JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

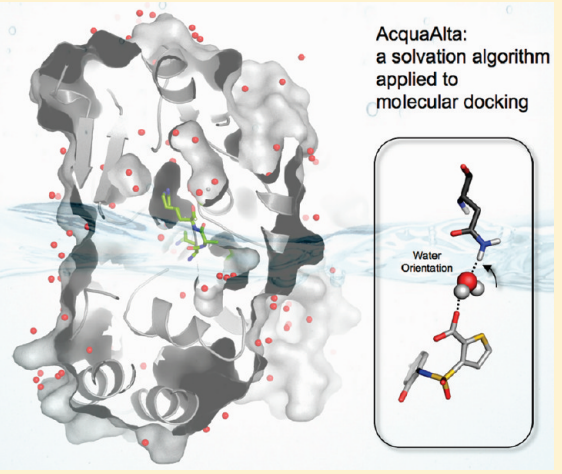
AcquaAlta: A Directional Approach to the Solvation of Ligand−Protein Complexes

Gianluca Rossato,[†] Beat Ernst,[†] Angelo Vedani,[†] and Martin Smieško*[,†]

[†]Institute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

Supporting Information

ABSTRACT: Water molecules mediating polar interactions in ligand−protein complexes can substantially contribute to binding affinity and specificity. To account for such water molecules in computer-aided drug design, we performed an extensive search in the Cambridge Structural Database (CSD) to identify the geometrical criteria defining interactions of water molecules with ligand and protein. In addition, with *ab initio* calculations the propensity of ligand hydration was evaluated. Based on this information, we developed an algorithm (AcquaAlta) to reproduce water molecules bridging polar interactions between ligand and protein moieties. This approach was validated with 20 crystal structures and yielded a match of 76% between experimental and calculated water positions. When water molecules establishing only weak interactions with the protein were neglected, the match could be improved to 88%. Supported by a pharmacophore-based alignment tool, the solvation algorithm was then applied to the docking of oligopeptides to the periplasmic oligopeptide binding protein A (OppA). Calculated waters based on the crystal poses matched an average of 66% of the experimental waters. With water molecules calculated based on the docked ligands, the average match with the experimental waters dropped to 53%.



INTRODUCTION

Water molecules mediating ligand−protein interactions can affect ligand affinity as well as increase the specificity of the interaction (i) by forming bridging hydrogen bonds[1] or (ii) by increasing the flexibility of the binding site.[2] Water molecules can form both intramolecular (i.e waters between two ligand atoms or two protein atoms) and intermolecular bridges (i.e., waters between ligand and protein atoms). Hereafter, bridging water molecules will refer to solvent molecules linking polar functional groups of ligand and protein.

An analysis of 392 high-resolution complexes[3] retrieved from the Protein Data Bank (PDB)[4] showed that over 85% of the ligand−protein complexes include at least one water molecule bridging ligand and protein. In the analyzed complexes, the average number of ligand-bound water molecules was found to be 4.6. In addition, 76% of these water molecules were identified to participate in polar interactions with both ligand and protein. A statistical analysis on protein−protein complexes revealed that water-mediated interactions are as abundant as direct protein−protein hydrogen bonds.[5]

Previous studies reported the utmost importance of considering water molecules in structure-based drug design and pharmacophore modeling.[6−8] The displacement of a water molecule can substantially affect the binding free energy. In the case of the HIV protease,[9] the displacement of ordered water molecules by cyclic urea inhibitors led to a substantial gain in entropy.[10] On the contrary, in the case of the periplasmic oligopeptide binding protein A (OppA), an analogous displacement of water molecules decreased the binding affinity.[11] Consequently, in computer-aided drug discovery, substantial efforts are being taken to distinguish relevant and retainable water molecules from displaceable ones. Such a classification can be based on structural features such as hydrogen-bond counting,[12] the solvent-accessible polar surface area, the polarity of the withholding cavity, the characterization of the binding-pocket shape,[13] and water conservation in homologous proteins,[14] as well as on advanced statistical methods.[15] Water molecules located in polar cavities forming two or more hydrogen bonds are typically considered structural waters.[3,16] Such waters are therefore frequently conserved in protein complexes with different ligands and show thermal factors (B_{iso}) in the same range as the protein atoms they are hydrogen-bonded to.[3] As a thermodynamic consequence, binding energies of ligands displacing structural water molecules suffer from a loss in enthalpy, which may only be partially compensated by a gain in entropy.

Despite previous attempts using a hydration penalty score[17] or defining general rules for the treatment of water molecules

Received: March 29, 2011
Published: June 29, 2011

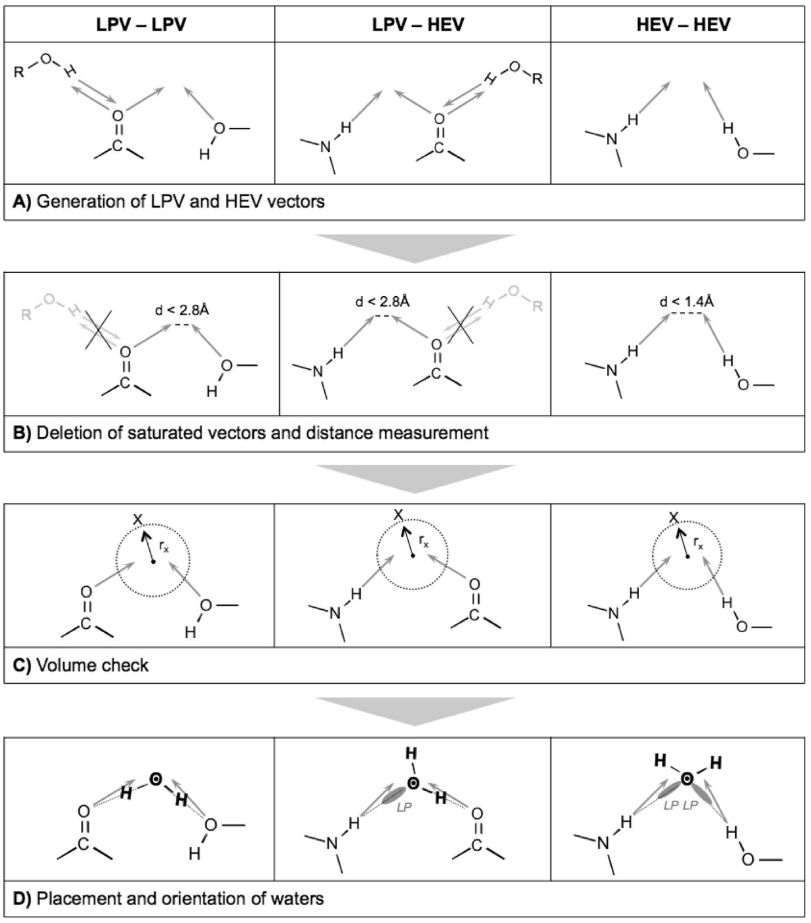


Figure 1. Figure displaying different steps of the vectors generation, scans, placement, and orientation procedure for the possible kinds of vectors combination (LPV−LPV, LPV−HEV, HEV−HEV): A) generation of the LPV and HEV vectors, B) scan for vectors not involved in other inter- or intramolecular hydrogen bonds and measurement of the distance to the unsaturated vectors; C) check of availability of space through distance measurement toward ligand and protein atoms (X) where r_x is the van der Waals radius of the atom +0.6 Å; D) placement and orientation of a water molecule toward the atoms origin of the vectors involved in the hydrogen bond.

binding at ligand-protein interfaces,[18] a thorough and general parametrization is difficult to extrapolate. Therefore, the solvation pattern of the binding site at each target protein should be analyzed individually[19] with particular attention to enthalpic and entropic contributions toward binding.[20]

Computational algorithms in molecular simulations use different approaches to hydration. Some concepts calculate free-energy changes characterizing the water contribution toward the binding affinity[21,22] or their relation to the hydration of binding sites.[15,23−25] Potential hydration sites can be evaluated by solvent mapping or by modular neural networks.[26,27] Furthermore, predictive tools to characterize and distinguish roles of water in biomolecules are available.[28−30] Hydration of molecular complexes is generally addressed either considering water implicitly[31] or explicitly.[32−35] Some studies emphasize the need of water molecules to improve accuracy in molecular docking,[36−39] while others claim that the presence of water improved the docking results only marginally.[40] However, since water molecules at binding interfaces are frequently observed, their presence along with their role must be considered in any molecular simulation.

In this account, we present a novel approach (AcquaAlta) where explicit water molecules are generated at the ligand−protein interface. The underlying algorithm relies on preferred positions and orientations of water molecules as extracted from structural information collected from the Cambridge Structural Database — CSD[41] (currently including the structures of more than 500,000 organic molecules). Specifically, we searched for water molecules interacting with generic functional groups (e.g., the carbonyl query is comprehensive of all aldehydes, ketones, carboxylic acids, esters, and amides) of small organic molecules. To establish a hydration-propensity ranking, water interaction energies were obtained from *ab initio* calculations on hydrated functional groups. In AcquaAlta, water molecules bridging interactions between ligand and protein partners are generated considering the hydration propensities of the involved functional groups and aromatic moieties.

For the validation of AcquaAlta, we attempted to reproduce bridging water molecules found in 20 mostly high-resolution crystal structures. This solvation approach was then applied to the docking of oligopeptides binding to OppA. We selected OppA because a large number of high-resolution ligand−protein complexes is available. Interestingly, different hydration patterns among apo OppA and ligand-bound holo OppA have been reported.[42,43]

The envisioned application area for AcquaAlta is ligand docking where generated water molecules (in the binding site) should increase the probability of finding the bioactive pose and therefore facilitate a more reliable ranking and scoring of the docked poses.

