Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

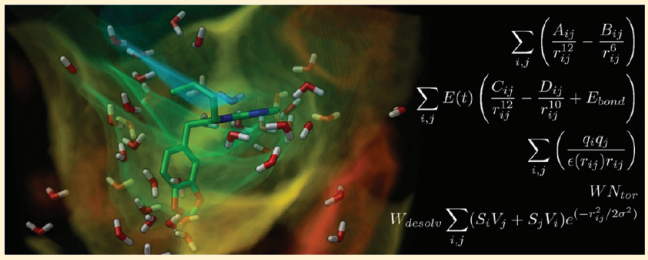
A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking

Stefano Forli and Arthur J. Olson*

Molecular Graphics Lab, Department of Molecular Biology, MB-112, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037-1000, United States

Supporting Information

ABSTRACT: In modeling ligand−protein interactions, the representation and role of water are of great importance. We introduce a force field 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 force field accounts for the entropic and enthalpic contributions of discrete waters to ligand binding, improving energy estimation accuracy and docking performance. The force field 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 force field 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 ligands[1] 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 protease (PR)[2] 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 inhibitors[5] and cyclic urea inhibitors of PR.[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 (PARP) 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 and then use an implicit solvent model in the form of a continuous 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 PR does not allow cyclic urea inhibitors to dock correctly.[12] Moreover, it is a nontrivial 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 the 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

Received: April 27, 2011
Published: December 9, 2011

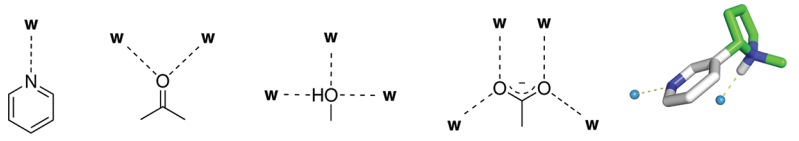
dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638

includes HINT/Rank (scoring function/geometric measurements[19]), WaterScore (multivariate regression[20]), Consolv (*k*-nearest neighborhood/GA[21]), and the method proposed by Barillari et al. (MC/replica exchange[22]). The second category includes WaterMap (short molecular dynamics analysis[23]) and JAWS (statistical thermodynamics[10]). 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* structures or among multiple *holo* structures, where waters are stabilized by the ligand interactions themselves. An example is water 52 in PARP structures.[7] The quality of the results is sensitive to the crystal structure resolution[19] 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, and DOCK)[15,24,25] or spherical particles (FlexX and 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., three waters for GOLD).[24] Also, the energetics related to water displacement is often ignored in the scoring function (FlexX, GLIDE, and DOCK).[24,25] FlexX does not require waters to be present in the target structure, but the accuracy in predicting their position and improvement in docking results has been reported to be very small.[24] Depending on the target, the inclusion of displaceable waters led to more consistent improvements with DOCK[15] 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 that we present here modifies the AutoDock 4.2 force field[29] 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* ligand, 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 monatomic pseudoatom (W) with properties designed to provide a simple description of water characteristics without compromising modeling accuracy. W atoms do not contribute to ligand intramolecular interactions. The intermolecular 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 a resolution of 2.5 Å or better, including more than 50 different protein families. The force field 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 the atom types in the ligand and the target structure are precalculated in a grid surrounding the binding region. Then, during the docking, grid interaction energies are used as a 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 1a). 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



(a) Schematic view of ligand hydration with W atoms. (b) Hydrated ligand before being docked.

Figure 1. Representation of the hydration procedure.

dx.doi.org/10.1021/jm2005145 | *J. Med. Chem.* 2012, 55, 623–638

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 sulfate 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 discrete spherical water molecule neutrally charged, presenting combined hydrogen and oxygen properties (Figure 1b). The spherical model avoids the extra degrees of freedom related to water hydrogen orientations with respect to both ligand and receptor. The ligand potential ΔG_{lig} for regular atoms is calculated in AutoDock according to eq 1:

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

The intermolecular 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] The entropy of ligand binding (ΔS_{tor}) is included to account for the loss of degrees of freedom upon binding, which 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. Because 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):

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

where the W atom, w, substitutes the regular ligand atom, i, in eq 1. The 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):

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

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 root-mean-square deviation (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 the second term (displacement). In particular, if one or more waters can stabilize a favorable ligand−receptor interaction, the water−receptor enthalpy is

