Published in final edited form as:

J Med Chem. 2012 January 26; 55(2): 623–638. doi:10.1021/jm2005145.

A forcefield with discrete displaceable waters and desolvation entropy for hydrated ligand docking

Stefano Forli and Arthur J. Olson*

Molecular Graphics Lab, Dept. Molecular Biology, MB-112, The Scripps Research Institute
10550, North Torrey Pines Road, La Jolla, CA 92037-1000, USA

Abstract

In modeling ligand-protein interactions, the representation and role of water is of great importance. We introduce a forcefield and hydration docking method that enables the automated prediction of waters mediating the binding of ligands with target proteins. The method presumes no prior knowledge of the *apo* or *holo* protein hydration state, and is potentially useful in the process of structure-based drug discovery. The hydration forcefield accounts for the entropic and enthalpic contributions of discrete waters to ligand binding, improving energy estimation accuracy and docking performance. The forcefield has been calibrated and validated on a total of 417 complexes (197 training set; 220 test set), then tested in cross-docking experiments, for a total of 1649 ligand-protein complexes evaluated. The method is computationally efficient and was used to model up to 35 waters during docking. The method was implemented and tested using unaltered AutoDock4 with new forcefield tables.

Introduction

In physiological environments, proteins and other biological structures are surrounded by water molecules. When a small-molecule binds to a protein, it must displace most of the waters occupying the binding cavity. However, rarely are all water molecules displaced. A comprehensive analysis of 392 high-resolution crystal structures of proteins with interacting ligands¹ showed that in most of the complexes (>84%) one or more waters are present and interact with the ligand mediating its interaction with the protein. Some waters can be so strongly bound and conserved among similar proteins that from a ligand-docking perspective they are considered a part of the target structure, altering the binding site topography. Classical examples are HIV-1 protease² and acetylcholine receptors,³ where stable waters are targeted to increase inhibitor affinity^{2,4} or contribute to the pharmacophore definition.³ Stable waters can also be displaced to improve ligand affinity by the entropy gain resulting from the release of ordered water to the bulk solvent. These strategies have been successfully applied in designing scytalone dehydratase inhibitors⁵ and cyclic urea inhibitors of HIV-1 protease.⁶ Weakly bound waters are more likely to play varying roles depending upon the nature of the bound ligand. In fact, the same water can be stabilized by one ligand and displaced by another, as with poly(ADP-ribose) polymerase inhibitors.^{7,8} However water displacement does not always lead to affinity improvement.⁹ Thus the choice of which water to displace can be important for drug design.⁵ The usual protocol with most docking software, including AutoDock, is to remove all explicit waters from a protein structure before docking, then use an implicit solvent model in the form of a continuous

*To whom correspondence should be addressed: olson@scripps.edu, Phone: +1-(858)-784-9702. Fax: +1-(858)-784-2860.

Supporting Information Available

Target preparation and docking parameters; experimental data for complexes in training, test and cross-docking sets; details on protonation and scoring function issues. This material is available free of charge via the Internet at <http://pubs.acs.org>.

desolvation potential.^{10–12} For cases in which the presence of one or more waters is known to be relevant, multiple forms of the target can be modeled by including selected explicit waters.^{13–15} However keeping them in the same orientation for all ligands can produce a bias, which would be an issue for ligands binding with different water patterns. For example, the presence of structural water 301 in the HIV-1 protease does not allow cyclic urea inhibitors to dock correctly.¹² Moreover, it is a non-trivial task to compare results obtained with differing water occupancies, due to the difficulty in accounting for entropic contributions resulting from their displacement in the protein target.¹⁶ Another issue is the classification of waters that can be present in a crystal structure. Usually well-defined structural waters are included in the resolved structures, while weakly bound waters are more likely to be ignored. With the rise of computational docking as a common ligand screening methodology for drug design,¹⁷ the availability of a fast and accurate model for binding site hydration of multiple targets is crucial for binding energy estimation and result accuracy. Different qualitative and quantitative methods have been developed to predict the energetic contribution implied by presence or displacement of water molecules in protein structures.¹⁸ They can be divided into structure-based methods, where they rely on experimentally identified water positions (e.g. from high resolution crystal structures), and predictive, where their positions are determined by computational methods. The first category includes HINT/Rank (scoring function/geometric measurements¹⁹), WaterScore (multivariate regression²⁰), Consolv (k-nearest-neighborhood/GA²¹) and the method proposed by Barillari et al. (MC/replica exchange²²). The second category includes WaterMap (short molecular dynamics analysis²³), and JAWS (statistical thermodynamics¹⁰). The main disadvantage involving structure-based methods is that they are limited to waters from the *apo* (unbound) or available *holo* forms (ligands bound) of the protein. Hydration patterns can differ from *apo* to *holo* structure, or among multiple *holo* structures, where waters are stabilized by the ligand interactions themselves. An example is water 52 in Poly(ADP-ribose) polymerase structures.⁷ The quality of the results is sensitive to the crystal structure resolution¹⁹ or temperature factor,²⁰ and these methods cannot be applied to low resolution structures (>2.5 Å) or NMR models where waters are typically not resolved. Limited accuracy in predicting ligand-displaced waters has been also reported.^{22,24} Predictive methods have been reported to be more reliable^{10,22} and have been successfully applied to rationalize affinity data involving water displacement for drug design,^{10,22} but speed is traded for accuracy. Indeed, the determination of water positions and interaction energies requires complex calculations not suitable for high throughput deployment. Also, even small ligand modifications require the full calculation to be repeated. Within a docking context, several programs can handle explicit waters in the binding site during the ligand docking.²⁴ Water models include full atom representations (GOLD, GLIDE, SLIDE, DOCK)^{15,24,25} or spherical particles (FlexX, FITTED).²⁴ Most of these rely on position of predetermined and oriented waters, but the number of explicit waters that can be modeled is limited by the consequent increase in the number of degrees of freedom (e.g. 3 waters for GOLD).²⁴ Also, the energetics related to water displacement is often ignored in the scoring function (FlexX, GLIDE, DOCK).^{24,25} FlexX does not require waters to be present in the target structure, but accuracy in predicting their position and improvement in docking results have been reported to be very small.²⁴ Depending on the target, the inclusion of displaceable waters led to more consistent improvements with DOCK¹⁵ and GOLD.²⁶ However, the authors of the latter have speculated that inclusion of the switchable crystallographic waters can enrich results by reducing the search space,²⁷ and this could bias *de-novo* dockings with different ligands than those present in the experimental structures. AutoDock does not support explicit displaceable waters, although it has been successfully used to simulate the presence of known waters in specific targets.^{14,24,28}

The method we present here modifies the AutoDock 4.2 forcefield²⁹ to include explicit displaceable waters during docking to improve both docking accuracy and energy estimation

without excessive computational complexity. Instead of placing the waters in specific positions on the target surface, waters are attached to ligands before docking, and their presence is continuously evaluated during the search. If the mediation of a water molecule stabilizes the ligand-receptor interaction it is kept, otherwise it is displaced. Entropy and enthalpy are calculated separately for every water modeled. The balance between enthalpic and entropic contributions is evaluated to discern between displaced and conserved waters. The main advantage of the method is that no prior knowledge about water placement is required. The hydration of ligands instead of target structures is also convenient for several reasons. First, the number of waters binding to a ligand is limited and their positions are easier to predict than the complex arrangements of waters possible inside a given protein binding site. Second, the same binding site can present different hydration patterns when interacting with different ligands and indeed the presence (or absence) of a water molecule is often a consequence of the ligand binding. Moreover, not *every* water in a binding site will necessarily interact with *every* the ligands, and novel or structurally diverse ligands could interact through waters not present in the *apo* or known *holo* structures. Thus, this method represents an efficient way of modeling only waters directly interacting with the ligands to be docked. Water molecules are represented by a mono-atomic pseudo-atom (W) with properties designed to provide a simple description of water characteristics without compromising modeling accuracy. W atoms do not contribute to ligand intra-molecular interactions. The inter-molecular interactions of W atoms are described as a combination of hydrogen bond acceptor (OA) and donor (HD) properties. The force field parameters were calibrated on a training set of 197 ligand-protein crystallographic structures with resolution of 2.5 Å or better, including more than 50 different protein families. The forcefield was then validated on a test set of 220 complexes and cross-docking experiments. The interaction energy of conserved W atoms was also used to classify them as strongly or weakly bound. Docking results analysis and prediction of position and stability of known crystallographic waters are discussed in more detail for four case studies.