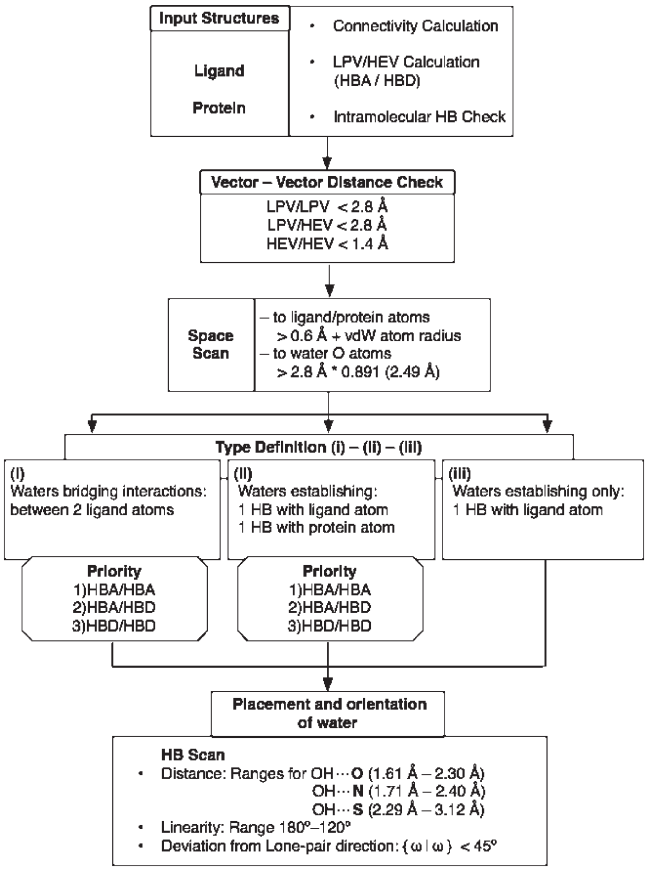


Figure 2. Flowchart representing the algorithm organization. Calculations and procedures are performed in the same order as shown in the flowchart from the top to the bottom and from left to right. HBA stands for hydrogen bond acceptors and HBD stands for hydrogen bond donors. Criteria and ranges for hydrogen bond (HB) acceptance are included.

METHODS

Cambridge Structural Database (CSD) Search. The analysis of small-molecule crystal structures regarding hydrogen bond geometries (distances, angles and torsions) with water molecules was performed in the CSD (v5.31 — November 2009 + 3 updates).[41] The queries (cf. Results section) were submitted using ConQuest, and the histograms in the Supporting Information were plotted using Vista, included in the CSD software suite. The following filters were applied in the search of crystal structures of small molecules:

- All 3D coordinates determined (no calculated atom positions)
- Crystallographic R factor ≤0.05
- No disordered, or polymeric structures, no metal ions present
- Solely organic compounds (no organometallic compounds)

Terminal hydrogen positions were normalized during the search. Hits refer to structures matching the query search. Multiple fragments (e.g., two carboxylates interacting with water molecule(s) in the same crystal structure) for one hit were possible. In order to produce the 3D plots (see, for example, Table 3 and Table 4) we wrote a program in C++ for transforming bond distances, bond angles, and torsion angles into Cartesian coordinates.

Ab Initio Analysis of Ligand–Water Interactions. Preferred interaction geometries and associated energies of generic functional groups of the ligand with a single water molecule were obtained from *ab initio* calculations using Gaussian 03.[44] The molecular complexes as well as the isolated entities were fully optimized in the gas phase using the Møller–Plesset perturbation theory (with energy corrections truncated at the second order, MP2) with 6-311++G(d,p) basis set.[45,46] At the optimized geometries, an analysis of the vibrational modes was performed to confirm that they represent true minima (without imaginary frequencies detected). The final interaction energy, ΔE, was calculated as the difference between the energy of the complex and the sum of energies of its respective isolated entities. Finally, ΔE was corrected for the superposition error of the basis set using the counterpoise method.[47,48] The generic functional groups were then ranked according to their hydration propensity based on the interaction energy (cf. Figure 4).

The setup with only a single water molecule in the gas phase was chosen to obtain an unbiased comparison of the strength of interaction (ΔE) with a given functional group for the hydration propensity scale. In reality the hydrogen-bond strength and the position of the water are also affected by nearby solvent, ligand or protein atoms.

AcquaAlta Organization. The solvation of ligand–protein complexes can be done using two different concepts: a) the solvation is performed before the docking and b) the ligand is docked and then the complex is solvated. In the former case waters are generated and oriented based solely on either ligand or protein atom positions. In the latter case (in AcquaAlta) the waters are generated and oriented based on the positions of both ligand and protein atoms simultaneously, which should lead to a better identification of waters bridging ligand–protein interactions.

Using a concept similar to the solvation module of Yeti,[49] vectors originating from hydrogen-bond donors (marking the ideal position of a hydrogen bond acceptor and termed hydrogen extension vectors, HEVs) and hydrogen-bond acceptors (marking the ideal position of a hydrogen-bond donor; lone pair vectors, LPVs) are calculated to identify ideal positions for water molecules. These vectors are generated using specific geometries for each functional group based on data retrieved from the CSD (cf. Tables 3 and 4).

In 3D space, halogen atoms present both an electropositive region (also called "corona") along the extension of the C–X (X = Cl, Br, I) bond and an electronegative region along the axis perpendicular to this bond (also called "belt"), displaying an amphoteric character with respect to polar interactions, in ligand–protein complexes.[50] It has been observed that intermolecular distances increase with polarizability of the halogen atom involved in the interaction, i.e. larger for iodine than for fluorine.[50] In AcquaAlta, halogen atoms are considered as potential hydrogen-bond acceptors and apolar hydrogen atoms as potential hydrogen-bond donors.[51] For both cases, in order to identify the starting position of the search for an interaction counterpart, the vectors are considered as the extensions of the carbon–halogen and carbon–hydrogen bond, respectively.

Figure 1 depicts the concept of AcquaAlta. First, HEV and LPV vectors are generated (Figure 1A), then, those vectors associated with atoms already engaged in inter- and intramolecular hydrogen bonds are marked as "saturated" and are not further considered (Figure 1B). Distances between unsaturated vectors are calculated and considered as potential water sites if the following criteria are met: distances of LPV–LPV and LPV–HEV below 2.8 Å, which refer to an O–H···O

Table 1. List of the 20 X-ray Structures Used for Validation of AcquaAlta along with PDB Codes, Resolutions, R-Factors, and Number of Binding Site Waters

#	biological target	PDB code	res. [Å]	R-factor	# of waters
1	trypsin	2ayw	0.97	0.138	8
2	dihydrofolate reductase	3dfr	1.70	0.152	23
3	thymidin kinase	1e2k	1.70	0.209	5
4	VEGFR2	1ywn	1.71	0.206	2
5	glycogen phosporylase	1a8i	1.78	0.182	13
6	human phospodiesterase	1xp0	1.79	0.194	1
7	beta trypsin	1bju	1.80	0.171	2
8	holo-glyceraldehyde 3P dehydrogenase	1gd1	1.80	0.177	15
9	Hsp90	1uy6	1.90	0.184	4
10	AmpC beta-lactamase	1xgj	1.97	0.168	2
11	2CDK2	2b53	2.00	0.223	3
12	ACE	1o86	2.00	0.180	6
13	COMT	1h1d	2.00	0.174	4
14	HIV-1 protease	1hpx	2.00	0.170	4
15	non-nucleoside adenosine deaminase	1ndw	2.00	0.206	6
16	ACK1	1u4d	2.10	0.205	5
17	coagulation factor XA	1f0r	2.10	0.216	2
18	thymidin kinase	1kim	2.14	0.209	3
19	EGFR	1xkk	2.40	0.209	2
20	EGFR	1m17	2.60	0.251	1

hydrogen-bond length, and HEV−HEV distance below 1.4 Å (Figure 1B).

To ensure that an appropriate free volume for a potential water molecule is available, distances from the midpoint of the two vectors tips are checked against any ligand and protein atom. The distance for acceptance must be larger than the sum of the individual van der Waals radius of the ligand or protein atom (X) plus 0.6 Å (at this cutoff value the highest number of experimental waters was matched by the algorithm).

If the space criterion is met, the water molecule is placed at the midpoint of the vector tips (Figure 1C) into an averaged position determined by the two vectors and then oriented toward its partners (Figure 1D). In the case that a water molecule is establishing just one interaction with a ligand atom, the unengaged hydrogen or oxygen atom is oriented toward the closest free protein HEV or LPV.

The detailed organization of the algorithm is shown in Figure 2.

Input Structures. The structures of ligand and protein atoms are supplied in PDB format in order to determine connectivities for all functional groups of ligand and protein and to generate vectors (HEV, LPV, and originating from halogen and apolar hydrogen atoms). The presence of vectors already engaged in inter- and intramolecular hydrogen bonds (both in ligand and protein) is checked, and associated vectors are deleted.

Vector−Vector Distance Check. Distances are checked for LPV/LPV, LPV/HEV, and HEV/HEV combinations. Distances must remain under the defined thresholds.

Space Scan. Distances toward any ligand and protein atom are calculated to ensure that a sufficiently large volume to accommodate a water molecule is available. To check whether a water molecule has been already placed in the vicinity, distances between water oxygen atoms must be larger than 2.49 Å [sum of van der Waals radii (2.8 Å) multiplied by a factor of 0.891 (threshold value below which the Leonard-Jones 6−12 potential becomes repulsive)].