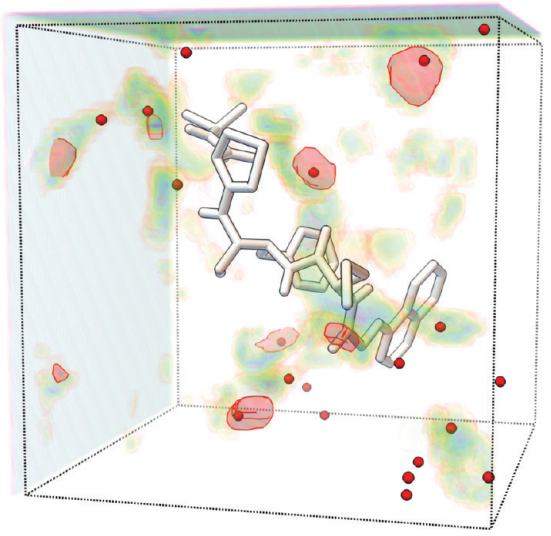
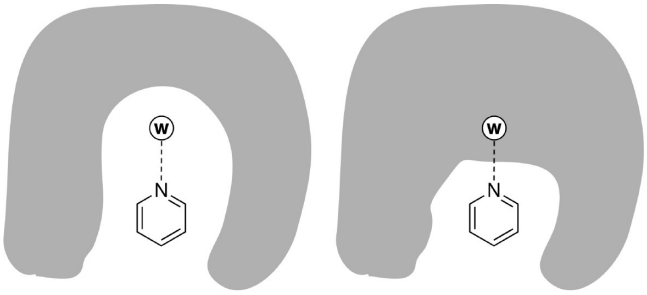


Figure 2. Volumetric rendering of the calculated *W* affinity map (HIV protease, 2zye). Ligand coordinates of **7** (white) and waters (red) are shown as ball and sticks. Predicted positions for strongly bound waters are highlighted with transparent red volumes.

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638

summed in the ligand-binding energy (Figure 3a). Conversely, when a tighter ligand−receptor interaction is more energetically favorable and



(a) $\Delta H_{vdw} + \Delta H_{hbond}$ (b) ΔS_{desolv}

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

requires a water displacement, the desolvation entropy resulting from its release to bulk solvent rewards ligand binding (Figure 3b). 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 force field eq 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 force field 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. To establish a reliable reference set, complexes matching the following structural and experimental criteria were selected:

- X-ray diffraction with resolution ≤2.5 Å
- noncovalent 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 (1add, 1bra, 1hsl, 1tng, 1tnh, 1tni, 1tnl, 2gbp, and 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 druglike molecules (MW up to 620 Da) with 0−12 rotatable bonds. The number of experimentally determined water molecules contained in grid boxes defining the binding sites ranges from 4 (1gwx) to 72 (1c5p). 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., cofactors, 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, and rotatable bonds ranged 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 protocol[40] (see the Supporting Information for details). Only protein residues and enzyme cofactors 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. OA and HD were added if not already present in the ligand atom type set. Grid maps were calculated for ligand atom types, and W maps were generated according to eq 4. Search parameters were adapted to ligand complexity (see the Supporting Information). Every ligand was docked in its dehydrated form with the standard force field and then docked in the hydrated form with the new force field, 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 were 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

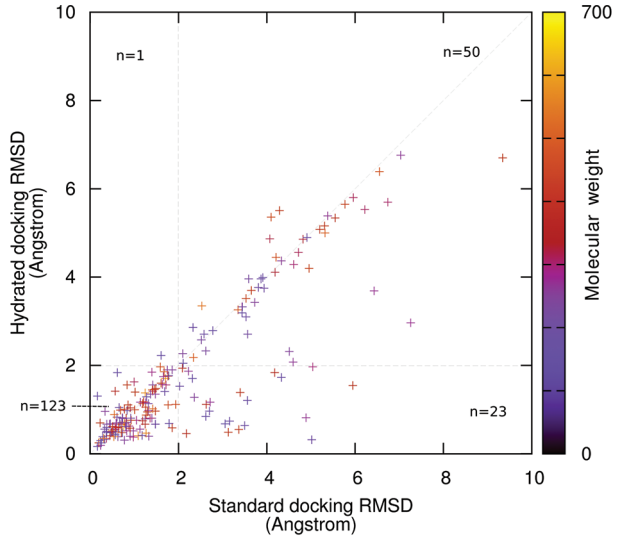
Two key requirements were defined for this method to make it suitable for potential use in drug design and virtual screening. First, the force field must be generally applicable and be independent of any previous knowledge of water presence. Thus, it must cover a wide range of ligand sizes (from small fragments to druglike 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