Methods

An AutoDock calculation is essentially a two-step process in which first the interactions between atom types in the ligand and the target structure are pre-calculated in a grid surrounding the binding region. Then during the docking, grid interaction energies are used as look-up table to speed ligand energy evaluations.³⁰ Among the different search methods available in AutoDock, the most efficient is the Lamarckian Genetic Algorithm (LGA).³¹ Our implementation enables the inclusion of explicit waters in docking by adding them as part of the ligand and using a special grid map calculated to describe water-protein interactions. All hydrogen bond donors and acceptor groups in the ligand are saturated with special W atoms placed along hydrogen bond vectors from the heavy atoms (Figure 1(a)). Bond angles are assigned according to experimentally determined values.^{32–35} A distance of 3.0 Å has been chosen as the most representative of the interactions found in crystallographic complexes. Each W atom is kept fixed during the entire calculation with respect to the hydrated ligand atom to which it is bound. If a hydrated atom is part of a flexible segment, the movement of the associated W atoms will be fixed in relation to it. Thus, no extra degrees of freedom are added because of the hydration process. Phosphate and sulphate oxygens were excluded from the hydration process due to the limitation of the rigid model in describing their variable hydration geometry. Hydrated ligands are then docked as regular ligands. The W atom represents a discreet spherical water molecule neutrally charged, presenting combined hydrogen and oxygen properties (Figure 1(b)). The spherical model avoids the extra degrees of freedom related to water hydrogen orientations with respect to both ligand and receptor. The ligand potential $\Delta G_{(lig)}$ for regular atoms is calculated in AutoDock according to eq. 1:

$$\begin{aligned}
\Delta G_{(lig)} = & \Delta H_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\
& + \Delta H_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\
& + \Delta H_{elec} \sum_{i,j} \left(\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) \\
& + \Delta S_{tor} N_{tor} \\
& + \Delta G_{desolv} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)}
\end{aligned} \tag{1}$$

The inter-molecular potentials are calculated by summations over all pairs of ligand atoms, i , and protein atoms, j , as a function of their distances, r . The potentials include a Lennard-Jones 12-6 dispersion/repulsion term (ΔH_{vdW}); a directional 12-10 hydrogen bonding term (ΔH_{hbond}), where $E(t)$ is the directional angle-based weight; and a Coulombic electrostatic potential (ΔH_{elec}) with a distance-dependent dielectric screening (ϵ). The potentials have been parametrized and optimized in earlier versions of the software.³⁶ Entropy of ligand binding (ΔS_{tor}) is included to account for the loss of degrees of freedom upon binding, that is proportional to the number of sp^3 bonds in the ligand (N_{tor}). The desolvation term ($\Delta G_{desolvation}$) is a function of the solvent accessible surfaces of ligand (S_i) and protein (S_j),³⁷ and accounts for the implicit bulk waters present in the docking volume. The W atom potential combines van der Waals and hydrogen bond potentials of hydrogen and oxygen atoms. In particular, the hydrogen bond term includes both acceptor and donor interactions. As a water molecule, a W atom is neutrally charged and the spherical model ignores the molecular dipole, consequently the electrostatic component ΔH_{elec} is neglected. Since the enthalpic contribution of water to desolvation is null, we also eliminate the ΔG_{desolv} . In our model, W atoms do not interact with each other nor with other ligand atoms (thus not limiting ligand flexibility), but if they overlap they are allowed to coalesce, summing their respective contributions to binding (see PR case study). The entropy contribution resulting from displacement of a water molecule upon binding is approximated as a constant (K_{wat}). Thus, the energy potential of W atoms is described by the sum of weighted van der Waals and hydrogen bond enthalpies and the desolvation entropy constant (eq. 2):

$$\begin{aligned}
\Delta G_{(wat)} = & \Delta H_{vdW} W_{wat} \sum_{w,j} \left(\frac{A_{wj}}{r_{wj}^{12}} - \frac{B_{wj}}{r_{wj}^6} \right) \\
& + \Delta H_{hbond} W_{wat} \sum_{w,j} E(t) \left(\frac{C_{wj}}{r_{wj}^{12}} - \frac{D_{wj}}{r_{wj}^{10}} \right) \\
& + \Delta S_{desolv} K_{wat}
\end{aligned} \tag{2}$$

where the W atom, w , substitutes the regular ligand atom, i , in eq. 1. Weight coefficient W_{wat} is an optimized parameter that scales W atom interaction with respect to regular atoms to include an entropy penalty to account for the loss of degrees of freedom of a ligand-bound water molecule. Finally, the total binding energy of a hydrated ligand is calculated from the sum of regular and W atoms potentials (eq. 3):

$$\Delta G = \Delta G_{(lig)} + \Delta G_{(wat)} \tag{3}$$

The grid map corresponding to interactions of W atoms with the protein is obtained by the contingent combination of oxygen-acceptor (OA) and hydrogen-donor (HD) grid maps. In particular, a W grid map point x_W will be equal to: a) the combination of acceptor (x_{OA}) and

donor (x_{HD}) enthalpies, if both are energetically favorable; or *b*) desolvation entropy if any clashes are found (eq. 4):

$$x_w = \begin{cases} \Delta H_{vdW} + \Delta H_{hbond} & x_{OA} < 0; x_{HD} < 0; \\ \Delta S_{desolv} & \text{otherwise.} \end{cases} \quad (4)$$

Several methods to mix and balance the acceptor and donor components components were tested. The one providing the best RMSD improvement was found to be the selection of the larger of the two absolute values (eq. 5):

$$x_w = \begin{cases} W_{wat}(x_{OA}) & \text{if } x_{OA} < x_{HD}; \\ W_{wat}(x_{HD}) & \text{if } x_{HD} < x_{OA}. \end{cases} \quad (5)$$

The resulting map represents the energy potential of a spherical water probe with both acceptor and donor properties. Characteristics such as directionality, strength and number of hydrogen bonds are encoded in the grid maps during receptor probing prior to docking. As depicted in Figure 2, the map is in very good agreement with experimental positions of water molecules. The determinant of water fate (bridging or displaced) is the balance between its energetic contributions to ligand binding and the stabilization of a given ligand pose through its displacement. The positional information of a given W atom during the docking simulation is used to establish which of the energetic contributions described in eq. 4 is considered for a particular pose, either the first term (protein-bridging) or second term (displacement). In particular, if one or more waters can stabilize a favorable ligand-receptor interaction, the water-receptor enthalpy is summed in the lig-and binding energy (Figure 3(a)). Conversely, when a tighter ligand-receptor interaction is more energetically favorable and requires a water displacement, the desolvation entropy resulting from its release to bulk solvent rewards ligand binding (Figure 3(b)). In this way, enthalpic (water conserved) and entropic (water displaced) contributions are continuously re-evaluated during the same docking run. Waters exposed to the solvent and not occupying any grid hot spots are considered bulk, and thus provide no energetic contribution.

Two sets of protein-ligand complexes were curated from the Protein Data Bank (PDB):^{38,39} more stringent resolution requirements have been used in the definition of a *training set*, where accurate water position placements were necessary for the calibration of weight coefficient (W_{wat}) and displacement entropy constant (K_{wat}) in the forcefield equation 2. Conversely, no resolution constraints were applied for the *test set* complexes. The goal was to test the general applicability of the method in a real-life scenario, where water positions are often not available. Finally, the forcefield has been tested on cross-docking experiments, where ligands have been docked to several conformations of the target protein different than the ligand-bound ones (i.e. *apo* conformations).

Training set

In order to establish a reliable reference set, complexes matching the following structural and experimental criteria were selected:

- X-ray diffraction with resolution ≤ 2.5 Å
- non-covalent ligand binding
- only one ligand in the binding site (including buffers, salt ions and solute molecules)
- no alternate or distorted configurations

- no crystallographic cell packing interactions with ligand atoms.

To limit the search space complexity, ligands with more than 12 rotatable bonds were excluded. The training set includes 197 complexes. Nine of these (*ladd*, *lbra*, *lhsl*, *ltng*, *ltnh*, *ltni*, *ltnl*, *2gbp*, *3ptb*) were also part of the AutoDock training set.¹² The training set is built to provide reasonable coverage of both ligand and binding site diversity. Training set complexes include more than 50 different protein classes, and ligand sizes that range from fragments (MW <250 da.) to drug-like molecules (MW up to 620 da.) with 0 to 12 rotatable bonds. The number of experimentally determined water molecules contained in grid boxes defining the binding sites ranges from 4 (*lgwx*) to 72 (*lc5p*). Coordinate files were analyzed to evaluate the role of well defined waters (occupancy ≥ 0.95) in ligand binding. Distance cutoffs were determined to include waters reported to interact with ligands. In particular, water oxygens within 3.7 Å from any ligand heavy atoms were considered “contact waters”; water oxygens within 3.55 Å from any ligand polar atoms were considered “bound waters”; bound waters within 3.55 Å or less from protein polar atoms were considered “bridging waters”.

Test set

More relaxed criteria were used in defining the test set, containing 221 complexes. The resolution was allowed to exceed the 2.5 Å cutoff; no rotatable bond limit was set; the occasional presence of molecules other than the ligand (i.e., co-factors, salt ions) was tolerated. Also, metal-coordinating complexes and mixed protein-DNA targets were considered only in the test set. Ligand molecular weights range from 79 to 667 da., rotatable bonds from 0 to 19. A summary of properties of training and test sets is given in Tables S1 and S2 in the Supporting information.