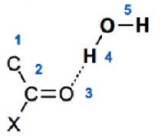
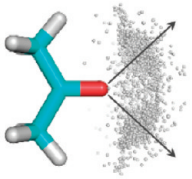
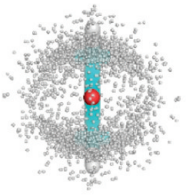
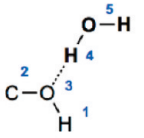
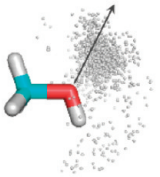
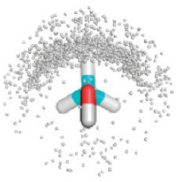
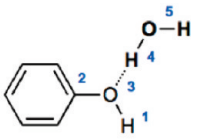
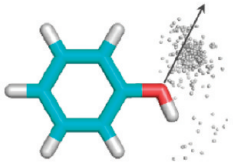
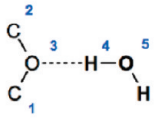
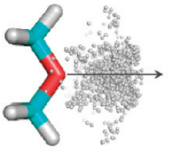
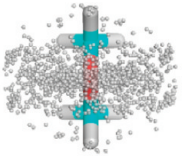
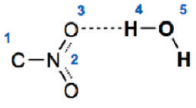
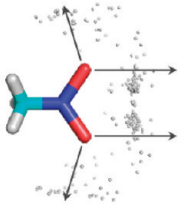
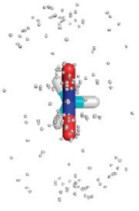
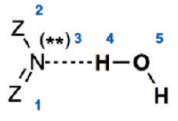
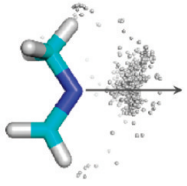
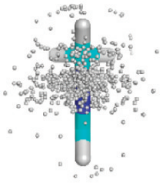
Table 2. List of the OppA Crystal Structures Used in the Automated Docking Procedure along with PDB Codes, Resolutions, R-Factors, Oligopeptide Sequences, and Number of Binding Site Waters

#	PDB code	resolution [Å]	R-factor	oligopeptide sequence	# of waters
1	1JET	1.2	0.229	KAK	7
2	1JEU	1.25	0.224	KEK	9
3	1JEV	1.30	0.203	KWK	6
4	1B4Z	1.75	0.179	KDK	10
5	1B5I	1.9	0.182	KNK	7
6	1B32	1.75	0.182	KMK	7
7	1B3F	1.8	0.177	KHK	7
8	1B46	1.8	0.177	KPK	6
9	1B51	1.8	0.179	KSK	9
10	1B58	1.8	0.179	KYK	7
11	1B5J	1.8	0.182	KQK	10
12	1B9J	1.8	0.179	KLK	6
13	1QKA	1.8	0.179	KRK	6
14	1QKB	1.8	0.181	KVK	6

Type Definition. While generating the positions of bridging water molecules, three different types, with decreasing priority, are defined: i) waters bridging interactions within two ligand atoms, ii) waters interacting with both ligand and protein, and iii) waters establishing only a single hydrogen bond with the ligand.

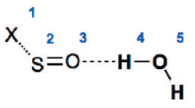
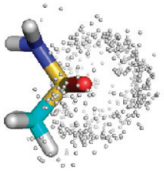
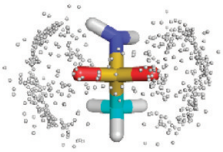
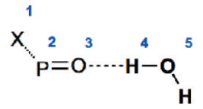
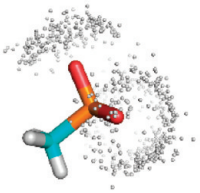
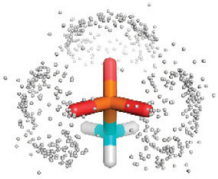
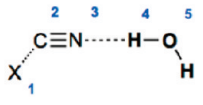
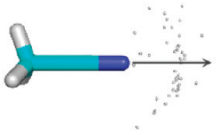
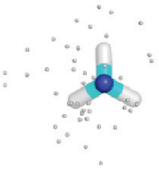
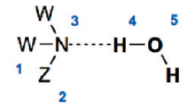
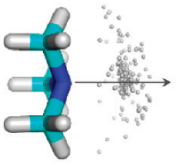
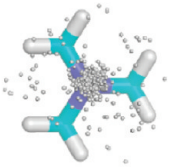
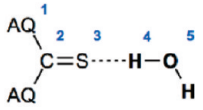
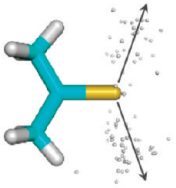
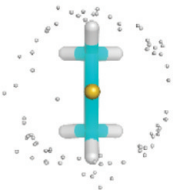
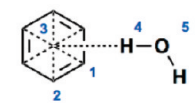
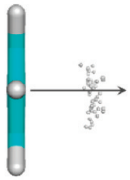
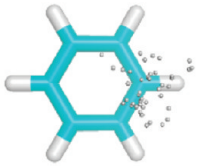
Water Placement and Orientation. Finally, the water molecule is placed and oriented toward the origin of the closest unsaturated vector. Vectors are considered based on their hydration propensity as ranked from the *ab initio* calculations

Table 3. Summary of the Crystallographic Data Search for Functional Groups Acting As Hydrogen Bond Acceptors (HBA)[a]

Entry	Query:	Side view	Front view	Measurements
	HBA: Functional groups acting as hydrogen bond acceptors			
1	Carbonyl  Hits/Fragments: 997/1378			d [3-4] (Å) / angle [3-4-5] (°) Mean 1.94±0.21 / 160.0±17.5 Median 1.87 / 165.8 LQ/HQ 1.76/2.24 / 134.1/175.1
2	Aliphatic hydroxyl  Hits/Fragments: 861/1478			d [3-4] (Å) / angle [3-4-5] (°) Mean 2.03±0.28 / 151.9±23.0 Median 1.91 / 159.7 LQ/HQ 1.78/2.54 / 115.2/173.6
3	Phenolic hydroxyl  Hits/Fragments: 168/267			d [3-4] (Å) / angle [3-4-5] (°) Mean 2.08±0.27 / 148.8±24.2 Median 1.96 / 156.0 LQ/HQ 1.81/2.55 / 112.0/172.7
4	Ether  Hits/Fragments: 375/645			d [3-4] (Å) / angle [3-4-5] (°) Mean 2.24±0.31 / 141.4±25.1 Median 2.20 / 146.4 LQ/HQ 1.86/2.66 / 105.7/171.9
5	Nitro  Hits/Fragments: 48/97			d [3-4] (Å) / angle [3-4-5] (°) Mean 2.33±0.26 / 112.9±23.3 Median 2.33 / 113.7 LQ/HQ 1.92/2.68 / 78.3/146.1
6	Imine  Hits/Fragments: 451/671			d [3-4] (Å) / angle [3-4-5] (°) Mean 2.1±0.30 / 157.9±17.6 Median 1.97 / 164.0 LQ/HQ 1.84/2.64 / 130.6/175.1 (**) nitrogen not charged

dx.doi.org/10.1021/ci200150p |J. Chem. Inf. Model. 2011, 51, 1867–1881

Table 3. Continued

Entry	Query:	Side view	Front view	Measurements		
	HBA: Functional groups acting as hydrogen bond acceptors					
7	Sulfoxide /Sulfone  Hits/Fragments: 139/271				d [3-4] (Å)	angle [3-4-5] (°)
				Mean	1.96±0.23	159.0±16.3
				Median	1.78	164.6
				LQ/HQ	1.78/2.34	134.0/174.1
8	Posphine oxide / Posphone  Hits/Fragments: 133/271				d [3-4] (Å)	angle [3-4-5] (°)
				Mean	1.88±0.18	162.3±15.9
				Median	1.84	166.4
				LQ/HQ	1.73/2.0	145.5/174.9
9	Nitrile  Hits/Fragments: 30/36				d [3-4] (Å)	angle [3-4-5] (°)
				Mean	2.10±0.23	160.4±17.4
				Median	2.02	167.0
				LQ/HQ	1.85/2.60	126.3/176.0
10	Tertiary amine  Hits/Fragments: 206/259				d [3-4] (Å)	angle [3-4-5] (°)
				Mean	2.13±0.33	153.9±24.2
				Median	1.97	164.3
				LQ/HQ	1.83/2.7	112.8/174.4
11	Thiokethon  Hits/Fragments: 59/76				d [3-4] (Å)	angle [3-4-5] (°)
				Mean	2.45±0.18	101.9±13.9
				Median	2.40	98.9
				LQ/HQ	2.26/2.70	86.9/123.0
12	Aromatic ring  Hits/Fragments: 40/43				d [3-4] (Å)	angle [3-4-5] (°)
				Mean	2.59±0.15	86.3±12.6
				Median	2.57	86.6
				LQ/HQ	2.40/2.81	71.2/106.3

[a] An entry number indicates each query search. In the left column, the query as submitted to the CSD; the number of hits and the number of fragments found. In the central column a typical molecule with 3D representation of the positions of water hydrogen atoms (gray spheres) and the positions of the LPVs (gray arrows). Light-blue sticks represent general fragments resembling the submitted query. In the right column, statistical results (mean, median, lower, and higher quantile) for the values of distance and linearity of hydrogen bonds between functional-group fragments and water molecules. In the 3D-plots of sulfoxide/sulfone and phosphine/phosphone the queries represent all acceptor atoms to resemble real functional groups. For clarity, in the 3D-plots of the aromatic ring as HBA, the water-hydrogen positions are displayed only for one carbon atom. In the query representation X stands for any kind of atom, Z for any kind of atom with exception of hydrogen, W for C and H atoms, AQ for C, N, O, S atoms.

(Figure 4 — Results). The geometric criteria for acceptance are loosened from very stringent to moderate giving priority to distance, linearity, and deviation from the lone-pair direction.

Predicting Bridging Water Molecules in High-Resolution Crystal Structures. Table 1 lists the proteins retrieved from the PDB for the purpose of algorithm validation. The crystal structures were selected according to their resolution, the number of bridging waters or because they were the object of previously published studies dealing with water molecules bridging polar interactions.

The PDB files were split into protein and ligand. Next, atom types based on the AMBER* force-field and atomic partial charges were assigned. Protonation and tautomeric states for ligand and protein structures were determined using Epik[52] from the Schrodinger software suite.