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638

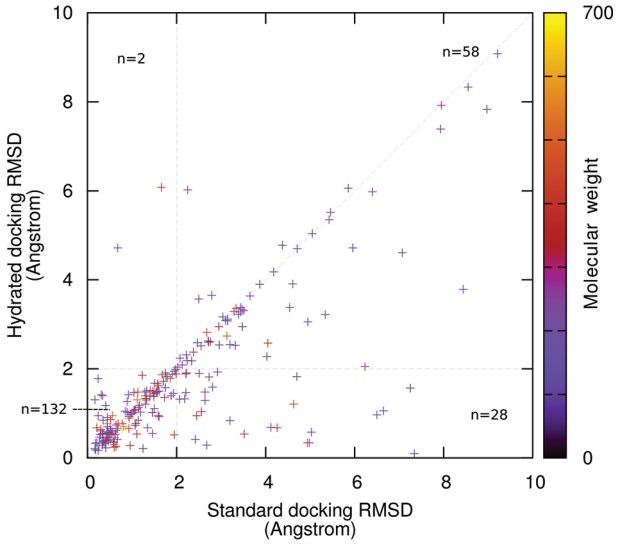
structures. Second, the method must be fast enough to be used with a large number of ligands in a high throughput fashion. Performance has been measured during the calibration on both the training set and the validation runs on the test set, and the computational overhead required by the new force field 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. Because 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, the calculation time increases depending on the number of W atoms in every ligand, ranging from 0.5 (1 water: 2pcp and 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 1/10 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 according to eq 5 in the W maps of the training set varied in the following ranges: 70−97% OA and 30−3% HD. In the training set, the standard AutoDock force field 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 4a). Improvements can be ascribed to enrichment of statistical relevance (clustering) of poses poorly represented in the standard protocol dockings (1efy and 3std) or to the sampling of poses otherwise not energetically accessible (1wcc, 1n2v, and 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 the rmsd accuracy and the number of waters modeled. The new force field appears to be able to predict the buried waters but also superficial ones stabilized by ligand binding (1efy and 1yc4). In some cases (e.g., 1u4d and 1yc4), 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 4b).

Several cases (1uy7, 1uy8, 1uy9, and 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 (1efy, 3ce0, and 1n2v) or to nondirectional interactions (i.e., van der Waals interactions: 1tng), 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 the Supporting Information. In two cases in the test set (2x8d and 2f7x), the new force field lead to rmsd >2.0 Å where the standard force field was successful. A different target protonation procedure restored accuracy in one case (2x8d). The issue is further analyzed in the Supporting Information.



(a) Training set (197 complexes)

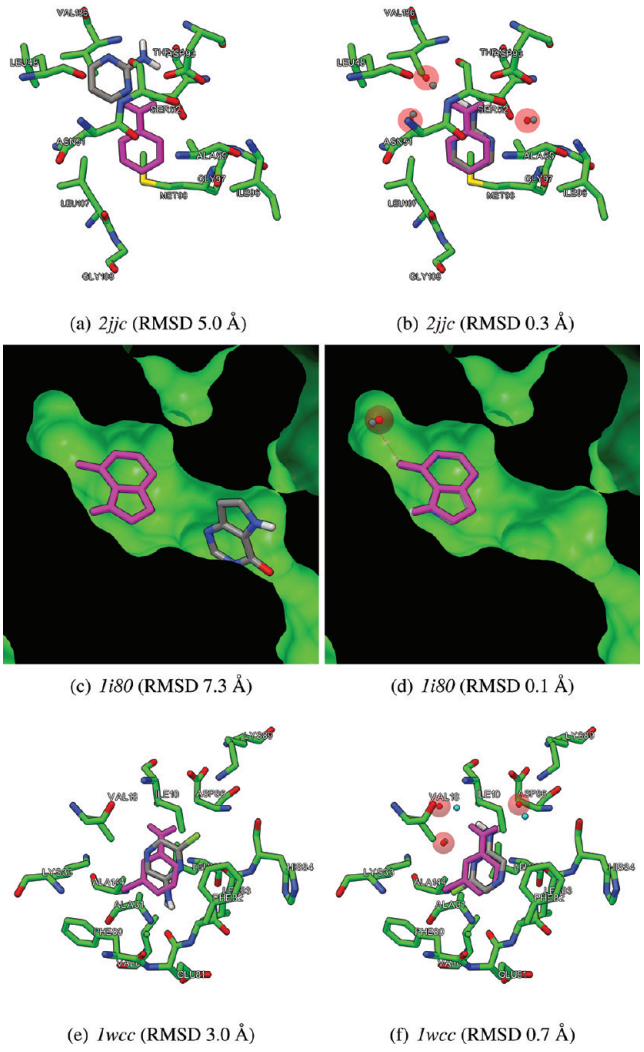


(b) Test set (220 complexes)

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.

