largely achieved. For the two test sets (90 water molecules) the success rate of the model was more than 90% on structures with ≤2.0 Å resolution. The clear dependence of these results on crystal structure resolution, even between 2.5 and 2.0 Å, again highlights the difficulties in experimentally placing waters in crystal structure models and the resulting hazards in relying on these water positions in modeling biomacromolecular structure and function. This predictive tool for water relevance can be easily applied to any crystallographic or other biomacromolecular model, as it requires as input data only readily available structural information.

Experimental Section

Data Set Selection Criteria. A number of criteria were applied in selecting protein structures for inclusion in the data sets. First, uncertainties associated with water determination in protein crystallography were considered[28–30] and only structures with resolution better than 2.5 Å were examined, with those having ≤2.0 Å resolution preferred (73% of structures had ≤2.0 Å resolution). At least three structures (one uncomplexed and two complexed) were examined for each protein, except for concanavalin-A, retinol binding protein-II, and cholesterol oxidase, where only two were available (all having ≤2.0 Å resolution). A further check on the positions of some "suspect" waters was performed using GRID,[31] version 22a (Tables 1 and 2). Only proteins presenting little or no binding pocket conformational changes were considered, while some waters in binding pocket regions displaying little variation between the complexed and uncomplexed structure were excluded (Tables 1 and 2).

Visual Analysis of Water Molecules within the Protein Binding Site. The visual analysis of all the water molecules located in the protein binding sites was performed with the criteria described in a previous work aimed at the energetic description of water molecules bound to proteins.[22]

Molecular Models. The three-dimensional coordinates of protein and protein–ligand complex structures were retrieved from the Protein Data Bank and imported into the molecular modeling program Sybyl, version 7.2. All structures were checked for chemically consistent atom and bond type assignment. Hydrogen atoms were added using Sybyl Biopolymer and Build/Edit menu tools and then energy-minimized using the Powell algorithm with a convergence gradient of 0.5 kcal (mol Å)$^{-1}$ for 1500 cycles. This procedure does not affect heavy-atom positions. Water molecules were exhaustively optimized using the HINT tool that finds a global minimum for the orientation of each water molecule with respect to its environment.[27]

HINT Score Calculations. The HINT score is a double sum over all atom−atom pairs of the product (b_{ij}) of the hydrophobic atom constants (a_i, partial log $P_{\text{octanol/water}}$) and atom solvent accessible surface areas (S_i) for the interacting atoms, mediated by a function of the distance between the atoms:

$$\sum_i \sum_j b_{ij} = \sum_i \sum_j (a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}) \tag{4}$$

R_{ij} is usually a simple exponential function, r_{ij} is an adaptation of the Lennard-Jones function, and T_{ij} is a logic function assuming +1 or −1 values, depending on the polar nature of interacting atoms.[21] Partition calculations were performed with the "dictionary" method for the proteins and the "calculate" method for ligands. The "all" partition mode that treats all the hydrogens explicitly was used.[22] Hydrogens bonded to unsaturated carbons were allowed to act as weak hydrogen bond donors. This is in accordance with several recent observations suggesting that some C−H···O hydrogen bonds are possible.[33] The HINT option that corrects the S_i

terms for backbone amide nitrogens by adding 30 Å[22] was used in this study to improve the relative energetics of inter- and intramolecular hydrogen bonds involving these nitrogens.

Rank Algorithm. Rank, which represents the weighted number of potential hydrogen bonds for each water molecule with respect to target molecule(s) surrounding the water, is calculated as

$$\mathrm{Rank} = \sum_n \left\{ (2.80\ \text{Å}/r_n) + \left[\sum_m \cos(\theta_{\mathrm{Td}} - \theta_{nm}) \right] \Big/ 6 \right\} \qquad (5)$$

where r_n is the distance between the water oxygen atom and the target heavy atom n (n is the number of valid targets or a maximum of 4). This is scaled relative to 2.8 Å, the presumed ideal hydrogen bond length. θ_{Td} is the ideal tetrahedral angle (109.5°) and θ_{nm} is the angle between targets n and m ($m = n$ to number of valid targets). The algorithm allows a maximum number of 4 targets (≤ 2 donors and ≤ 2 acceptors). To properly weight the geometrical quality of hydrogen bonds, any angle less than 60° is rejected along with its associated target.[22,27]

Acknowledgment. This work was partially supported by U.S. NIH Grant GM71894 (G.E.K.) and funds from the Italian Ministry of Instruction, University and Research within an Internationalization project (A.M.). J.A.S., from Hampden-Sydney College, was a participant in, and supported by, the VCU Bioengineering and Bioinformatics Summer Institute funded by an NIH/NSF grant. F.S. was supported with a fellowship from the Italian National Institute for Biostructures and Biosystems.