The criteria defining a bridging water molecule were arbitrarily set to a distance up to 3.3 Å from any ligand atom and interaction with any kind of polar protein atom within the threshold 2.7—3.3 Å. An energy filter — a water molecule was considered "bridging", if its total interaction energy, E_{tot} (E_{ele}+E_{vdW}+E_{HB}), was equal or lower than −1.0 kcal/mol after rotational optimization (oxygen atom kept fixed) based on the Yeti force field[53] — was applied in addition. For the comparison of experimental and calculated waters, the distance threshold for oxygen-water matching was defined as 1.4 Å (50% of the O—O distance of an OH···O hydrogen bond).

Application to Docking. AcquaAlta was applied to ligand—protein complexes in which the ligand poses were generated by Alignator.[54] This pharmacophore-based alignment tool takes advantage of a conformer pool generated by a conformational search. In our case, we used Macromodel[52] and performed in implicit water, using the OPLS2005 force-field, 20,000 search iterations (1000 steps for each rotatable bond), using an energy window of 100 kcal/mol. Each accepted conformer was then aligned to a template molecule (i.e., Lys-Asn-Lys for OppA) based on the matching pharmacophores. Of all possible solutions, only those having no close contacts with the protein and maximum number of superimposed pharmacophores were retained. Further information on this docking approach together with its application are published elsewhere.[54] Protein—ligand complexes were stripped of the solvent. Then, both crystal poses and poses produced by Alignator (14 OppA complexes — Table 2) were resolvated by AcquaAlta. Finally, the water molecules were optimized using Yeti.[53] The deviation between crystal waters and calculated waters was monitored for each oligopeptide—protein complex using the same distance criteria as in the validation step (cf. above).

RESULTS AND DISCUSSIONS

Water Directionality. In our CSD searches we obtained detailed information on distance, linearity, and deviations from lone-pair vectors for 16 generic functional groups interacting with water molecules (Figure S1 — Supporting Information). From these data, we extracted preferred geometries for water molecules interacting with such entities for our solvation algorithm. Tables 3 and 4 list the results for the different functional groups included in the search. For each of the 16 generic functional groups the following information are shown: the query, the number of hits and fragments obtained, a 3D representation (side and front view) of water hydrogen atoms (in the case of functional groups acting as HBA) and water oxygen atoms (in the case of functional groups acting as HBD), together with relevant statistical data for distance and angle of interaction.

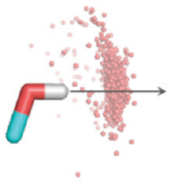
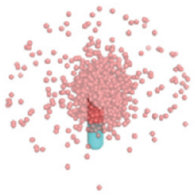
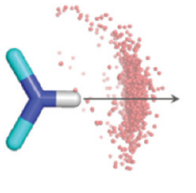
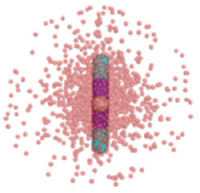
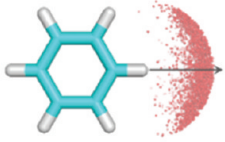
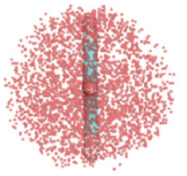
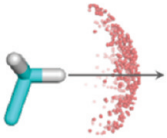
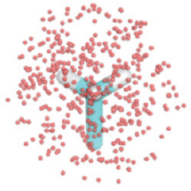
We used general queries to compile rules applicable to a wide variety of similar functional groups (e.g., carbonyl oxygen in aldehyde, ketone, carboxylic acid, ester, and amide; carboxylate groups were treated separately: cf. Figure S1 — Supporting Information) and to monitor geometries of interaction of waters involved in polar interactions. Quantitative information on specific moieties, (e.g., pyridines vs pyrimidine) and the effects of neighboring atoms of the functional groups are not extractable from the collected data (e.g., ketone vs ester).

The median length of a hydrogen bond is required to be shorter than the sum of the individual van der Waals radii. The difference between strong and weak hydrogen bonds is documented with the following examples. The mean distances for carbonyl (1.94 Å) and for phenolic hydroxyl (2.08 Å) are substantially, although not significantly shorter than the mean distances for the ones established with ether (2.24 Å) and nitro groups (2.33 Å). In addition, the spatial distribution of the water molecules interacting with a carbonyl or a phenolic/aliphatic hydroxyl positions is considerably narrowed compared to ether or nitro groups. In AcquaAlta, the acceptable lower limit for the hydrogen-bond distances is shifted compared to the values obtained from the CSD (Figure 2). This allows for a reasonable treatment of interactions between water and charged atoms where the distance is even shorter than the average hydrogen bond distance (the mean distance for the $-COO^{-}\cdots HOH$ interaction is 1.92 ± 0.26 Å and the mean distance for the $-NR_2H^{+}\cdots OH_2$ interaction is 1.86 ± 0.21 Å).

Water prefers hydrogen bonds with a linear X-H···Y arrangement. Significant deviations from this geometry are typically caused by additional interactions with nearby atoms.[55] For oxygen acceptor atoms in carbonyls, hydroxyls, sulfoxides/sulfones, phosphines/phosphones, and nitrogen acceptor atoms in tertiary amines, imines, and nitriles as well as all the hydroxyl and amino donor atoms the median values for angles vary from 156° to 180° (angle statistics in Table 3). In contrast, for acceptors such as ethers, nitro groups, thioketones, aromatic rings, and aromatic and aliphatic "hydrogen-bond donor" groups (Table 4 — entries 15, 16) the median values for angles substantially deviate from linearity, ranging from 87° (aromatic ring) to 148° (aliphatic CH). AcquaAlta allows for the X-H···Y angle to range from 180° (linear) to 120°, which therefore covers most cases listed above.

Hydrogen bonds are usually directed along the lone pair(s) of the acceptor atom.[56] Since CSD/Conquest does not allow to define lone pairs, the torsion angle defined by the three atoms of the acceptor group (denoting the lone-pair plane) and the water hydrogen atom were analyzed (Figure S1 — Supporting Information). The 3D plot in Table 3 (entries 1—6, 9—12) shows the distribution of the hydrogen atoms of water molecules for hydrogen-bond acceptors and how water is arranged along the LPVs. Interestingly, water hydrogen atoms in aliphatic and phenolic hydroxyls (entries 2, 3) are also positioned in between the two LPVs. Sulfoxide/sulfone and phosphine/phosphone (entries 7, 8) do not show a well-defined distribution because the S—O and P—O bond have only partial double-bond character.[57] In the latter two cases (entries 7, 8) water hydrogen atoms are distributed equally (i.e., without any spatial preference) forming a "corona"-like pattern around the acceptor oxygen atom. The deviation from the acceptor plane is generally below 30°.[58] In our algorithm, the optimal positions for interaction with water are at the end points of LPVs and HEVs. Deviations from the LPV or HEV of up to 45° are allowed (Figure 3).

Table 4. Summary of the Crystallographic Data Search for Functional Groups Acting As Hydrogen Bond Donors (HBD)[a]

Entry	Query	Side view	Front view	Measurements
	HBA: Functional groups acting as hydrogen bond acceptors			
13	OH - Hydroxyl Hits/Fragments: 1266/1814			d [3-4] (Å) / angle [2-3-4] (°) Mean: 1.80±0.21 / 162.70±15.0 Median: 1.76 / 166.7 LQ/HQ: 1.6/1.99 / 146.2/175.6
14	NH - Amino Hits/Fragments: 1138/1599			d [3-4] (Å) / angle [2-3-4] (°) Mean: 2.0±0.23 / 156.7±16.6 Median: 1.94 / 160.8 LQ/HQ: 1.75/2.34 / 134.3/173.3
15	CH - Aromatic Hits/Fragments: 758/1213			d [3-4] (Å) / angle [2-3-4] (°) Mean: 2.54±0.13 / 145.6±16.1 Median: 2.56 / 145.8 LQ/HQ: 2.35/2.69 / 124.3/167.6
16	CH - Aliphatic Hits/Fragments: 305/433			d [3-4] (Å) / angle [2-3-4] (°) Mean: 2.59±0.10 / 146.9±15.0 Median: 2.61 / 147.8 LQ/HQ: 2.46/2.70 / 125.4/167.1

[a] The organization is identical to Table 3 with the only exception of water oxygen atoms, which are displayed with red spheres and the gray arrows, which refer to the HEVs.

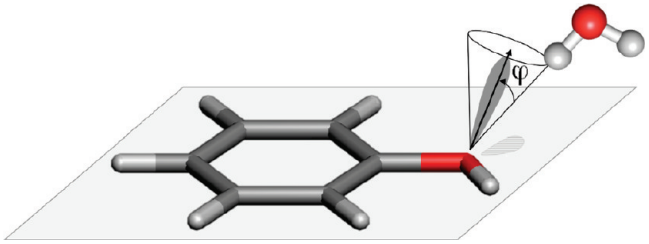


Figure 3. Definition of the allowed deviation (φ) from an ideal LPV (black arrow), in this case assumed at the lone-pair position. The same criteria are used for HEVs.

Despite similar studies,[56−61] no analysis focusing exclusively on water molecules interacting with generic functional groups is available. Consistent with a previous study,[59] our analysis revealed a clear sp^2 lone-pair preference of the carbonyl when interacting with water but not for the sp^3 lone pair for ether oxygen atom. A previous study showed hydrogen-bonding preference for heteroaromatic nitrogen acceptors compared to heteroaromatic oxygen acceptors.[60] Our results showed that water hydrogen atoms are less scattered around the lone pair of the nitrogen (entry 6) when compared to the distribution around the lone pairs of oxygen acceptor atom (entry 4).[60]

Forty well-defined query hits (entry 12) for the interaction between water (as HBD) and aromatic rings suggest that even fragments generally considered as lipophilic can form favorable interactions with water molecules. In this case, the aromatic-ring atoms behave as HBAs and water molecules orient themselves to maximize the interactions of their hydrogen atoms with the π-electron cloud of the aromatic carbons, in a manner comparable with a hydrocarbon $\sigma-\pi$ interaction.