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

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638



(a) *2jjc* (RMSD 5.0 Å) (b) *2jjc* (RMSD 0.3 Å)

(c) *1i80* (RMSD 7.3 Å) (d) *1i80* (RMSD 0.1 Å)

(e) *1wcc* (RMSD 3.0 Å) (f) *1wcc* (RMSD 0.7 Å)

Figure 5. Comparison of fragments docking with the standard (a, c, and e) and the new hydrated force field (b, d, and f). Crystallographic positions of ligands are shown as purple sticks; binding site residues are shown in green (sticks or MSMS surfaces[49]). Waters are shown as spheres red (experimental) or cyan (predicted).

number of cases where the correct binding location was found only in the hydrated docking (e.g, training set: 2jjc, 1f3e, 2wi2, 1enu, 1wcc, and 1wbu; test set: 1i80, 3bsf, and 1oty). In Figure 5 are shown some representative docking results of fragment-sized ligands.

Metal-Coordinating Complexes. The training set did not contain proteins with catalytic metals in the binding sites, while the test set contains eight matrix metalloproteinases (MMPs) (1biw, 1d8m, 1g4k, 1zp5, 2yig, 3ehx, 3ehy, and 3kry), seven phosphodiesterases (PDEs) (1xlx, 1xm6, 1xmu, 3g45, 3g4k, 3hmv, and 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 the 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 force field 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 (1b6l:HOH320, 1c83:HOH370) or which interact through van der Waals or electrostatic forces (1pax). Rigidity of the W atom position is also responsible for the drop in accuracy with hydrated phosphate and sulfate oxygens (e.g., training, 1rbo; test, 1xgj); 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.

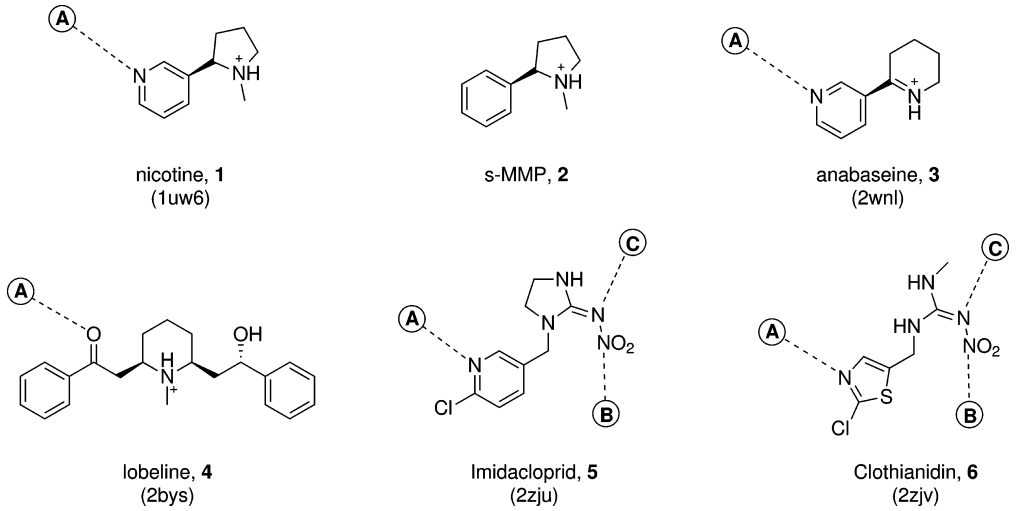
External Data Sets and Case Study Cross-Docking Validations. Table 1 shows a summary of external set and cross-docking validations. Standard and hydrated force fields performed equally well when docking complexes in the Astex Diverse Set[27] (85 ligand−protein complexes) reproducing ligand binding poses 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 set[50] (67 ligands and 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 acetylcholine binding protein (AChBP) protein structures (34 *holo* and 2 *apo*). For PARP, compound **27** was docked against 17 protein structures (15 *holo* and 2 *apo*). For both targets, ligands with known common binding mode for all subtypes (AChBP, *Lymnaea stagnalis* and *Aplysia californica* structures; PARP, subtypes 1 and 2) were selected for docking. Ligand coordinates were generated from SMILES strings deposited in the PDB entries, and protonation states were visually inspected and corrected when necessary. Ligands were docked following the protocol described in the Methods section.