Ligand preparation and atom parameters

The crystallographic poses of ligands were extracted from the PDB structures, and hydrogens were manually added to match the deposited chemical structure and correct protonation state. Prior to docking, ligand translation, orientation and rotatable bonds angles were randomized, and the AutoDock input ligand format was generated following the standard procedure.⁴⁰ From these, hydrated versions of the ligands were generated by adding W atoms. The number of W atoms added per ligand ranges from 1 to 35 with an average of 9.6 for the training set.

Grid maps and docking calculation

Target protein structures were prepared following the standard AutoDock protocol⁴⁰ (see Supporting information for details). Only protein residues and enzyme co-factors were kept and all structural waters were removed. For every complex, the grid box was automatically placed at the geometric center of the experimental ligand pose, sized to encompass all ligand heavy atoms, and then expanded by 10 extra grid points (3.75 Å) in each direction. Oxygen-acceptor (OA) and hydrogen-donor (HD) were added if not already present in the ligand atom-type set. Grid maps were calculated for ligand atom types, and W maps generated according to equation 4. Search parameters were adapted to ligand complexity (see Supporting information). Every ligand was docked in its de-hydrated form with the standard forcefield, then docked in the hydrated form with the new forcefield, using the same search parameters in both cases. For every ligand complex, 100 docking poses were generated using the same search parameters. Results were clustered with 2.0 Å RMSD tolerance calculated on heavy atoms (excluding W atoms). The pose corresponding to lowest energy in the most populated cluster was selected as the docking result. We and others^{36,41–43} have found that clustering consistency is related to the conformational entropy of the system, and provides better estimations of the free energy of binding for configurations

corresponding to the experimental crystallographic structure.⁴² The docking was considered successful if the RMSD between the docking result and the experimental coordinates was ≤ 2.0 Å. Conserved W atoms were scored based on their corresponding grid affinity value. In particular, a W atom was considered a strongly bound water when affinity ≤ -0.5 Kcal/mol; a weakly bound water when affinity > -0.5 Kcal/mol and ≤ -0.3 Kcal/mol; and ignored for affinities > -0.3 Kcal/mol. The main goal of this work was not to predict the position of crystallographic waters *per-se*, but the presence and influence of those in close contact with docked ligands. Moreover, even for training set complexes it was not possible to assess the quality of the water placement due to the lack of density maps for most of them. Although, for systems where curated data was available (i.e. Case studies) and water positions were extensively evaluated in previous studies^{7,10,44,45} we reported the accuracy of in their predictions.

Results and discussion

Two key requirements were defined for this method to make it suitable for potential use in drug design and virtual screening. First, the forcefield must be generally applicable and be independent of any previous knowledge of water presence. Thus, it must cover a wide range of ligand sizes (from small fragments to drug-like molecules or peptidomimetics) and targets with different hydrophobic/hydrophilic properties with an arbitrary number of waters involved (including none). This first issue is addressed by the combined use of a ligand hydration model and grid maps to describe target water affinity. No previous knowledge is required because information on the potential presence and position of water molecules in the binding site can be obtained from grid analysis during the docking itself. Thus, the role of every water is evaluated on a per-ligand basis, with no bias from previously determined structures. Second, the method must be fast enough to be used with large number of ligands in a high throughput fashion. Performance has been measured during both the calibration on the training set and validation runs on the test set, and the computational overhead required by the new forcefield was very limited. Ligand hydration requires a negligible amount of time (fractions of second per ligand) and grid maps are already calculated for regular atoms. Since no extra degrees of freedom are added, the number of waters per ligand explored can be rather high (up to 35 in the training set and 23 in the test set). The docking computation scales nearly linearly with the increase in number of atoms evaluated by the scoring function at every step. Consequently, for complexes considered in this study, calculation time increases depending on the number of W atoms in every ligand, ranging from 0.5% (1 water: *2pcp*, *3gup*) to 73% (35 waters: *1rbo*) longer than the corresponding standard docking. On average, calculation times are about 30% longer.

Training set calibration of the terms of eq. 2 provided the optimal values for W_{wat} coefficient and ΔS_{desolv} term, respectively 0.6 and 0.2 Kcal/mol. This latter value corresponds roughly to one tenth of the upper limit of the entropy gain for transferring a water molecule from a protein to water reported by Dunitz.⁴⁶ The relative components of acceptor and donor character calculated accordingly to eq. 5 in the W maps of the training set varied in the following ranges: 70–97% OA; 30–3% HD. In the training set, the standard AutoDock forcefield docking was successful in 124 cases (62.9%) while the new hydrated docking was successful in 143 complexes (73.6%), giving a performance improvement of 17% (*p*-value: 0.007). Interestingly, in 123 complexes the calculated RMSD values were lower than those obtained with the standard protocol (Figure 4(a)). Improvements can be ascribed to enrichment of statistical relevance (clustering) of poses poorly represented in the standard protocol dockings (*1efy*, *3std*), or to the sampling of poses otherwise not energetically accessible (*1wcc*, *1n2v*, *2jjc*). Both can be traced back to the stabilizing effect of W atoms to ligand poses closer to experimental ones. No detriment of performance was found in both sets for cases where water molecules were not involved in docking. There was no significant

correlation between RMSD accuracy and the number of waters modeled. The new forcefield appears to be able to predict the buried waters but also superficial ones stabilized by ligand binding (*lefy*, *lyc4*). In some cases (e.g. *lu4d*, *lyc4*) the presence of simulated waters provides more subtle improvements like alignments of imidazole groups reflecting interactions present in the crystal structures. The increase in accuracy was demonstrated in the test set, where standard and hydrated dockings were successful in 135 (61.1%) and 161 (72.9%) cases, respectively, showing 19% performance improvement (*p*-value 0.001, Figure 4(b)).

Several cases (*luy7*, *luy8*, *luy9*, *2wi7*) where no improvement was found seem to be related to limitations of the scoring function. Four cases in the training set were identified as false negatives due to either flexible portion misalignment (*lefy*, *3ce0*, *ln2v*) or to non-directional interactions (i.e. van der Waals interactions: *lmg*), resulting in RMSD values higher than the cutoff despite reproducing of the correct binding mode. For a detailed summary of results of the training and test sets refer to Tables S1 and S2 in Supporting Information. In two cases in the test set (*2x8d*, *2f7x*), the new forcefield lead to RMSD >2.0 Å where the standard forcefield was successful. A different target protonation procedure restored accuracy in one cases (*2x8d*). The issue is further analyzed in the Supporting Information.

Fragment-sized ligands—Factors such as binding site topography and ligand shape can mitigate the effect of water molecules to ligand docking. From the analysis of complexes considered in this work, we note that within the same binding site, smaller ligands appear to be more likely to be influenced by waters. In particular, we found marked improvements in cases where waters are essential, such as fragment-size ligands (MW <250 da.). Fragment screening is a well established technique^{47,48} but their small size and low binding affinity make fragment hit identification difficult.⁴⁷ This challenge also exists in docking where the restricted number of interactions established with the target protein can limit scoring accuracy. For this reason, a single water molecule could have a dramatic influence in docking performance leading to either inaccurate scoring and/or incorrect pose. With hydrated docking, fragment-sized ligands show RMSD improvement in about 10% of the cases in both training and test sets, comparable to the overall sets, but the average RMSD improvement with respect to the standard docking is larger. This improvement is due to the number of cases where the correct binding location was found only in the hydrated docking (e.g. training set: *2jjc*, *lf3e*, *2wi2*, *lenu*, *lwcc*, *lwbu*; test set: *li80*, *3bsf*, *loty*). In Figure 5 are shown some representative docking results of fragment-size ligands.

Metal-coordinating complexes—The training set did not contain proteins with catalytic metals in the binding sites, while the test set contains 8 matrix metalloproteinases (MMP, *lbiw*, *ld8m*, *lg4k*, *lzp5*, *2yig*, *3ehx*, *3ehy*, *3kry*), 7 phosphodiesterases (PDE, *lxlx*, *lxm6*, *lxmu*, *3g45*, *3g4k*, *3hmv*, *3ly2*), and a protein-DNA intasome complex (*3oya*). Results showed no degradation of performance of the hydrated dockings with respect to the standard method, with in fact a significant improvement in one case (*3oya*, Table S2 in Supporting information).

Limitations—The model applies some approximations to simplify the implementation. The first is that only the first hydration shell is modeled. This is sufficient in most of the cases, but when more complex water networks involving second or third hydration shells are present (e.g. *2acj*), the new forcefield does not provide any advantage over the standard one. In addition, a direct consequence of W atoms being rigidity attached to ligand polar atoms is the difficulty of predicting waters whose position deviates significantly from ideal hydrogen bond geometry (*lb6l*:HOH320, *lc83*:HOH370) or which interact through van der Waals or electrostatic forces (*lpax*). Rigidity of the W atom position is also responsible for the drop

in accuracy with hydrated phosphate and sulphate oxygens (e.g. training: *lrbo*; test: *lxgi*) hence the choice to ignore them in the ligand hydration procedure; although it should be noted that the hydration of the other polar atoms in the ligands seemed to be sufficient to improve performance. Another limiting approximation is treating all W atoms equivalently, while a specific desolvation potential could lead to a more precise binding energy estimation.