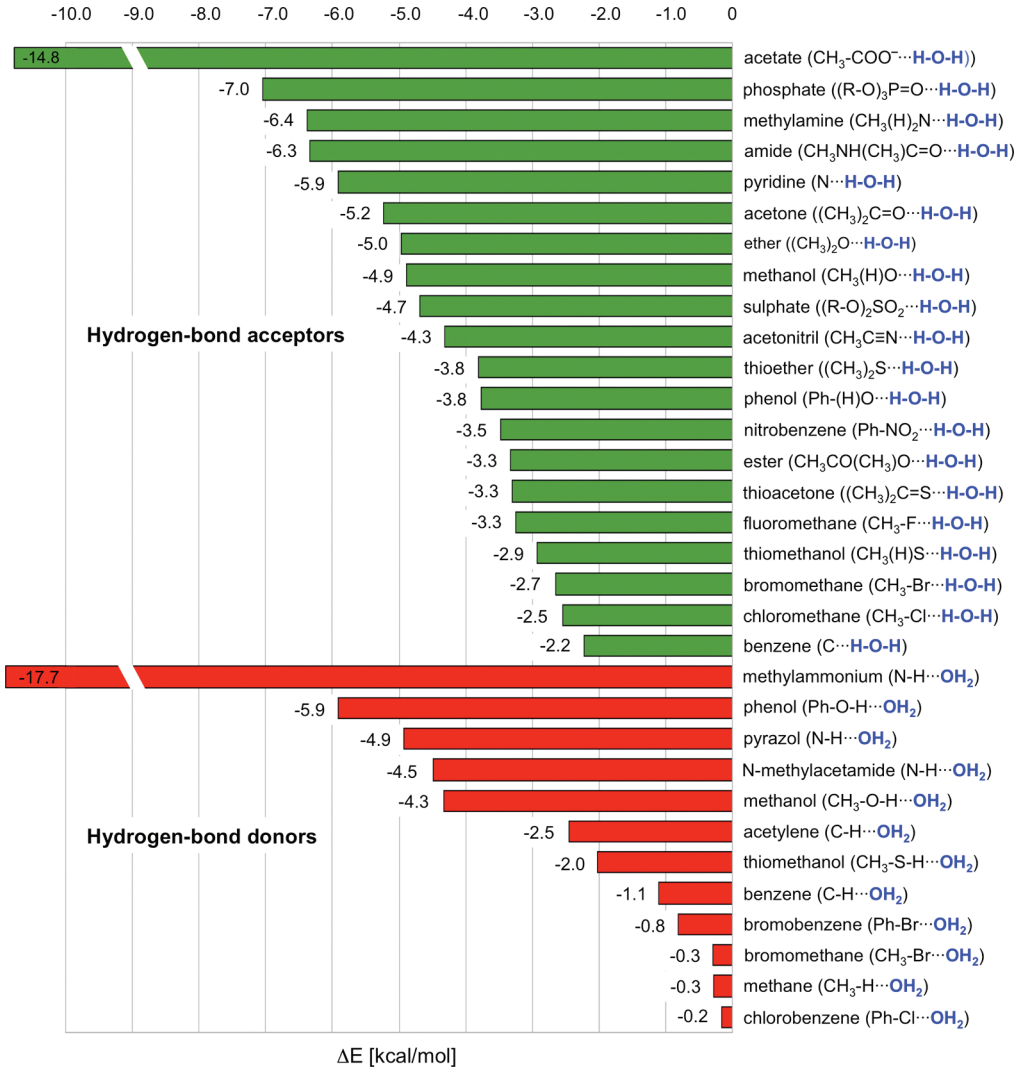


Figure 4. *Ab initio* calculated interaction energies for selected compounds interacting with water. Hydrogen-bond acceptors are depicted in green, hydrogen-bond donors in red.

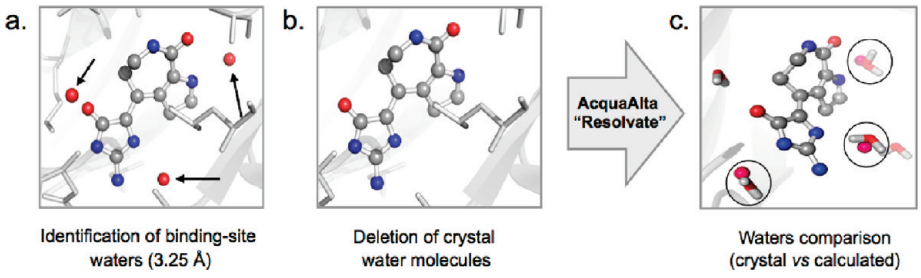


Figure 5. Algorithm validation. a) Identification of water molecules within 3.25 Å from each heavy atom of the ligand. b) Deletion of the water molecules. Application of the AcquaAlta solvation algorithm (gray arrow). c) Comparison between experimental and calculated waters (oxygen atom only).

Water oxygen atoms interacting with hydroxyl and amino HBDs (Table 4 — entries 13, 14) are distributed along the HEVs (OH and NH). Interestingly, in the case of the water oxygen interacting with apolar hydrogen atoms (Table 4 — entries 15, 16), the average median distance of 2.59 Å is lower than the sum of the van der Waals radii for hydrogen and oxygen (2.72 Å). This value, together with the range between lower and higher quantile of the two queries (2.35−2.70 Å), shows that a weak interaction is established between the apolar aromatic and aliphatic hydrogen atoms and the oxygen of the water and that the water oxygen position is not a consequence of packing effects.[51] [Quantiles: values marking equally sized consecutive subsets of an ordered sample of population (in our search analysis the quantile is set to 10).] The localization of the oxygen atoms of water shows no

dx.doi.org/10.1021/ci200150p |*J. Chem. Inf. Model.* 2011, 51, 1867–1881

Table 5. Results from the Validation Process[a]

entry	PDB code	experimental binding site waters	match by AcquaAlta	identified bridging waters	match by AcquaAlta
1	2ayw	8	6 (75%)	3	3 (100%)
2	3dfr	23	12 (52.2%)	12	7 (58.3%)
3	1e2k	5	4 (80%)	2	2 (100%)
4	1ywn	2	1 (50%)	2	1 (50%)
5	1a8i	13	10 (77.0%)	9	6 (66.7%)
6	1xp0	1	1 (100%)	1	1 (100%)
7	1bju	2	2 (100%)	1	1 (100%)
8	1gd1	15	11 (73.3%)	8	6 (75%)
9	1uy6	4	4 (100%)	4	4 (100%)
10	1xgj	2	2 (100%)	2	2 (100%)
11	2b53	3	2 (66.7%)	1	1 (100%)
12	1o86	6	5 (83.4%)	4	4 (100%)
13	1h1d	4	3 (75%)	2	2 (100%)
14	1hpx	4	2 (50%)	3	2 (66.7%)
15	1ndw	6	3 (60%)	2	2 (100%)
16	1u4d	5	3 (60%)	3	2 (66.7%)
17	1f0r	2	1 (50%)	1	1 (100%)
18	1kim	3	2 (66.7%)	3	2 (66.7%)
19	1xkk	2	2 (100%)	1	1 (100%)
20	1m17	1	1 (100%)	1	1(100%)

[a] Listed are 20 PDB structures ordered by resolution (see Table 1) and identified by an entry number. Examined are (a) number of experimental binding site waters and corresponding match by AcquaAlta, (b) number of identified bridging waters and corresponding match by AcquaAlta. No direct correlation was found between water matching and B-factor range of the experimental waters.

preferred distribution. Based on atomic distances and lone pair deviations, we deduce that water molecules are indeed interacting with these fragments through electrostatic interactions and that their spatial orientation is mainly determined by the neighboring atoms.

Finally, one aim of the CSD search was to assess whether the small size and relatively high mobility of a water molecule has an impact on the geometry of hydrogen bonds. Our results (Table S2 − Supporting Information) are generally in agreement with values obtained from queries where water is represented only as a generic hydroxyl group. However, a detailed analysis of all the interactions shows that water acting as acceptor systematically forms shorter hydrogen bonds especially if interacting with strong donors (OH, NH) when compared to the generic hydroxyl query. In contrast, water acting as hydrogen bond donor forms longer hydrogen bonds than those found for the generic hydroxyl.

Thresholds and value ranges used in our approach are in agreement with the geometries obtained in the CSD search. In order to process protein structures with lower resolution and possibly less accurate geometries,[62,63] AcquaAlta softens the corresponding geometric criteria during an iterative scan.

Functional-Group Hydration Propensity. Despite the availability of geometric criteria for describing hydrogen bonds (see Tables 3 and 4), rules to calculate associated energies (enthalpy, entropy) are scarce.[64,65] In addition, the treatment of hydrogen-bond networks has only recently been addressed.[66,67] Up to date, the NIST database[68] lists only a few experimentally determined parameters for strength of hydrogen bonds between water and small charged species like methylammonium and acetate.

Different hydration propensities of ligand functional groups and protein side chains have been elucidated through an extensive statistical analysis performed on 392 high-resolution ligand−protein structures.[3] A recent study[69] experimentally determined the relative basicity and therefore the strength of hydrogen-bond acceptors using Fourier transform infrared spectrometry (FTIR). When compared to NMR, UV, and IR techniques, this method allows to identify and to rank HBAs of polyfunctional bases. Due to the extreme diversity of the structures deposited in the CSD, hydrogen-bond frequencies and geometries can hardly be used to quantify their strengths for establishing quantitative structure−activity relationships. Nonetheless, the agreement between the abundance of certain hydrogen bonds in the CSD and the acceptor strength[69] can indicate the likelihood of hydrogen-bond formation.