Table 1. Success Rates of External Set Dockings and Cross-Dockings within 2.0 Å rmsd

		success rate (%)	
docking set	structures	standard	hydrated
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

dx.doi.org/10.1021/jm2005145 | *J. Med. Chem.* 2012, 55, 623−638

Scheme 1. AChBP Inhibitor Structures with Experimentally Determined Ligand-Bound Waters



Receptor structures were kept rigid during docking. Further details are available in the 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-loop[54] 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%). Compound **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 (waters 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 subtypes.[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 cricual 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.

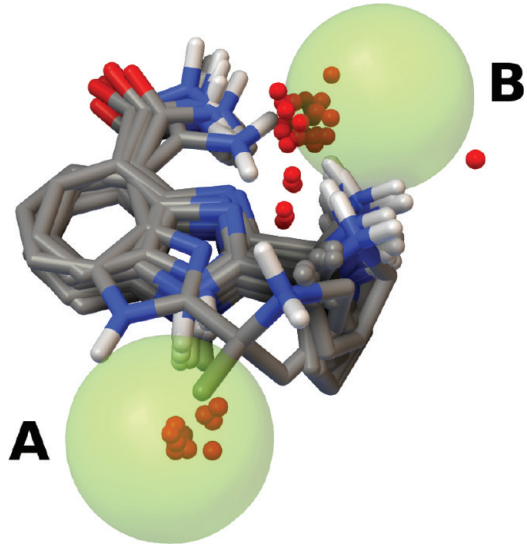


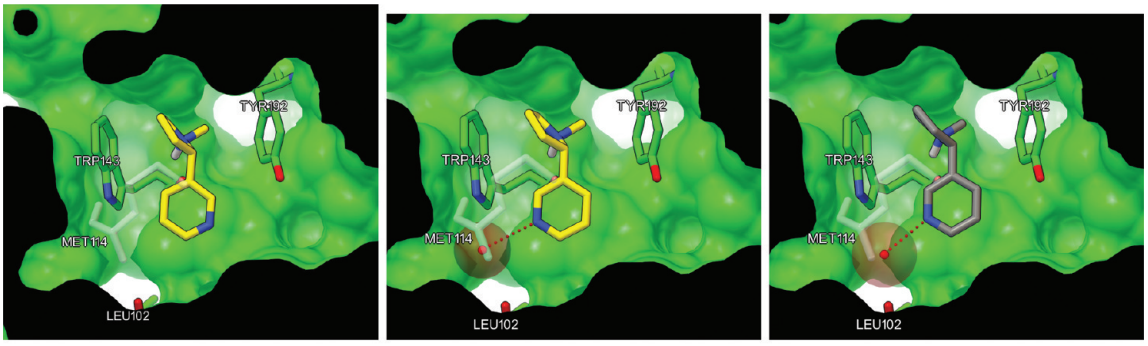
Figure 6. Representation of 15 successful cross-docking results of **27**. Green spheres show a 2.0 Å range around experimental coordinates of waters A and B. The rightmost red sphere is water E (Scheme 4), predicted only in one docking.

■ 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.

AChBP (1uw6, 2zju, 2zjv, 2byq, 2bys, 2wn9, 3c79, 3c84, and 2wnl). Rationalization of the activity of agonists of the neuronal 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 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

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638



(a) standard (RMSD 1.18 Å) (b) hydrated (RMSD 0.42 Å) (c) experimental

Figure 7. Comparison of experimental structure of **1** and dockings obtained with standard and hydrated force fields (1uw6). Docked poses are colored in yellow; the experimental pose is in gray.

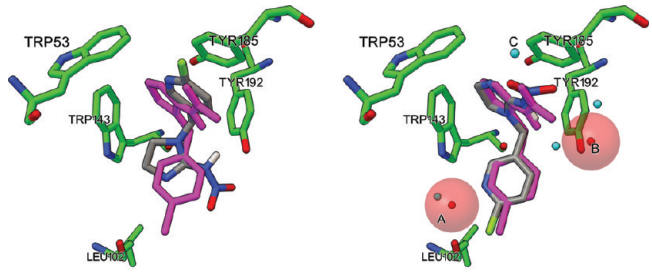
L. stagnalis (training set, 1uw6; test set, 2zju and 2zjv) and six from *A. californica* (training set, 2bys and 2wn9; test set, 3c79, 3c84, 2wnl, and 2byq). Structure 1uw6 contains **1** bound,[54] where the pyrrolidine charged group is interacting with Trp147 (cation–π and hydrogen bond interactions), and water HOH2063 (water A) mediates hydrogen bonds between the Met114 and Leu102 backbone and the pyridine ring (Figure 7c). Both the standard and the new hydration dockings can reproduce binding mode of **1** with satisfactory accuracy (1.18 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 the 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 force field were tested by docking **1** derivative s-MMP (*N*-methyl-2-phenylpirroline,[3] **2**) to the same structure (1uw6). Ligand **2** has 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 force fields are able to reproduce the binding mode, although in the absence of the water the standard force field does not correctly rank the two ligands (**2** > **1**). The new force field 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.