External data sets and case study cross-docking validations—Table 1 shows a summary of external set and cross-docking validations. Standard and hydrated forcefields performed equally well when docking complexes in the Astex Diverse Set²⁷ (85 ligand-protein complexes) reproducing ligand binding poses within 2.0 Å RMSD in 78% of the cases. No performance degradation was found due to waters simulation. More challenging cross-docking experiments on the derived Astex Non-native set⁵⁰ (67 ligands, 1112 protein conformations) presented smaller but equally close success rate (standard: 40.4%; hydrated: 40.6%). Performance degradation was comparable to results reported in several cross-docking experiments by different Authors.^{12,50,51} The performance similarity of the two docking methods is compatible with limited involvement of water molecules reported for Astex set complexes.²⁷ In fact, cross-docking performance degradation has been ascribed mainly to large structural deviations from *holo* conformations in target proteins.⁵⁰ In these cases various degrees of success can be achieved with techniques for simulating protein flexibility,^{12,52,53} while water inclusion by itself provides negligible advantage. However, in complexes where water plays a crucial role in ligand binding,^{2–8} docking performance may benefit from modeling the placement of waters even among different protein conformations. To test this hypothesis we ran cross-docking experiments on some of the protein families discussed in detail below in the case studies. Two of them not present in the Diverse set,⁵⁰ had a sufficiently large number of protein conformations available from the PDB to perform cross-docking. Nicotine (**1**) and imidacloprid (**5**) were docked against 36 AChBP protein structures (34 *holo*, 2 *apo*). For PARP, compound **27** was docked against 17 protein structures (15 *holo*, 2 *apo*). For both targets, ligands with known common binding mode for all subtypes (AChBP: *L. stagnalis* and *A. californica* structures; PARP: subtypes 1 and 2) were selected for docking. Ligand coordinates were generated from SMILES strings deposited in the PDB entries, protonation states were visually inspected and corrected when necessary. Ligands were docked following the protocol described in the Methods section. Receptor structures were kept rigid during docking. Further details are available in Supporting Information.

AChBP binding site conformation can change dramatically, accommodating a wide variety of ligand sizes ranging from fragments to poly-peptides.⁵⁴ Most of the conformational changes occur in the Cys-loop⁵⁴ region. For both **1** and **5**, prediction of water positions improved the success rate, although to different extents (Table 1). In agreement with training and test set results, fragment-sized ligand **1** improved more dramatically (+30%) than larger **5** (+5%). **5** binds by interacting more extensively with the Cys loop therefore it is more sensitive to loop movements.⁵⁵ Docking with waters partially compensated for the loss of interactions due to protein movements. Prediction of a strongly bound water (water A, Scheme 1) was essential for the correct binding pose of both ligands, while other waters (water B and C) were predicted for **5** only when the Cys-loop conformation was compatible with their interaction. Dockings of **27** to the PARP structures showed the largest difference between standard and hydrated docking success rates (+76%), with the latter being able to reproduce the crystallographic pose in 15 out of 17 protein conformations. When binding, **27** establishes a hydrogen bond interaction with Glu335,⁵⁶ peculiar of PARP-1 and PARP-2, among few other receptor sub-types.⁵⁶ This interaction is mediated by a water molecule (water A, Scheme 4) described to be transient and ligand-dependent.⁷ Modeling this water was crucial in stabilizing the docking poses, and its position was consistently predicted

within 2 Å RMSD from the experimental coordinates in both subtypes (see Figure 6), including one *apo* conformation (2paw).

Water B was also predicted in most of the results, although with less accuracy than water A. Overall the cross-docking results confirm our findings from the calibration and validation dockings: *a*) simulation of water molecules with the hydrated docking method does not degrade results when waters are not involved in ligand binding; *b*) when waters are directly involved in ligand binding, modeling them can significantly improve the success rate, partially compensating for protein movements. A more detailed analysis of the role of water in AChBP and PARP structures is presented in the Case studies.

Case studies

Four protein families represented in both training and test sets were selected as case studies. Details of water predictions, scoring and ligand placement of representative complexes were analyzed and discussed.

Acetylcholine binding protein (1uw6, 2zju, 2zjv, 2byq, 2bys, 2wn9, 3c79, 3c84, 2wnl)—Rationalization of the activity of agonists of the nicotinic acetylcholine receptor (nAChR) has been an ongoing effort since 1970.⁵⁷ The pharmacophore of the nicotinic agonist contains a cationic group and a hydrogen bond acceptor (Scheme 1).⁵⁸ A conserved tryptophan was found to be involved in a cation- π interaction with the charged moiety,⁵⁹ while the elusive hydrogen bond-acceptor counterpart was eventually identified as a water molecule in crystallographic studies on homologous water-soluble acetylcholine binding proteins (AChBP).⁴⁵ This interaction pattern is conserved in receptors of the same family and among different organisms, including human neuronal receptors (nAChR).³ AChBP and AChR are pentameric structures.^{3,59} In this study we considered nine AChBP structures: three from *Lymnaea stagnalis* (training set: 1uw6; test set 2zju, 2zjv) and six from *Aplysia californica* (training set: 2bys, 2wn9; test set: 3c79, 3c84, 2wnl, 2byq). Structure 1uw6 contains **1** bound,⁵⁴ where the pyrrolidine charged group is interacting with Trp147 (cation- π and hydrogen bond interactions), and water HOH2063 (water A) mediates hydrogen bonds between Met114 and Leu102 backbone and the pyridine ring (Figure 7(c)). Both the standard and new hydration dockings can reproduce binding mode of **1** with satisfactory accuracy (1.18 Å and 0.42 Å, respectively, Figure 7). Although results obtained with the former in the absence of water, show a flipped orientation of the pyridine ring resulting in a weak interaction between the nitrogen and Tyr192 hydroxyl. Hydrated dockings reproduce the experimental orientation of pyridine ring and predict the position of the crystallographic water A with 0.9 Å RMSD. Scoring capabilities of the new forcefield were tested by docking **1** derivative s-MMP (N-methyl-2-phenylpyrrolidine,³ **2**) to the same structure (1uw6). Ligand **2** was been designed to bind in place of **1** in mutagenesis studies investigating the role of water.³ In fact, despite sharing the same binding mode, **2** is a weaker binder than **1**, lacking the nitrogen atom involved in the water-bridged interaction. Again, both forcefields are able to reproduce the binding mode, although in the absence of the water the standard forcefield does not correctly rank the two ligands (**2**>**1**). The new forcefield predicts the correct trend (**1**>**2**) providing an estimated energy contribution comparable to results obtained by standard dockings in the presence of the explicit water optimally oriented to maximize the hydrogen bond. A RMSD comparison of dockings on AChBP is summarized in Table 2. Improvement shown using the hydrated docking are larger for other ligands requiring water A or other waters present in the binding site which interact with the receptor, such as alkaloids anabaseine (**3**, 2wnl) and lobeline (**4**, 2bys), and neonicotinoids imidacloprid (**5**, 2zju) and chlotianidin (**6**, 2zjv). In particular, the standard docking was not able to reproduce the binding mode of the last class without crystallographic waters. The hydrated docking predicts the position of water A producing the correct binding mode

(Figure 8). For **5** the position of weakly bound water HOH337 (water B) bridging the interaction between a nitro oxygen and Glu190 is predicted within 2.0 Å RMSD. Another water (water C) was predicted to bridge the interaction with Gln55 and Lys34 side-chains. Due to the low crystallographic resolution of 2jzu, these two waters are not defined in all the pen-tamer binding sites, but they are identified separately at interfaces between chains A and B, and chains C and D, respectively. For the same reason, it was not possible to match positions of waters B and C with experimental waters in 2zjv. Position of water A was predicted with <1 Å RMSD approximation in all cases, including complexes where it was not identified due to very low resolution (2bys: 3.40 Å; 2wnl: 2.7 Å). By aligning⁶⁰ these results on the highest resolution structure (2wn9, 1.75 Å), the position of water A was predicted with less than 1 Å approximation. Results of the application of the hydrated docking to the AChBP family show high accuracy in recognizing the role of the nicotine receptor pharmacophoric water, improving ligand scoring. Notably, the structural water was correctly predicted to be always conserved with all ligands considered in the docking experiments. All predicted waters matched the experimentally determined positions in high resolution structures within a range of 2.0 Å. Moreover, we found that docking with W atoms is capable of predicting important water positions in low resolution structures, making it a powerful drug design tool. In fact, water A was also predicted with high consistency in different protein conformations as reported in the cross-docking results (Table 1).