In order to obtain relative interaction energies as well as geometric information, particularly valuable for interactions that are not well represented in the CSD, we performed a series of *ab initio* calculations on hydrated functional groups (Figure 4). Here, calculated hydrogen-bond interaction energies were used as a direct measure of a given functional group's propensity to hydration. A subsequent ranking based on those interaction energies inspired by a statistical analysis of 17 ligand atom types[3] was used to fine-tune our own hydration-propensity scale for the AMBER* force-field atom types.[70]

Oxygen atoms of carboxyls are most likely to be involved in hydrogen bond interactions. Their *ab initio* interaction energies compared to those of sp^3 oxygen atoms of ethers and especially esters are substantially higher (ΔE ranges from −14.8 to −5.2 kcal/mol vs −5.0 to −3.3 kcal/mol). A similar trend for these two groups is observed when frequency of hits and number of fragments found in the CSD search are compared. Due to the low electronegativity of the phosphorus atom, the oxygen atom of a P═O group is also a strong acceptor (ΔE = −7.0 kcal/mol), while the same atom in S═O or N═O groups forms comparably weaker hydrogen bonds (ΔE of −4.7 and −3.5 kcal/mol,

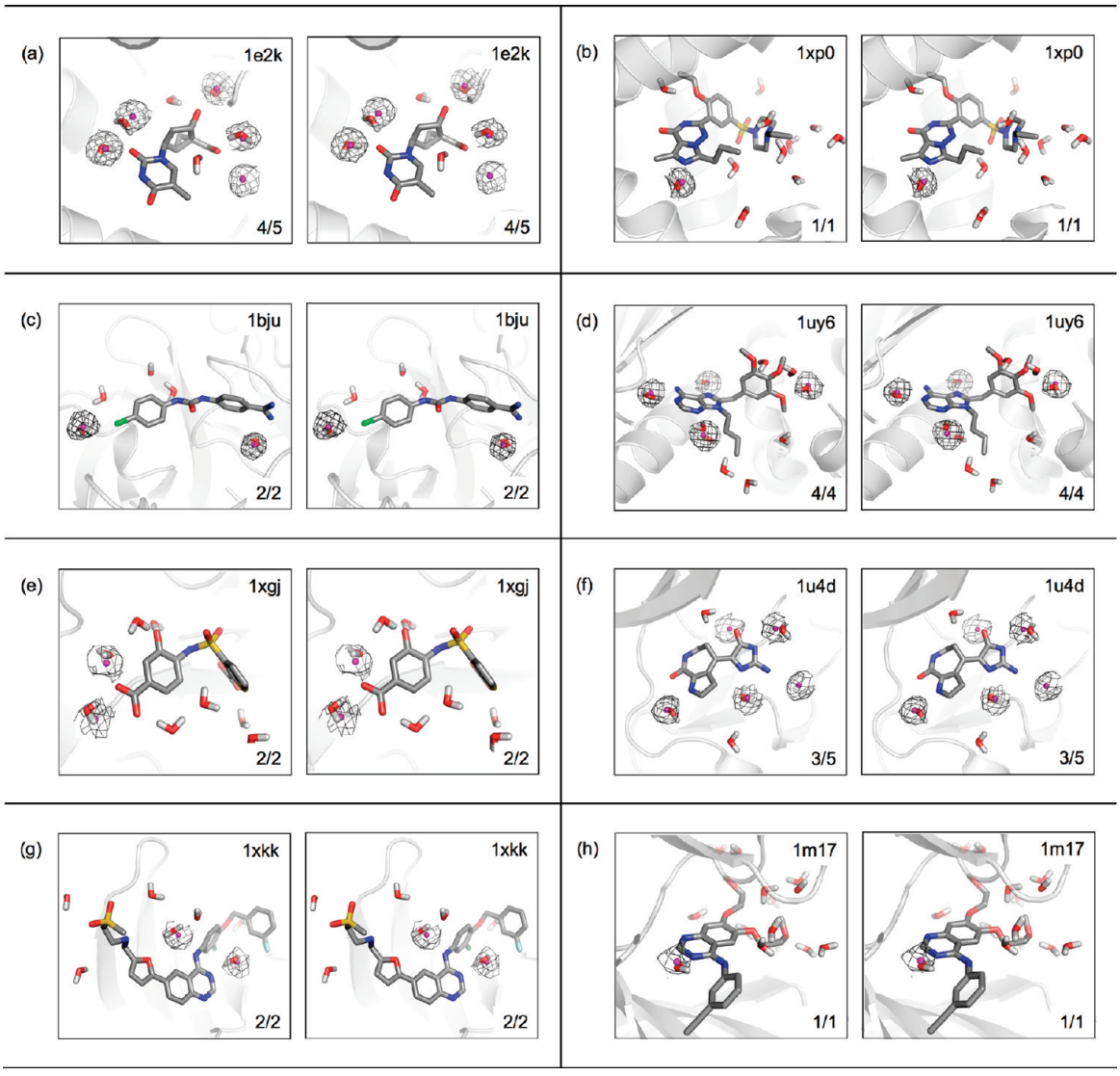


Figure 6. 3D representation (stereo view) of selected structures (PDB code in the top right corner) used in the validation. Ligand carbons are displayed in gray sticks, other ligand atoms are colored by atom type. Experimental waters are displayed as magenta spheres, water electron-density map as gray mesh and calculated waters in sticks representation. The match between experimental and calculated waters is given in the bottom right corner.

respectively). A nitrogen sp^2 atom incorporated in pyridine rings (−5.9 kcal/mol) forms a slightly stronger hydrogen bond than the *sp* hybridized nitrogen atom in acetonitrile (−4.3 kcal/mol). A sulfur atom, regardless of its hybridization state, was found to be a weak hydrogen bond acceptor, comparable to halogen atoms, where the fluorine atom was the most willing hydrogen bond acceptor. In the ideal case, aromatic carbon atoms can form a well-defined $\sigma-\pi$ interaction with a water molecule, contributing as much as −2.2 kcal/mol, which is roughly 40% of a strong hydrogen bond (e.g., $O-H\cdots O_{acetone}$ = −5.2 kcal/mol).

When acting as a hydrogen-bond acceptor, water forms the strongest interactions with charged nitrogen groups. Otherwise no substantial difference in interaction strength (ΔE) can be found between N−H and O−H groups, with both being frequent and potent hydrogen-bond donors. Thiols form only weak interactions with water oxygen (ΔE = −2.0 kcal/mol), as a matter of fact, they are even weaker than the best interaction formed with some hydrocarbons (e.g., for acetylene, ΔE = −2.45 kcal/mol). Our calculation further suggests that halogens form only weak interactions with water. The strongest halogen donor···water interaction in Figure 4 is the one of bromobenzene featuring a ΔE of only −0.8 kcal/mol. Although, some of the listed interactions are quite rare (e.g., esters,[56] sulfones, and sulfonamides[58]), they are included in Figure 4 to cover fragments as found in some of the CSD structures.

Validation of AcquaAlta. The AcquaAlta algorithm was validated on a data set of suitable crystal structures retrieved from the PDB. The structures belong to different pharmacological and biological target families. The resolution of the selected 20 crystal structures (Table 1) ranges from 0.97 Å for the bovine β-trypsin (PDB code: 2ayw) to 2.60 Å for the epidermal growth factor receptor (PDB code: 1m17). A subset of 15 structures (75%) showed resolutions lower or equal to 2.0 Å.

The most probable protonation state of both ligand and protein at physiological pH (7.4) was calculated using Epik.[52] Then, calculated hydrogen atoms (as they are not typically resolved in the experimental structure) were added to the ligand and the protein structures by using the Yeti software. In the

dx.doi.org/10.1021/ci200150p |*J. Chem. Inf. Model.* 2011, 51, 1867–1881

Table 6. Results from Automated Docking: The 13 Peptides Aligned to the Template KNK Are Listed along with the "Self-Fit" Experiment (Entry # 5)[a]

#	peptide aligned to KNK	matching possible PP	*rmsd* PP	*rmsd* aligned to crystal	match with crystal	match with aligned pose
1	KAK	26/28	0.486	0.623	4/7 (57.1%)	4/7 (57.1%)
2	KEK	27/31	0.768	0.634	5/9 (55.6%)	6/9 (66.7%)
3	KWK	27/31	0.923	0.916	5/6 (83.3%)	3/6 (50%)
4	KDK	28/30	0.805	0.619	5/10 (50%)	4/10 (40%)
5	KNK	32/32	0.550	0.495	5/7 (71.4%)	3/7 (42.9%)
6	KMK	26/31	0.811	0.513	3/7 (42.9%)	3/7 (42.9%)
7	KHK	28/32	0.975	0.933	5/7 (71.4%)	2/7 (28.6%)
8	KPK	24/25	0.597	0.831	5/6 (83.3%)	5/6 (83.3%)
9	KSK	28/29	0.721	0.660	7/9 (77.8%)	6/9 (66.7%)
10	KYK	27/30	1.068	2.885	5/7 (71.4%)	3/7 (42.9%)
11	KQK	30/33	1.271	1.280	6/10 (60%)	5/10 (50%)
12	KLK	27/30	0.839	1.230	5/6 (83.3%)	4/6 (66.7%)
13	KRK	27/37	0.533	0.855	5/6 (83.3%)	2/6 (33.3%)
14	KVK	25/30	0.768	0.540	2/6 (33.3%)	4/6 (66.7%)

[a] Included are (a) number of the matched pharmacophore centers (PP) out of possible pharmacophores for each tripeptide, (b) the *rmsd* of the pharmacophores between template and aligned tripeptides, (c) the *rmsd* between the aligned tripeptide and the corresponding tripeptide from the crystal structure, (d) match of calculated to experimental waters when tripeptide from the X-ray structure and from the docking was used.

validation, we identified all experimental waters located within 3.25 Å from any ligand heavy atom (Figure 5a, waters indicated with arrows). To allow for an unbiased procedure, the experimental water coordinates were deleted (Figure 5b), and in the main step, were recalculated by AcquaAlta, oriented and minimized using the Yeti software, featuring a directional force-field. In the last step (Figure 5c), the positions of calculated waters were compared to the experimental position.

The match between experimental and calculated waters is summarized in Table 5. A filter (filtering criteria listed in the Methods section) to discriminate between binding site waters and bridging waters was applied. The match refers to calculated waters compared with experimental waters for each structure. Two different matches are provided considering experimental binding site waters or identified bridging waters.