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	1uw6	1/1
3	2.42	0.42	2.70	2wnl	1/1
4	2.18	0.46	2.05	2bys	1/1
5	5.03	0.58	2.58	2zju	3/3[a]
6	6.23	2.05	2.70	2zjv	3/1[a]

[a]Ambiguous/problematic water positions in the experimental structure.

Improvements 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



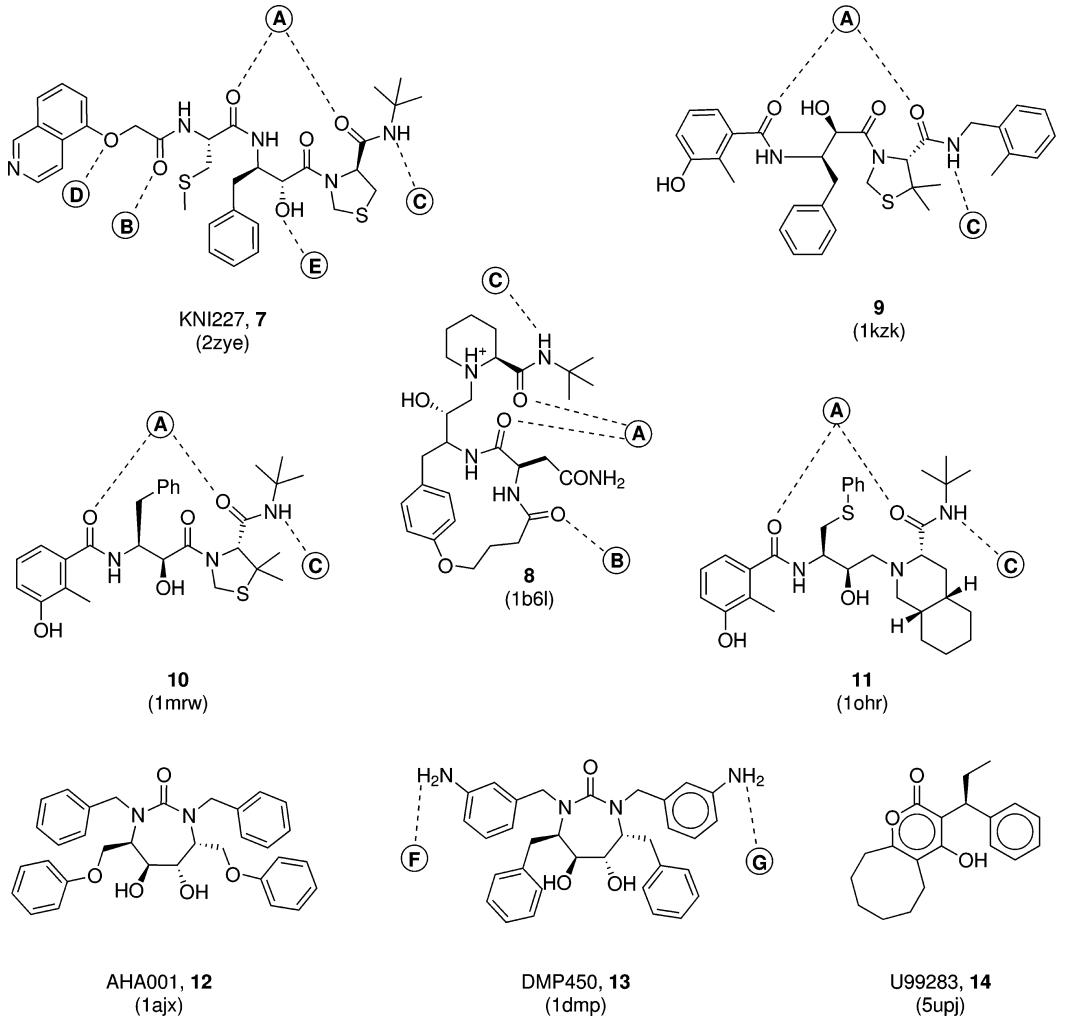
(a) standard (RMSD 5.03 Å) (b) hydrated (RMSD 0.58 Å)

Figure 8. Comparison of experimental structure of **5** and dockings obtained with standard and hydrated force fields (2zju). The experimental structure is colored in purple; the docked poses are colored by atom type.