HIV-1 protease (1b6l, 1kzk, 1mrw, 1ohr, 2zye, 1ajx, 1dmp, 5upj, 1hxx)—HIV-1 pro-tease (PR) is an essential enzyme for the reproductive cycle of HIV, the virus responsible of AIDS.⁶¹ Structural studies have shown that the activity of potent PR inhibitors, like KNI-272⁴⁴ (**7**, Scheme 2) and similar binders (**8–11**), is mediated by structural water HOH301² (water A), which establishes a hydrogen bond bridge between the ligand and the backbone of Ile50A and Ile50B residues in the flaps that cover the active site.⁴ The position of water A has been also exploited to design cyclic urea inhibitors (**12–14**) able to bind by taking advantage of its displacement entropy.⁶ Other stable waters surrounding **7** are responsible for stabilizing ligand binding, in particular HOH566 (water B) and HOH608 (water C)⁴ and can be potentially displaced for increased ligand affinity.⁴ Cross-docking experiments with PR show that while unhydrated structures can be used for docking both ligand classes, the standard protocol fails on docking cyclic urea derivatives in presence of water A (RMSD >5.0 Å).¹² Therefore for accurate energy estimation, water A should be kept in PR structures used to dock ligands that bind via its mediation (i.e. **7**), while unhydrated structures should be used for docking ligands that displace it (e.g. cyclic urea derivatives). Predicting the presence or the absence of waters in the PR binding site can thus be a crucial point for obtaining correct docking poses and affinity estimation of inhibitors. Docking results and ligand-bound predicted waters are summarized in Table 3. The training set contained 4 complexes of PR with inhibitors, all binding via water A: **8** (1b6l), **9** (1kzk), **10** (1mrw), **11** (1ohr) (Scheme 2). In all cases, both standard and hydrated dockings provided correct placement for all ligands within <2.0 Å RMSD from the crystallographic pose. The absence of waters did not affect the standard dockings, because the size of the ligands and the binding site topography are sufficient to induce the correct poses. Although, by lacking HOH301 none of the ligands are able to interact with the flap residues. Hydrated dockings show a subtle RMSD improvement and predict a water chelated by conserved ligand oxygens and bridging interaction with the flaps. This water results from the coalescence of two W atoms bound to carbonyl oxygens on the ligand backbone, mimicking the chelating geometry found in the experimental structure. The distance between the predicted water and the experimental position is <2.0 Å RMSD as shown on Figure 9. To test the accuracy of the prediction of other waters, **7** was re-docked in the neutron diffraction structure of the complex with the PR (2zye).⁴ Four waters were identified in the dockings as being involved in ligand-protein interaction, matching with very high

accuracy the positions of stable waters A, B, C, and HOH322 (water D) with RMSD of 0.9 Å, 0.9 Å, 1.2 Å and 0.7 Å, respectively (Figure 9). A weakly bound water was predicted to be bound to Gly27B and Asp25A. Its position differs by 0.9 Å from that of water HOH607 (water E) bound to two-fold related Gly27A and Asp25B (Figure 9). Notably, no complexes with ligands displacing water A are present in the training set, but the new hydrated docking predicts its displacement in three complexes from the test set: two structurally similar cyclic urea inhibitors AHA001⁶² (**12**, *1ajx*), DMP450⁶³ (**13**, *ldmp*) and a 4-hydroxycumarin derivative U99283⁶⁴ (**14**, *5upj*). Results are comparable to the standard docking in reproducing the experimental poses within 1 Å RMSD. Water A is displaced and its position is correctly occupied by the ligand oxygens in all cases. No waters were found to mediate interactions of **12** and **14**, in accord with experimental results. On the other hand, two waters were predicted to bridge interactions with aniline nitrogens of **13**, although no waters were reported in the crystal structure. Due to the high structural similarity between ligands and protein conformation,⁶⁵ positions of waters present in *1ajx* were used to measure RMSDs of predicted waters in *ldmp* upon alignment.⁶⁰ Waters predicted with **13** correspond to those bound to Gly48 (water F, 1.7 Å RMSD) and Asp29 (water G, 1.1 Å RMSD). Thus, we find in a strong agreement with the experimental data on PR inhibitors that hydrated docking can predict either the position of structural water A or its displacement, depending on the nature of the ligand. Additionally, calculations on different ligands have shown the possibility of predicting position and classification of extra waters (B-E) with very good accuracy (<2.0 Å). All predicted water positions matched the experimentally determined positions within 2.0 Å. Only water B bound to **8** (*1b6l*:W320) was not predicted due to its deviation from ideal hydrogen bond geometry. Results of this case study confirm that hydrated docking can also be applied predictively to structures where crystallographic resolution is not sufficient to determine their position.

Scytalone dehydratase (3std, 4std, 5std, 6std, 7std)—Scytalone dehydratase (SD) is an essential enzyme of the biosynthetic pathway of the fungal plant pathogen *Magnaporthe grisea* infecting rice crops.⁶⁶ The absence of this biosynthetic pathway in the host organism and mammals makes SD an interesting target for fungicide design.⁶⁶ In the binding site, two conserved water molecules establish hydrogen bond networks with Tyr30/Tyr50 (water A) and His85/His110 (water B). Water displacement has been exploited in designing new inhibitors derived from a salicylamide inhibitor **15**⁴⁸ (Scheme 3). Five SD-inhibitors, **15** (*4std*), **16** (*5std*), **17** (*6std*) and **18** (*7std*), bind with both conserved waters A and B, while the cyano-cinnoline derivative (**19**, *3std*) is designed to displace water A. The energetics of this displacement upon ligand binding have been investigated by Michel et al. by means of statistical thermodynamics.¹⁰

Docking results and ligand-bound predicted waters are summarized in Table 4. Without waters in the binding site, the standard forcefield docking fails to reproduce the binding pose in two cases (**15**, **19**, Figure 10). In **15**, the small salicylamide ligand overlaps water A with the amide oxygen. In **19** the lack of the water B prevents the correct alignment of the amino-biaryl group of the cyano-cinnoline inhibitor. Standard docking in the presence of both waters results in >6 Å RMSD for **19**. Conversely, all hydrated dockings were successful. Water positions were predicted with deviations <1.5 Å in all structures, helping to orient ligands correctly in the binding site. In particular, when docking **19**, the hydrated docking is able to distinguish between the two waters, predicting water A to be displaced by the cyano group and water B to be conserved. It is interesting to note that protein affinity map values calculated for both waters are very similar, but the choice of which one to displace is made on the basis of the nature of the ligand, in agreement with Michel et al.¹⁰ One of the assumptions upon which the hydrated docking is based is that different ligands can bind the same target through mediation of different waters. Indeed, docking of SD inhibitors demonstrates that the new hydrated forcefield is able to predict water displacement when it

is more favorable for binding, in agreement with experiment. The process is handled in an automated manner, calculating the balance between entropic and enthalpic contributions during the docking.