The comparison (Table 5 and Figure 6) shows a 76% match between experimental and calculated binding site waters, while the match for the identified bridging waters was 87.5%. Analysis of the subset of 15 crystal structures with resolutions equal to or lower than 2.0 Å provided a match of 76.2% for binding site waters and a better prediction of 87.8% for bridging waters, showing a small but not significant improvement of the results.

In 6 out of 20 structures (entries 2, 4, 5, 14, 16, 18), the match of the bridging waters is lower or equal to 66.7%. Two structures (entries 2 and 5) contain rather high numbers of waters in the binding site (23 and 13, respectively). In both structures, the high number of water molecules originates from the presence of cofactors, which are also included in the solvation process, as well as from the ligands being localized in solvent accessible areas. Despite the high resolution of entry 2 (1.70 Å), only a low match (52.2% and 58.3% for binding site and bridging waters, respectively) is obtained, a fact that can probably be attributed to the localization of both ligand and cofactor on the enzyme surface where water networks can easily be formed. The poor result obtained for this structure can be explained by the fact that water-network geometries are not implemented in our approach.

The poor results obtained for entries 4, 14, 16, and 18, despite the ligand not being localized on the protein surface, are related to the concept of the algorithm. AcquaAlta aims to generate waters engaging in polar interactions with the ligand, therefore bridging interaction to polar atoms of the protein. Waters that were not matched by our algorithm (e.g., W163 of entry 4, W566 and W607 of entry 14, W503 and W603 of the chain b of entry 16, and W498 of entry 18) are bridging hydrogen bonds between the protein atoms and have only some electrostatic interactions with ligand atoms. Most of such unmatched waters are positioned beyond geometrical limits implemented in the AcquaAlta program, and their positions are primarily influenced by interactions with the protein atoms or "second shell" waters (e.g., W504 interacting with W503 of entry 16).

Selected examples of the results obtained in the validation phase (Table 5) are depicted in Figure 6. In all eight cases the match of the experimental binding site waters is higher or equal to the 60%. The calculated waters both match the experimental oxygen positions and fit the experimental water electron-densities.

APPLICATION OF ACQUAALTA TO AUTOMATED DOCKING

In order to challenge the solvation algorithm in a molecular docking application, we chose the oligopeptide binding protein (OppA) as a target. Fourteen crystal structures of the OppA with resolutions lower than 2.0 Å are available in the PDB (Table 2). In these structures, OppA is complexed with different tripeptides of the type lysine-X-lysine (K-X-K), where X refers to a variable amino-acid residue.

Molecular docking was performed in two steps. First, the prealignment/docking tool Alignator[54] was employed to align the tripeptides. The underlying protocol aligns low-energy conformers of the docked molecule, obtained from conformational searching in aqueous solution, to a template molecule, based on pharmacophores matching. Solutions having unfavorable close contacts with the protein, which may require substantial induced fit for a proper accommodation, are discarded.

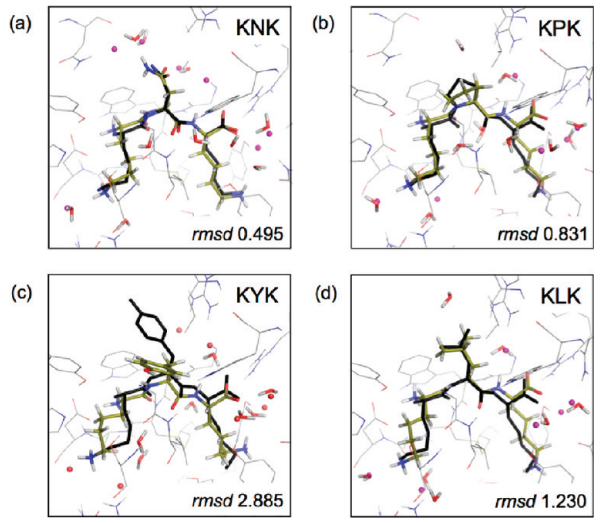


Figure 7. 3D representation of selected PDB structures (tripeptide sequence, upper right corner) from the automated docking. Crystal ligands are displayed in black sticks without oxygen atoms. Experimental waters are displayed as magenta spheres and calculated waters, based on the docked result, in sticks representation. In the bottom right corner the heavy-atoms *rmsd* of the oligopeptide (conformations between the crystal and the docked result) is given.

For this purpose, we selected the tripeptide KNK as template, which contains the highest number of diverse pharmacophore centers. The structural similarity of the docked tripeptides and the template is high, although the scaffold of KXK features a substantial number of rotatable bonds (>10). Moreover, for KXK motifs with X = D (aspartate) or X = E (glutamate), most of the low energy conformers have the two lysine side chains folded back to form an intramolecular stabilization. For this reason, the conformational search was set up with a rather large energy window (100 kcal/mol) and up to 12,000 conformers for each tripeptide were retained and aligned by Alignator. Accepting a wider variety of conformers increases the probability of identifying extended conformer geometries similar to those bound in the binding pocket.

The oligopeptides aligned to KNK displayed a high match of pharmacophore centers and in the case of the "self-fit" (i.e., conformers of KNK aligned to itself) all the pharmacophores between the docked tripeptide and the template were matched. The *rmsd* of the matched pharmacophores (Table 6) is lower than 1.0 Å in all the cases except of KYK (Figure 7c) where the tyrosine side chain adopts an alternative conformation compared to the crystal structure (χ_2 crystal = −61°, χ_2 docked = 58°, i.e. gauche$^-$ vs gauche$^+$) resulting in an apparent high *rmsd* value. The *rmsd* calculated on the heavy-atoms for aligned oligopeptides with the corresponding crystal conformation was always found to be lower than 1.0 Å, with the exception of KYK, KLK (Figure 7c,d), and KQK. High *rmsd* for KYK and KQK reflects different side chain orientation of the X amino acid. Alignments of tripeptides with long side chains (e.g., the two terminal lysine residues) are remarkably complex due to high number of possible combinations of their dihedral angles. Examples of aligned poses compared to the crystal structures, and experimental versus calculated water molecules, are shown in Figure 7.

After minimization, water molecules generated with AcquaAlta were compared with the experimental waters for both crystal and Alignator-generated ligand poses, using the same distance criteria as in the validation phase. Neither geometry nor energy filters were applied. The average match for the 14 crystal poses is 66%, which is by 10% lower than the match obtained for the 20 complexes used in the validation. When the best *rmsd* pose from Alignator was used as input for the solvation algorithm, the resulting average match of water positions was 52.7%. A slight drop of accuracy can be expected, since in the adopted alignment protocol the conformers are aligned to the pharmacophore centers of a single oligopeptide template (KNK). In addition, as one can deduce from the *rmsd* scores, the matches are not perfect particularly for solvation sensitive hydrogen atoms in the terminal NH_3^+ group of the lysine side chain.

CONCLUSIONS

A correct simulation of ligands binding to proteins should account for structural and bridging solvent molecules. Based on a survey on 392 high-resolution crystal structure, an average of 4.6 water molecules are found at the ligand−protein interface, and three-quarters of them are involved in polar interactions with both ligand and protein.[3]

A rigorous analysis of small-molecule crystal structures from the CSD was performed to collect data on the geometry of interactions of water molecules when interacting with generic functional groups. *Ab initio* interaction energies between water and selected functional group representatives were calculated to construct an empiric hydration propensity scale. This information serves the basis for an algorithm (AcquaAlta), which solvates ligand-protein interfaces. AcquaAlta is based on the directionality of hydrogen bonds and can solvate both classical hydrogen bond moieties as well as halogen and hydrophobic functionalities of the ligand.

The approach was thoroughly validated on 20 X-ray structures with resolution ranging from 0.97 to 2.60 Å, checking the match of experimental and calculated water molecules. The match for binding site waters was 76%. The algorithm accuracy was not influenced substantially by the resolution of the crystal structures. When we applied geometry and energy filters to identify only the water molecules having polar contacts with both ligand and protein (i.e., bridging waters), the match rose to 87.5%. WATGEN,[71] an algorithm to model water networks at protein−protein surfaces, yielded comparable results predicting 72% and 88% of water sites placed within 1.5 and 2.0 Å, respectively.

Subsequently, we combined our solvation approach with a pharmacophore-based alignment tool and applied it to 14 structures of OppA. This approach yielded poses within a rmsd of 1.0 Å from the crystal structures for 13 out of 14 complexes. The match, without any filtering criteria, between experimental and calculated water molecules was 66% and 53% when based on the crystal and docked poses, respectively. This drop in performance highlights the importance of a reasonable starting position to correctly reproduce water molecule positions. In a control test for the previously mentioned WATGEN algorithm,[71] solvent molecules randomly placed at protein−protein interfaces were able to match only 22% and 40% of the experimental water sites within 1.5 and 2.0 Å.

In our approach, the solvated ligand-protein complex is not evaluated based on energy or on modification of entropy and enthalpy of the binding but rather on evaluation of the presence of hydrogen bond partners that a water molecule could bridge and thus bring to favorable interaction. The aim of this solvation algorithm is to produce crystal-like binding poses with optimally

arranged bridging waters, expecting that further refinement using force-field minimization routine would benefit and lead to a more accurate evaluation of thermodynamics of ligand binding. AcquaAlta is thought to be applied to docking studies where the presence or absence of a water molecule, calculated on each ligand conformation, can substantially affect pose scoring. On one side this could help obtaining reasonable binding modes; on the other one, the geometric criteria used do not give any information whether the water will enhance or worsen the binding. AcquaAlta is available on request through the Web site http://www.modeling.unibas.ch/AcquaAlta

■ ASSOCIATED CONTENT

Supporting Information. Further details of the results from the CSD search, *ab initio* calculations, along with additional CSD searches. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Fax: +41 61 267 1552. E-mail: martin.smiesko@unibas.ch. Corresponding author address: Department of Pharmaceutical Sciences, Pharmazentrum, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland.

■ ACKNOWLEDGMENT

Financial support by the Swiss National Foundation is gratefully acknowledged (grant number: 200021_119817).