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. Because of the low crystallographic resolution of 2jzu, these two waters are not defined in all of the pentamer 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. The 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 aligning[60] 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

dx.doi.org/10.1021/jm2005145 | *J. Med. Chem.* 2012, 55, 623–638

Scheme 2. HIV-1 PR Inhibitor Structures with Experimentally Determined Ligand-Bound Waters



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).

PR (1b6l, 1kzk, 1mrw, 1ohr, 2zye, 1ajx, 1dmp, 5upj, and 1hxw). PR is an essential enzyme for the reproductive cycle of HIV, the virus responsible for AIDS.[61] Structural studies have shown that the activity of potent PR inhibitors, like KNI-272[44] (7, Scheme 2) and similar binders (**8**–**11**), is mediated by structural water HOH301[2] (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

Table 3. rmsd of Dockings of PR 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]	2zye	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 the experimental structure.

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623–638

contained four complexes of PR with inhibitors, all binding via water A: **8** (1b6l), **9** (1kzk), **10** (1mrw), and **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 in Figure 9. To test the

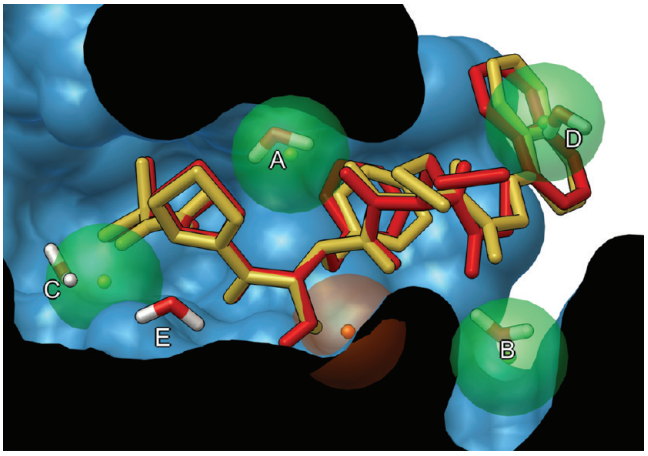


Figure 9. Comparison of hydrated docking pose (rmsd, 0.65 Å) and waters prediction with the neutron diffraction structure of **7** complexed in 2zye. Experimental coordinates are in gold sticks, and docked ones are in red. Experimental waters are shown as stick water molecules, and predicted water positions are shown as green spheres. The red sphere water shows the predicted 2-fold symmetry position of water E (see the text).

accuracy of the prediction of other waters, **7** was redocked 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 values of 0.9, 0.9, 1.2, 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 2-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 AHA001[62] (**12**, 1ajx), DMP450[63] (**13**, 1dmp), and a 4-hydroxycumarin derivative U99283[64] (**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. Because of the high structural similarity between ligands and protein conformation,[65] positions of waters present in 1ajx were used to measure rmsd values of predicted waters in 1dmp 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 (SD) (3std, 4std, 5std, 6std, and 7std). 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 Michel2009 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 force field docking fails to reproduce the binding pose in two cases (**15** and **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 force field 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.

PARP (1pax, 2pax, 3pax, 1efy, 1wok, 1uk1, 2rcw, 2rd6, 3c49, 3c4h, 3ce0, 3fhb, 3gjw, 3gn7, 3goy, 3hkv, 3kcz, 3kjd, 3l3l, 3l3m, 3p0n, 3p0p, and 4pax). PARP is a family of enzymes involved in the process of post-translational

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638

Scheme 3. SD Inhibitor Structures with Experimentally Determined Ligand-Bound Waters