Supporting Information Available: A list of all the analyzed waters with Rank, HINT score, B-factor, and probability; nonlinear regression parameters for weighting (i.e., for eq 3 and Figure 2c). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. *Curr. Med. Chem.* **2004**, *11*, 3093–3118.

(2) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. Binding mode prediction of cytochrome p450 and thymidine kinase protein−ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308–2318.

(3) 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.

(4) Garcia-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624–633.

(5) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.

(6) 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.

(7) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein−ligand docking predictions. *Proteins* **1999**, *34*, 17–28.

(8) Wang, J.; Chan, S. L.; Ramnarayan, K. Structure-based prediction of free energy changes of binding of PTP1B inhibitors. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 495–513.

(9) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.

(10) Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein−ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.

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

(12) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445–464.

(13) 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.

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

(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) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Ericksonviitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380–384.

(17) Coburn, C. A.; Stachel, S. J.; Li, Y. M.; Rush, D. M.; Steele, T. G.; Chen-Dodson, E.; Holloway, M. K.; Xu, M.; Huang, Q.; Lai, M. T.; DiMuzio, J.; Crouthamel, M. C.; Shi, X. P.; Sardana, V.; Chen, Z. G.; Munshi, S.; Kuo, L.; Makara, G. M.; Annis, D. A.; Tadikonda, P. K.; Nash, H. M.; Vacca, J. P.; Wang, T. Identification of a small molecule nonpeptide active site beta-secretase inhibitor that displays a nontraditional binding mode for aspartyl proteases. *J. Med. Chem.* **2004**, *47*, 6117–6119.

(18) 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.

(19) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(20) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.

(21) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: is LogP(o/w) more than the sum of its parts? *Eur. J. Med. Chem.* **2000**, *35*, 651–661.

(22) Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the energetics of water−protein and water−ligand interactions with the "natural" HINT forcefield: predictive tools for characterizing the roles of water in biomolecules. *J. Mol. Biol.* **2006**, *358*, 289–309.

(23) Burnett, J. C.; Botti, P.; Abraham, D. J.; Kellogg, G. E. Computationally accessible method for estimating free energy changes resulting from site-specific mutations of biomolecules: systematic model building and structural/hydropathic analysis of deoxy and oxy hemoglobins. *Proteins* **2001**, *42*, 355–377.

(24) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Simple, intuitive calculations of free energy of binding for protein−ligand complexes. 1. Models without explicit constrained water. *J. Med. Chem.* **2002**, *45*, 2469–2483.

(25) Spyrakis, F.; Cozzini, P.; Bertoli, C.; Marabotti, A.; Kellogg, G. E.; Mozzarelli, A. Energetics of the protein−DNA−water interaction. *BMC Struct. Biol.* **2007**, *7*, 4.

(26) Spyrakis, F.; Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Computational titration analysis of a multiprotic HIV-1 protease-ligand complex. *J. Am. Chem. Soc.* **2004**, *126*, 11764–11765.

(27) Kellogg, G. E.; Chen, D. L. The importance of being exhaustive. Optimization of bridging structural water molecules and water networks in models of biological systems. *Chem. Biodiversity* **2004**, *1*, 98–105.

(28) Carugo, O.; Bordo, D. How many water molecules can be detected by protein crystallography? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 479–483.

(29) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem., Int. Ed.* **2003**, *42*, 2718–2736.

(30) Levitt, M.; Park, B. H. Water: now you see it, now you don't. *Structure* **1993**, *1*, 223–236.

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

(32) Dunitz, J. D. The entropic cost of bound waters in crystals and biomolecules. *Science* **1994**, *264*, 670.

(33) Vargas, R.; Garza, J.; Dixon, D. A.; Benjamin, P. H. How strong is the $\mathrm{C}^{\alpha}\mathrm{-H}\cdots\mathrm{O{=}C}$ hydrogen bond? *J. Am. Chem. Soc.* **2000**, *122*, 4750–4755.

(34) Pillai, B.; Kannan, K. K.; Hosur, M. V. 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *Proteins* **2001**, *43*, 57–64.

(35) Withlow, M.; Howard, A. J.; Stewart, D.; Hardman, K. D.; Kuyper, L. F.; Baccanari, D. P.; Fling, M. E.; Tansik, R. L. X-ray crystallographic studies of *Candida albicans* dihydrofolate reductase. High resolution structures of the holoenzyme and an inhibited ternary complex. *J. Biol. Chem.* **1997**, *272*, 30289–30298.

JM701023H