Poly(ADP-ribose) polymerase (1pax, 2pax, 3pax, 1efy, 1wok, 1uk1, 2rcw, 2rd6, 3c49, 3c4h, 3ce0, 3fhb, 3gfw, 3gn7, 3goy, 3hkv, 3kcz, 3kjd, 3l3l, 3l3m, 3p0n, 3p0p, 4pax)—Poly(ADP-ribose) polymerase (PARP) is a family of enzymes involved in the process of post-translational modifications triggered by DNA damage. They are involved in many major pathological conditions such as cancer, ischemia, inflammation and diabetes,^{7,67} making them a very important target for drug design.⁶⁸ Moreover, specificity among the different sub-types is critical to their therapeutic application.⁶⁸ The activity of known inhibitors involves formation of a hydrogen bond with Gly863. The training set contains complexes of 4 ligands bound to PARP1 while the test set contains 18 complexes of six receptor subtypes (1,2,3,4,10 and 14), Scheme 4. Over training and test sets the ligands docked correctly using both methods, except in cases where waters are directly involved in ligand binding, like the benzimidazole inhibitor NU1098⁶⁹ (**20**, *1efy*, Figure 11) from the training set. In the crystal structure the interaction of **20** is mediated by two weakly bound waters HOH52 (water A) and HOH107 (water B) bridging the benzimidazole ring with the protein, plus at least 8 extra waters filling the cavity between Tyr899 and His862. When no waters were placed in the binding site, the standard docked ligand shifted to maximize contacts with the protein surface (4.6 Å RMSD). Conversely, the hydrated docking was able to reproduce the correct alignment of the ligand (2.08 Å RMSD) predicting the presence of waters A (0.6 Å RMSD) and B (2.2 Å RMSD), and two of the waters filling the underlying cavity, waters C and D (1.0 Å and 2.3 Å RMSD, respectively), ranking them correctly. The relatively high resulting RMSD arises from a different alignment of the methoxy-phenyl group, whose methyl overlaps with one of the waters in the underlying cavity (water 82). In the crystallographic pose, waters C and D are far from the methoxy-oxygen (~ 4 Å) that then rotates to reach them. The role of weakly bound waters appears to be important for the ligand binding. In a systematic study of PARP1 inhibitors, Bellocchi et al.⁷ suggested that water positions are stabilized by the ligand presence itself while they influence its activity by limiting desolvation energy penalties.⁷ Analysis of crystal structures of PARP1 supports this hypothesis linking the presence of water A to the characteristics of the ligands bound. Water A is not resolved in the *apo* structure (*2paw*), where it would be weakly bound and surface exposed. Then, depending on the ligand, it is stabilized by **21** (*1pax*), **22** (*3pax*), **28** (*2rd6*) and **23** (*3l3m*) or displaced by **24** (*2pax*). The new hydrated docking predicts weakly bound waters for ligands capable of binding them, in particular improving results of ligands structurally related to **20** and binding via the water A (**23**, **26**, **27**). The absence of water A is correctly predicted with benzindol-3-one inhibitor **25** (*3l3l*) because the ligand lacks the corresponding nitrogen responsible of stabilizing its presence in **20**. Other waters (water C, D, E) matching experimental coordinates are predicted to mediate the interactions of different ligand substituents with the protein (Figure 11). On the other hand, naphthyl-amino group of **24** is correctly predicted to displace water A, while water B is conserved. Predictions of the position of waters was tested on a low resolution (3.0 Å) structure of PARP complexed with **26** (*1wok*), a quinoxaline derivative of **20**. Using the reference position of waters from *1efy*,⁷⁰ the predicted waters are very close to the experimental positions of waters A (1.0 Å) and B (1.5 Å) and one approximately corresponding to water C (2.7 Å). Despite the fact that positional accuracy in predicting weakly bound waters is lower than that achieved with stable waters, determining their presence is valuable information for understanding and improving estimates of ligand binding energetics. Dockings of PARP inhibitors show the capabilities of the new hydration method to predict the position of weakly bound waters and rank them in accordance with experimental data, improving the docking accuracy. In

particular, transient waters stabilized by ligand binding are predicted with good accuracy (<2.0 Å). As shown in cross-docking results (Figure 6), weakly bound water A was predicted also among different protein conformations and subtypes, providing a dramatic improvement in docking results (Table 5). It must be noted that when **21** is bound (*Ipax*), water A seems to be stabilized by weak van der Waals interactions with a nearby ligand methyl group, but due to the nature of the interaction itself the hydration forcefield is not able to predict its presence.

Conclusions

We have developed a forcefield and a ligand hydration method capable of substantially improving docking results by predicting position and stability of water molecules during dockings. The force-field has been implemented by extending the standard AutoDock forcefield to include a spherical water model and a potential to estimate enthalpies and entropies resulting from conserved or displaced waters, respectively. Waters are attached to ligands before docking by hydrating all polar groups capable of hydrogen bond interactions. The forcefield separately calculates entropy and enthalpy for every water, enabling a fine-grained estimation of their contributions in ligand-protein interaction. Their state is evaluated dynamically at every step of the docking process. The energetic description does not imply a hard steric wall nor a binary state switch but models a smooth potential easily sampled during the conformational search. This new method allows sampling different hydration patterns for the same protein depending on the ligand docked. In case studies for which detailed information on important crystallographic waters is available, dockings performed with the new hydration forcefield are able to find experimentally determined waters and predict their stability using grid-based calculations. The new forcefield combines the two approaches of predicting water positions and use them during docking. As demonstrated in case studies and cross-dockings, the fact that the forcefield does not require previous knowledge of waters presence makes it suitable when information on ligand-induced hydration patterns is unavailable. No detriment of performance was found where water molecules were not involved in docking. Calibration of the parameters on 197 complexes of the training resulted in an accuracy increase of 10% in training set (performance increase: 17%) compared with the standard docking. Validation of the forcefield on a test set of 220 complexes showed a 11.7% improvement in results (performance increase: 19%). General RMSD improvements are distributed across the entire sets, while the results are more dramatic in cases where waters are important for ligand placement, such as with small fragments. Calibration and testing have been performed on diverse systems with no bias toward the presence of waters in ligand binding. The method has been further validated with cross-docking experiments confirming both general applicability and performance improvement when waters are directly involved in ligand binding. The maximum number of waters modeled per ligand was rather large (up to 35). Conversely, the computational effort required to account for waters during the docking is relatively small. The method's simple implementation, involving no source code changes to AutoDock,⁷¹ and speed make it suitable for virtual screening application. In addition to ligand RMSD improvements, the new hydrated docking method provides predictions on the presence of water molecules in the binding site and the interactions they mediate. This is important information in de-novo or early stages of drug design, to optimize ligands to better fit the binding site. Also due to its hydrogen bond bi-valency, a water molecule can also 'invert' a receptor's hydrogen bond acceptor region into a donor. The forcefield is able to model this behavior, thereby sampling a broader bio-isosteric spectrum. Examples reported in the case studies and cross-dockings confirm the hydrated docking's ability to predict structurally conserved waters and ligand-stabilized waters that are not present in the *apo* structure. The predictive capabilities of the method have been tested also by comparing results obtained in low resolution structures with corresponding high resolution ones.

Results showed that it is possible to use hydrated force-field dockings successfully to predict water positions when experimental conditions do not allow to resolve their position. Some limitations are present in the forcefield, which will also be addressed in future work. In particular, flexible W atoms or a soft potential might compensate for non-ideal water placement and phosphate/sulphate hydration shells. Energy estimation can be further improved by modeling specific hydration potentials for different chemical groups. Scoring might be improved by applying the new forcefield to explicitly hydrated target structures, increasing the desolvation potential accuracy and providing a model for waters displaced by non-polar group. This docking method can be used as an exploratory tool for probing the presence of water molecules in binding sites, being able to predict weakly bound waters in *apo* structures. Results obtained with this approach can provide valuable information to guide ligand design efforts. The next development of the method is to test its applicability in large virtual screening campaigns.

Due to the customizable features of AutoDock, the full protocol has been implemented with only parameter files and a few helper Python scripts, with no changes to the source code. The code for preparing and analyzing dockings will be made available via the internet at <http://autodock.scripps.edu>. All figures have been generated with Python Molecular Viewer v1.5.6.⁷²

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

S.F. thanks Dr. Luca Bellucci for the inspiring discussion, Prof. David Goodsell, Dr. Fabrizio Manetti, Dr. Garrett Morris and Dr. Oleg Trott for their valuable suggestions. We acknowledge funding support on NIH grant R01 GMO69832 (to A.J.O.). This is publication number 21222 of the Scripps Research Institute. During the revision of this manuscript, a paper using a similar approach has been published.⁷³ A brief comparison of the two methods is provided in Supporting information.

Abbreviations

LGA	Lamarckian Genetic Algorithm
PARP	poly(ADP-ribose)
RMSD	root mean square deviation
MC	Monte Carlo simulation
PDB	Protein Data Bank
AChBP	acetylcholine binding protein
nAChR	neuronal Acetylcholine receptor
s-MMP	<i>N</i> -methyl-2-phenylpirroline
PR	HIV-1 protease
SD	scytalone dehydratase
MMP	matrix metalloproteinase
PDE	phosphodiesterase
OA	oxygen acceptor probe
HD	hydrogen donor probe