■ ABBREVIATIONS:

CSD, Cambridge Structural Database; PDB, Protein Data Bank; HEV, Hydrogen-Extension Vector; LPV, Lone-Pair Vector; OppA, Oligopeptide binding protein A; HBA, Hydrogen Bond Acceptors; HBD, Hydrogen Bond Donors; QSAR, Quantitative structure−activity relationships

■ REFERENCES

(1) Quiocho, F.; Wilson, D.; Vyas, N. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **1989**, *340* (6232), 404–407.

(2) Poole, P. L.; Finney, J. L. Hydration-induced conformational and flexibility changes in lysozyme at low water content. *Int. J. Biol. Macromol.* **1983**, *5* (5), 308–310.

(3) Lu, Y.; Wang, R.; Yang, C.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47* (2), 668–675.

(4) ProteinDataBank. http://www.pdb.org/ (accessed June 10, 2011).

(5) Rodier, F.; Bahadur, R.; Chakrabarti, P.; Janin, J. Hydration of protein-protein interfaces. *Proteins* **2005**, *60* (1), 36–45.

(6) Mancera, R. L. De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (7), 479–499.

(7) García-Sosa, A. T.; Mancera, R. L. The effect of a tightly bound water molecule on scaffold diversity in the computer-aided de novo ligand design of CDK2 inhibitors. *J. Mol. Model.* **2006**, *12* (4), 422–431.

(8) Lloyd, D. G.; García-Sosa, A. T.; Alberts, I. L.; Todorov, N. P.; Mancera, R. L. The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comput.-Aided Mol. Des.* **2004**, *18* (2), 89–100.

(9) Lam, P.; Jadhav, P.; Eyermann, C.; Hodge, C.; Ru, Y.; Bacheler, L.; Meek, J.; Otto, M.; Rayner, M.; Wong, Y. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263* (5145), 380–384.

(10) Li, Z.; Lazaridis, T. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.* **2003**, *125* (22), 6636–6637.

(11) Ladbury, J. 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* (12), 973–980.

(12) Poornima, C.; Dean, P. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 500–512.

(13) Poornima, C.; Dean, P. Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 513–520.

(14) Poornima, C.; Dean, P. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 521–531.

(15) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129* (9), 2577–2587.

(16) Fischer, S.; Verma, C. S. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (17), 9613–9615.

(17) Mancera, R. L. A new explicit hydration penalty score for ligand−protein interactions. *Chem. Phys. Lett.* **2004**, *399* (1−3), 271–275.

(18) García-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* (3), 624–633.

(19) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: concanavalin A. *J. Phys. Chem. B* **2005**, *109* (1), 662–670.

(20) Michel, J.; Tirado-Rives, J.; Jorgensen, W. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.* **2009**, *131* (42), 15403–15411.

(21) Mancera, R. L. Molecular modeling of hydration in drug design. *Curr. Opin. Drug Discovery Dev.* **2007**, *10* (3), 275–280.

(22) García-Sosa, A. T.; Mancera, R. L. Free Energy Calculations of Mutations Involving a Tightly Bound Water Molecule and Ligand Substitutions in a Ligand-Protein Complex. *Mol. Inf.* **2010**, *29* (8−9), 589–600.

(23) Michel, J.; Tirado-Rives, J.; Jorgensen, W. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113* (40), 13337–13346.

(24) Abel, R.; Young, T.; Farid, R.; Berne, B.; Friesner, R. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817–2831.

(25) Pearlstein, R.; Hu, Q.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V.; Daniels, D.; Hanrahan, S.; Sherman, W.; Abel, R. New hypotheses about the structure-function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor-like repeat A docking site using WaterMap. *Proteins* **2010**, *78* (12), 2571–2586.

(26) Kortvelyesi, T.; Dennis, S.; Silberstein, M.; Brown, L. r.; Vajda, S. Algorithms for computational solvent mapping of proteins. *Proteins* **2003**, *51* (3), 340–351.

(27) Ehrlich, L.; Reczko, M.; Bohr, H.; Wade, R. Prediction of protein hydration sites from sequence by modular neural networks. *Protein Eng.* **1998**, *11* (1), 11–19.

(28) Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D.; Kellogg, G.; 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* (1), 289–309.

(29) García-Sosa, A.; Mancera, R.; Dean, P. 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* (3), 172–182.

(30) Raymer, M.; Sanschagrin, P.; Punch, W.; Venkataraman, S.; Goodman, E.; Kuhn, L. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265* (4), 445–464.

(31) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson−Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22* (6), 608–640.

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

(33) Huang, N.; Shoichet, B. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51* (16), 4862–4865.

(34) Friesner, R.; Murphy, R.; Repasky, M.; Frye, L.; Greenwood, J.; Halgren, T.; Sanschagrin, P.; Mainz, D. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.

(35) Corbeil, C.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47* (2), 435–449.

(36) de Graaf, C.; Oostenbrink, C.; Keizers, P.; van der Wijst, T.; Jongejan, A.; Vermeulen, N. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49* (8), 2417–2430.

(37) Thilagavathi, R.; Mancera, R. Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Model.* **2010**, *50* (3), 415–421.

(38) Yang, J.; Chen, C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* **2004**, *55* (2), 288–304.

(39) Roberts, B. C.; Mancera, R. L. Ligand-protein docking with water molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408.

(40) Birch, L.; Murray, C.; Hartshorn, M.; Tickle, I.; Verdonk, M. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (12), 855–869.

(41) Allen, F. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., B* **2002**, *58* (Pt 3 Pt 1), 380–388.

(42) Davies, T. G.; Hubbard, R. E.; Tame, J. R. Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes. *Protein Sci.* **1999**, *8* (7), 1432–1444.

(43) Tame, J. R.; 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* (12), 998–1001.

(44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J.A.; Vreven, T.; Kudin, T. K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, C.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. In Gaussian 2003; J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.

(45) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46* (7), 618–622.

(46) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153* (6), 503–506.

(47) Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen − bonded dimers? *J Chem. Phys.* **1996**, *105*, 11024.

(48) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553.

(49) Vedani, A.; Huhta, D. W. Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *J. Am. Chem. Soc.* **1991**, *113* (15), 5860–5862.

(50) Lu, Y.; Wang, Y.; Zhu, W. Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design. *Phys. Chem. Chem. Phys.* **2010**, *12* (18), 4543–4551.

(51) Taylor, R.; Kennard, O. Crystallographic Evidence for the Existence of C-H-0, C-H-N, and C-,H.-Cl Hydrogen Bonds. *J. Am. Chem. Soc.* **1982**, *104*, 5063–5070.

(52) Schrödinger. http://www.schrodinger.com/ (accessed June 10, 2011).

(53) Vedani, A.; Huhta, D. W. A New Force-Field for Modeling Metalloproteins. *J. Am. Chem. Soc.* **1990**, *112* (12), 4759–4767.

(54) Rossato, G.; Ernst, B.; Smiesko, M.; Spreafico, M.; Vedani, A. Probing Small-Molecule Binding to Cytochrome P450 2D6 and 2C9: An In Silico Protocol for Generating Toxicity Alerts. *ChemMedChem* **2010**, *5* (12), 2088–2101.

(55) Klemperer, W. The Potential to Surprise. *Nature* **1993**, *362*, 698.

(56) Murray-Rust, P.; Glusker, J. P. Directional hydrogen bonding to sp2- and sp3-hybridized oxygen atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.* **1984**, *106* (4), 1018–1025.

(57) Vedani, A.; Dunitz, J. D. Lone-pair directionality in hydrogen-bond potential functions for molecular mechanics calculations: the inhibition of human carbonic anhydrase II by sulfonamides. *J. Am. Chem. Soc.* **1985**, *107* (25), 7653–7658.

(58) Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, *53* (14), 5061–5084.

(59) Lommerse, J. P. M.; Price, S. L.; Taylor, R. Hydrogen bonding of carbonyl, ether, and ester oxygen atoms with alkanol hydroxyl groups. *J. Comput. Chem.* **1997**, *18*, 757–774.

(60) Nobeli, I.; Price, S. L.; Lommerse, J. P. M.; Taylor, R. Hydrogen bonding properties of oxygen and nitrogen acceptors in aromatic heterocycles. *J. Comput. Chem.* **1997**, *18*, 2060–2074.

(61) Taylor, R.; Kennard, O.; Versichel, W. Geometry of the imino-carbonyl (N-H···O:C) hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **1983**, *105* (18), 5761–5766.

(62) DePristo, M.; de Bakker, P.; Blundell, T. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **2004**, *12* (5), 831–838.

(63) Henrik, C.; Görbitz, C. H.; Etter, M. C. Hydrogen bonds to carboxylate groups. The question of three-centre interactions. *J. Chem. Soc., Perkin Trans. 2* **1992**, 131–135.

(64) Bohm, H. J.; Schneider, G. *Protein−Ligand Interactions from Molecular Recognition to Drug Design*; Wiley-VCH Verlag GmbH and Co.: Weinheim, Germany, 2003.

(65) Baron, R.; Setny, P.; McCammon, J. A. Water in cavity-ligand recognition. *J. Am. Chem. Soc.* **2010**, *132* (34), 12091–12097.

(66) Matthews, B. W.; Liu, L. A review about nothing: are apolar cavities in proteins really empty?. *Protein Sci.* **2009**, *18* (3), 494–502.

(67) Plumridge, T. H.; Waigh, R. D. Water structure theory and some implications for drug design. *J. Pharm. Pharmacol.* **2002**, *54* (9), 1155–1179.

(68) NIST Chemistry WebBook. http://webbook.nist.gov/ (accessed June 10, 2011).

(69) Laurence, C.; Brameld, K.; Graton, J.; Le Questel, J.; Renault, E. The pK(BHX) database: toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J. Med. Chem.* **2009**, *52* (14), 4073–4086.

(70) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784.

(71) Bui, H.; Schiewe, A.; Haworth, I. WATGEN: an algorithm for modeling water networks at protein-protein interfaces. *J. Comput. Chem.* **2007**, *28* (14), 2241–2251.