Notes and References

1. Lu Y, Wang R, Yang CY, 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. [PubMed: 17266298]
2. Baldwin ET, Bhat TN, Gulnik S, Liu B, Topol IA, Kiso Y, Mimoto T, Mitsuya H, Erickson JW. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure.* 1995; 3:581–590. [PubMed: 8590019]
3. Blum AP, Lester HA, Dougherty DA. Nicotinic pharmacophore: the pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proc Natl Acad Sci U S A.* 2010; 107:13206–13211. [PubMed: 20616056]
4. Adachi M, Ohhara T, Kurihara K, Tamada T, Honjo E, Okazaki N, Arai S, Shoyama Y, Kimura K, Matsumura H, Sugiyama S, Adachi H, Takano K, Mori Y, Hidaka K, Kimura T, Hayashi Y, Kiso Y, Kuroki R. Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc Natl Acad Sci U S A.* 2009; 106:4641–4646. [PubMed: 19273847]
5. Chen JM, Xu SL, Wawrzak Z, Basarab GS, Jordan DB. Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry.* 1998; 37:17735–17744. [PubMed: 9922139]
6. Lam PY, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science.* 1994; 263:380–384. [PubMed: 8278812]
7. Bellocchi D, Macchiarulo A, Costantino G, Pellicciari R. Docking studies on PARP-1 inhibitors: insights into the role of a binding pocket water molecule. *Bioorg Med Chem.* 2005; 13:1151–1157. [PubMed: 15670923]
8. García-Sosa AT, Firth-Clark S, Mancera RL. 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. [PubMed: 15921452]
9. Wissner A, Berger DM, Boschelli DH, Floyd MB, Greenberger LM, Gruber BC, Johnson BD, Mamuya N, Nilakantan R, Reich MF, Shen R, Tsou HR, Upeslakis E, Wang YF, Wu B, Ye F, Zhang N. 4-Anilino-6,7-dialkoxyquinoline-3-carbonitrile inhibitors of epidermal growth factor receptor kinase and their bioisosteric relationship to the 4-anilino-6,7-dialkoxyquinazoline inhibitors. *J Med Chem.* 2000; 43:3244–3256. [PubMed: 10966743]
10. Michel J, Tirado-Rives J, Jorgensen WL. Prediction of the water content in protein binding sites. *J Phys Chem B.* 2009; 113:13337–13346. [PubMed: 19754086]
11. Mysinger MM, Shoichet BK. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model.* 2010; 50:1561–1573. [PubMed: 20735049]
12. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009; 30:2785–2791. [PubMed: 19399780]
13. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins.* 2002; 46:34–40. [PubMed: 11746701]
14. Moitessier N, Westhof E, Hanessian S. Docking of aminoglycosides to hydrated and flexible RNA. *J Med Chem.* 2006; 49:1023–1033. [PubMed: 16451068]
15. Huang N, Shoichet BK. Exploiting ordered waters in molecular docking. *J Med Chem.* 2008; 51:4862–4865. [PubMed: 18680357]
16. Roberts BC, Mancera RL. Ligand-protein docking with water molecules. *J Chem Inf Model.* 2008; 48:397–408. [PubMed: 18211049]
17. Ripphausen P, Nisius B, Peltason L, Bajorath J. Quo vadis, virtual screening? A comprehensive survey of prospective applications. *J Med Chem.* 2010; 53:8461–8467. [PubMed: 20929257]
18. Wade R, Goodford P. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups

- with the ability to form more than two hydrogen bonds. *Journal of medicinal chemistry*. 1993; 36:148–156. [PubMed: 8421281]
19. Amadasi A, Surface JA, Spyraakis F, Cozzini P, Mozzarelli A, Kellogg GE. Robust classification of “relevant” water molecules in putative protein binding sites. *J Med Chem*. 2008; 51:1063–1067. [PubMed: 18232647]
 20. García-Sosa AT, Mancera RL, Dean PM. 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. [PubMed: 12756610]
 21. Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J Mol Biol*. 1997; 265:445–464. [PubMed: 9034363]
 22. Barillari C, Taylor J, Viner R, Essex JW. Classification of water molecules in protein binding sites. *J Am Chem Soc*. 2007; 129:2577–2587. [PubMed: 17288418]
 23. Young T, Abel T, Kim B, Berne BJ, Friesner RA. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *PNAS*. 2007; 104:808–813. [PubMed: 17204562]
 24. de Beer SBA, Vermeulen NPE, Oostenbrink C. The role of water molecules in computational drug design. *Curr Top Med Chem*. 2010; 10:55–66. [PubMed: 19929830]
 25. Villacanas O, Madurga S, Giralt E, Belda I. Explicit Treatment of Water Molecules in Protein-Ligand Docking. *Current Computer-Aided Drug Design*. 2009; 5:145–154.
 26. Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, Nissink JWM, Taylor RD, Taylor R. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem*. 2005; 48:6504–6515. [PubMed: 16190776]
 27. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, Murray CW. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem*. 2007; 50:726–741. [PubMed: 17300160]
 28. Minke WE, Diller DJ, Hol WG, Verlinde CL. The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: heat-labile enterotoxin, a multivalent test case. *J Med Chem*. 1999; 42:1778–1788. [PubMed: 10346930]
 29. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem*. 2007; 28:1145–1152. [PubMed: 17274016]
 30. Goodsell D, Olson A. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics*. 1990; 8:195–202.
 31. Morris G, Goodsell D, Halliday R, Huey R, Hart W, Belew R, Olson A. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*. 1998; 19:1639–1662.
 32. Wiberg K, Marquez M, Castejon H. Lone pairs in carbonyl-compounds and ethers. *Journal of Organic Chemistry*. 1994; 59:6817–6822.
 33. Taylor R, Kennard O, Versichel W. Geometry of the n-h= π c hydrogen-bond .1. lone-pair directionality. *Journal of the American Chemical Society*. 1983; 105:5761–5766.
 34. Nobeli I, Price SL, Lommerse JPM, Taylor R. Hydrogen bonding properties of oxygen and nitrogen acceptors in aromatic heterocycles. *Journal of Computational Chemistry*. 1997; 18:2060–2074.
 35. Lommerse JPM, Price SL, Taylor R. Hydrogen bonding of carbonyl, ether, and ester oxygen atoms with alkanol hydroxyl groups. *Journal of Computational Chemistry*. 1997; 18:757–774.
 36. Morris G, Goodsell D, Huey R, Olson A. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design*. 1996; 10:293–304. [PubMed: 8877701]
 37. Mehler E, Solmajer T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein engineering*. 1991; 4:903. [PubMed: 1667878]
 38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
 39. [Accessed: 08/09/2011.] The Protein Data Bank. <http://www.pdb.org/>

40. Morris GM, Goodsell DS, Pique ME, Lindstrom W, Huey R, Forli S, Hart WE, Hallyday S, Belew R, Olson AJ. AutoDock 4.2 User Guide. 2009
41. Illingworth CJR, Morris GM, Parkes KEB, Snell CR, Reynolds CA. Assessing the role of polarization in docking. *J Phys Chem A*. 2008; 112:12157–12163. [PubMed: 18986122]
42. Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin Drug Discov*. 2010; 5:597–607. [PubMed: 21532931]
43. Chang M, Belew R, Carroll K, Olson A, Goodsell D. Empirical entropic contributions in computational docking: Evaluation in APS reductase complexes. *Journal of computational chemistry*. 2008; 29:1753–1761. [PubMed: 18351616]
44. Kiso Y, Matsumoto H, Mizumoto S, Kimura T, Fujiwara Y, Akaji K. Small dipeptide-based HIV protease inhibitors containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic. *Peptide Science*. 1999; 51:59–68. [PubMed: 10380353]
45. Brejc K, van Dijk WJ, Klaassen RV, Schuurmans M, van Der Oost J, Smit AB, Sixma TK. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*. 2001; 411:269–276. [PubMed: 11357122]
46. Dunitz JD. The entropic cost of bound water in crystals and biomolecules. *Science*. 1994; 264:670. [PubMed: 17737951]
47. Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov*. 2007; 6:211–219. [PubMed: 17290284]
48. Chen Y, Shoichet BK. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat Chem Biol*. 2009; 5:358–364. [PubMed: 19305397]
49. Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*. 1996; 38:305–320. [PubMed: 8906967]
50. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW. Protein-ligand docking against non-native protein conformers. *J Chem Inf Model*. 2008; 48:2214–2225. [PubMed: 18954138]
51. Tuccinardi T, Botta M, Giordano A, Martinelli A. Protein kinases: docking and homology modeling reliability. *J Chem Inf Model*. 2010; 50:1432–1441. [PubMed: 20726600]
52. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc*. 2002; 124:5632–5633. [PubMed: 12010024]
53. Bottegoni G, Kufareva I, Totrov M, Abagyan R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J Med Chem*. 2009; 52:397–406. [PubMed: 19090659]
54. Celie PHN, van Rossum-Fikkert SE, van Dijk WJ, Brejc K, Smit AB, Sixma TK. Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron*. 2004; 41:907–914. [PubMed: 15046723]
55. Talley TT, Harel M, Hibbs RE, Radic Z, Tomizawa M, Casida JE, Taylor P. Atomic interactions of neonicotinoid agonists with AChBP: molecular recognition of the distinctive electronegative pharmacophore. *Proc Natl Acad Sci U S A*. 2008; 105:7606–7611. [PubMed: 18477694]
56. Karlberg T, Hammarström M, Schütz P, Svensson L, Schüler H. Crystal structure of the catalytic domain of human PARP2 in complex with PARP inhibitor ABT-888. *Biochemistry*. 2010; 49:1056–1058. [PubMed: 20092359]
57. Beers WH, Reich E. Structure and activity of acetylcholine. *Nature*. 1970; 228:917–922. [PubMed: 4921376]
58. Glennon RA, Dukat M. Central nicotinic receptor ligands and pharmacophores. *Pharm Acta Helv*. 2000; 74:103–114. [PubMed: 10812946]
59. Xiu X, Puskar NL, Shanata JAP, Lester HA, Dougherty DA. Nicotine binding to brain receptors requires a strong cation- π interaction. *Nature*. 2009; 458:534–537. [PubMed: 19252481]
60. Martínez L, Andreani R, Martínez JM. Convergent algorithms for protein structural alignment. *BMC Bioinformatics*. 2007; 8:306. [PubMed: 17714583]
61. Brik A, Wong CH. HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem*. 2003; 1:5–14. [PubMed: 12929379]

62. Bäckbro K, Löwgren S, Österlund K, Atepo J, Unge T, Hultén J, Bonham N, Schaal W, Karlén A, Hallberg A. Unexpected binding mode of a cyclic sulfamide HIV-1 protease inhibitor. *Journal of medicinal chemistry*. 1997; 40:898–902. [PubMed: 9083478]
63. Hodge C, Aldrich P, Bacheler L, Chang C, Eyermann C, Garber S, Grubb M, Jackson D, Jadhav P, Korant B. Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chemistry & Biology*. 1996; 3:301–314. [PubMed: 8807858]
64. Romines K, Watenpaugh K, Tomich P, Howe W, Morris J, Lovasz K, Mulichak A, Finzel B, Lynn J. Use of medium-sized cycloalkyl rings to enhance secondary binding: Discovery of a new class of human immunodeficiency virus (HIV) protease inhibitors. *Journal of medicinal chemistry*. 1995; 38:1884–1891. [PubMed: 7783120]
65. 0.55 Å RMSD on protein backbone.
66. Valent B, Farrall L, Chumley FG. Magnaporthe grisea genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics*. 1991; 127:87–101. [PubMed: 2016048]
67. Delaney CA, Wang LZ, Kyle S, White AW, Calvert AH, Curtin NJ, Durkacz BW, Hostomsky Z, Newell DR. Potentiation of temozolomide and topotecan growth inhibition and cytotoxicity by novel poly(adenosine diphosphoribose) polymerase inhibitors in a panel of human tumor cell lines. *Clin Cancer Res*. 2000; 6:2860–2867. [PubMed: 10914735]
68. Yap TA, Sandhu SK, Carden CP, de Bono JS. Poly(ADP-ribose) polymerase (PARP) inhibitors: Exploiting a synthetic lethal strategy in the clinic. *CA Cancer J Clin*. 2011; 61:31–49. [PubMed: 21205831]
69. White A, Almassy R, Calvert A, Curtin N, Griffin R, Hostomsky Z, Maegley K, Newell D, Srinivasan S, Golding B. Resistance-modifying agents. 9. Synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme poly (ADP-ribose) polymerase. *Journal of medicinal chemistry*. 2000; 43:4084–4097. [PubMed: 11063605]
70. 0.9 Å RMSD on protein backbone.
71. Although, minor changes to the source code can further improve efficiency by excluding W atoms from the ligand intra-molecular interaction calculation, removing almost entirely the overhead of simulating waters.
72. Sanner MF. Python: a programming language for software integration and development. *J Mol Graph Model*. 1999; 17:57–61. [PubMed: 10660911]
73. Lie MA, Thomsen R, Pedersen CNS, Schiøtt B, Christensen MH. Molecular docking with ligand attached water molecules. *J Chem Inf Model*. 2011; 51:909–917. [PubMed: 21452852]

Figure 1.
Representation of the hydration procedure.

Figure 2. Volumetric rendering of the calculated W affinity map (HIV protease, 2zye). Ligand coordinates of **7** (white) and waters (red) are shown as ball and sticks. Predicted positions for strongly bound waters are highlighted with transparent red volumes.

Figure 3. Energetic contributions in modeling bridging (a) or displaced (b) waters.

(a) Training set (197 complexes)

(b) Test set (220 complexes)

Figure 4. RMSD plots of hydrated over standard docking for training and test sets. RMSD values are in Å and points are colored by molecular weight. For each quadrant, the number of items is reported.

Figure 5.

Comparison of fragments docking with the standard (*a*, *c*, *e*) and the new hydrated forcefield (*b*, *d*, *f*). Crystallographic position of ligands are shown as purple sticks; binding site residues are shown in green (sticks or MSMS surfaces⁴⁹). Waters are shown as spheres red (experimental) or cyan (predicted).

Figure 6. Representation of 15 successful cross-docking results of **27**. Green spheres show a 2.0 Å range around experimental coordinates of waters A and B. The rightmost red sphere is water E (Scheme 4), predicted only in one docking.

Figure 7. Comparison of experimental structure of **1** and dockings obtained with standard and hydrated forcefields (*Iuw6*). Docked poses are colored in yellow; experimental pose is in gray.

Figure 8.

Comparison of experimental structure of **5** and dockings obtained with standard and hydrated forcefields (*2zju*). Experimental structure is colored in purple; docked poses are colored by atom type.

Figure 9.

Comparison of hydrated docking pose (RMSD 0.65 Å) and waters prediction with the neutron diffraction structure of **7** complexed in 2zye. Experimental coordinates are in gold sticks, docked ones in red. Experimental waters are shown as stick water molecules, predicted water positions are shown as green spheres. The red sphere water shows the predicted two-fold symmetry position of water E (see the text).

Figure 10.

Comparison of dockings of **19** with the standard and hydrated methods (*a*, *b*), and water predictions obtained by docking **15** (*c*). The crystallographic poses are represented as purple sticks; protein residues are as green sticks; docked results are as gray sticks; crystallographic (red) and predicted waters (cyan) are shown as spheres.

Figure 11.
PARP inhibitor **20** dockings with the standard and hydrated methods.

Scheme 1.
AChBP inhibitor structures with experimentally determined ligand-bound waters.

Scheme 2.
HIV-1 PR inhibitor structures with experimentally determined ligand-bound waters.

Scheme 3.

Scytalone dehydratase inhibitor structures with experimentally determined ligandbound waters.

Scheme 4.
PARP inhibitor structures with experimentally determined ligand-bound waters.

Table 1

Success rates of external set dockings and cross-dockings within 2.0 Å RMSD.

Docking set	Structures	Success rate %	
		<i>standard</i>	<i>hydrated</i>
Astex Diverse	85	77.6	77.6
Astex Non-Native	1112	40.4	40.6
AChBP (1)	36	0.0	30.6
AChBP (5)	36	11.1	16.7
PARP	17	11.8	88.2

Table 2

RMSD of dockings of AChBP inhibitors performed with the two methods and prediction accuracy of ligand-protein bridging waters within 2.0 Å.

	standard (Å)	hydrated (Å)	resolution (Å)	PDB id	waters (pred./exp.)
1	1.18	0.42	2.20	<i>1uw6</i>	1/1
3	2.42	0.42	2.70	<i>2wnl</i>	1/1
4	2.18	0.46	2.05	<i>2bys</i>	1/1
5	5.03	0.58	2.58	<i>2zju</i>	3/3 ^a
6	6.23	2.05	2.70	<i>2zjv</i>	3/1 ^a

^a Ambiguous/problematic water positions in experimental structure.

Table 3

RMSD of dockings of HIV-1 protease inhibitors performed with the two methods and prediction accuracy of ligand-protein bridging waters within 2.0 Å.

	standard (Å)	hydrated (Å)	resolution (Å)	PDB id	waters (pred./exp.)
7	0.78	0.65	n.a. ^a	2c3ve	5/5
8	1.09	0.41	1.75	1b6l	2/3
9	0.78	0.72	1.09	1kzk	2/2
10	0.54	0.39	2.00	1mrw	2/2
11	1.25	1.38	2.10	1ohr	2/2
12	0.94	0.89	2.00	1ajx	0/0
13	0.44	0.77	2.00	1dmp	2/2 ^b
14	1.11	0.54	2.30	5upj	0/0

^a Neutron diffraction structure.

^b Ambiguous/problematic water positions in experimental structure.

Table 4

RMSD of dockings of Scytalone dehydratase inhibitors performed with the two methods and prediction accuracy of ligand-protein bridging waters within 2.0 Å.

	standard (Å)	hydrated (Å)	resolution (Å)	PDB id	waters (pred./exp.)
15	4.18	1.84	2.15	<i>4std</i>	2/2
16	0.78	1.01	1.95	<i>5std</i>	2/2
17	1.77	1.11	1.80	<i>6std</i>	2/2
18	1.25	0.67	1.80	<i>7std</i>	2/2
19	3.13	0.49	1.65	<i>3std</i>	1/1

Table 5

RMSD of dockings of PARP inhibitors performed with the two methods and prediction accuracy of ligand-protein bridging waters within 2.0 Å.

	standard (Å)	hydrated (Å)	resolution (Å)	PDB id	waters (pred./exp.)
20	4.60	2.08	2.20	<i>Iefy</i>	4/2
21	0.47	0.48	2.40	<i>Ipxx</i>	0/0
22	0.81	0.62	2.40	<i>3pxx</i>	2/2
23	2.45	0.95	2.50	<i>3l3m</i>	4/4
24	0.38	0.51	2.40	<i>2pxx</i>	2/2
25	1.19	1.03	2.50	<i>3l3l</i>	3/3
26	3.52	0.54	3.00	<i>Iwok</i>	3/3 ^a
27	1.14	0.85	2.30	<i>2rd6</i>	2/3

^a Ambiguous/problematic water positions in experimental structure.