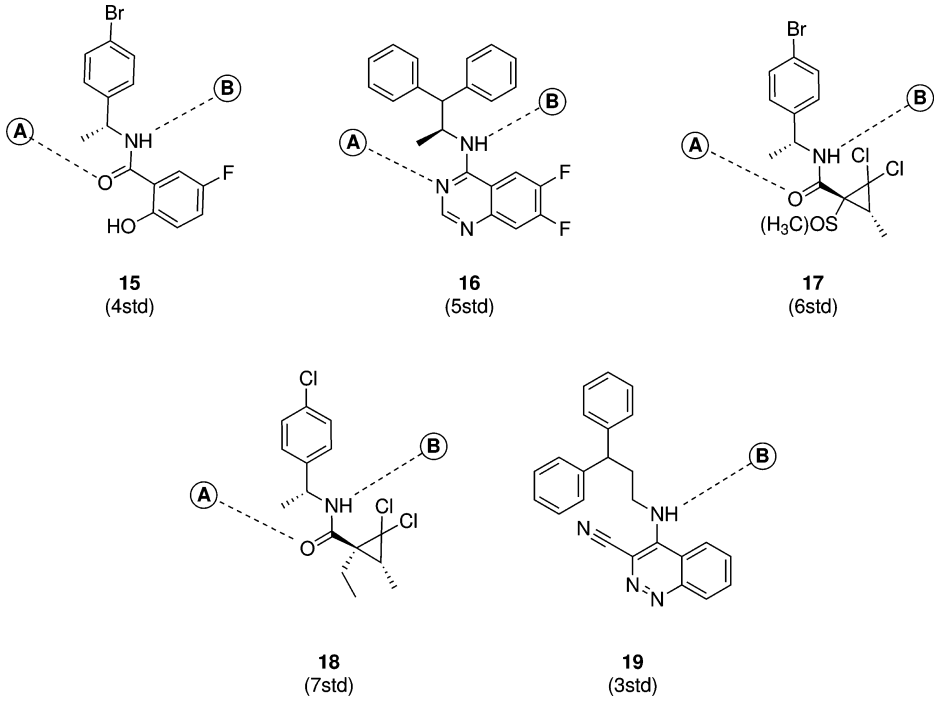


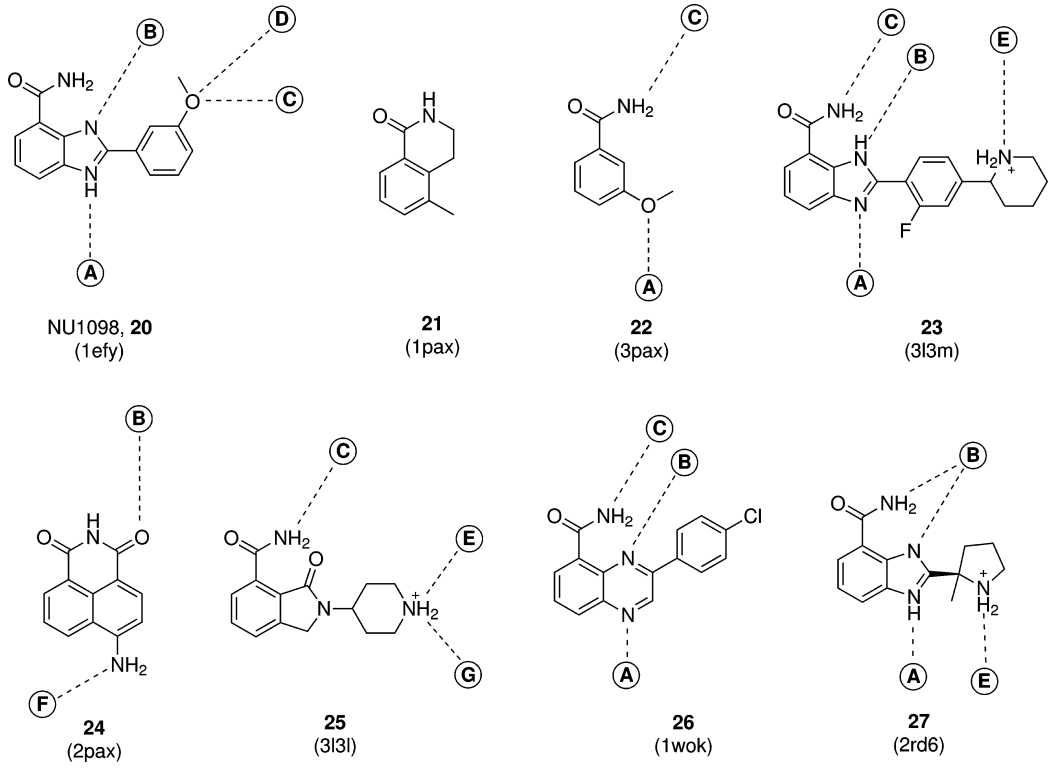
Table 4. rmsd of Dockings of SD Inhibitors Performed with the Two Methods and Prediction Accuracy of Ligand–Protein Bridging Waters within 2.0 Å

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

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 subtypes is critical to their therapeutic application.[68] The activity of known inhibitors involves the formation of a hydrogen bond with Gly863. The training set contains complexes of four 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 NU1098[69] (**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 eight 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 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**, and **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, and E) matching experimental coordinates are predicted to mediate the interactions of different ligand substituents with the protein (Figure 11). On the other hand, the naphthyl-amino group of **24** is correctly predicted to displace water A, while water B is conserved. Predictions of the position of waters were 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

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623–638

Scheme 4. PARP Inhibitor Structures with Experimentally Determined Ligand-Bound Waters

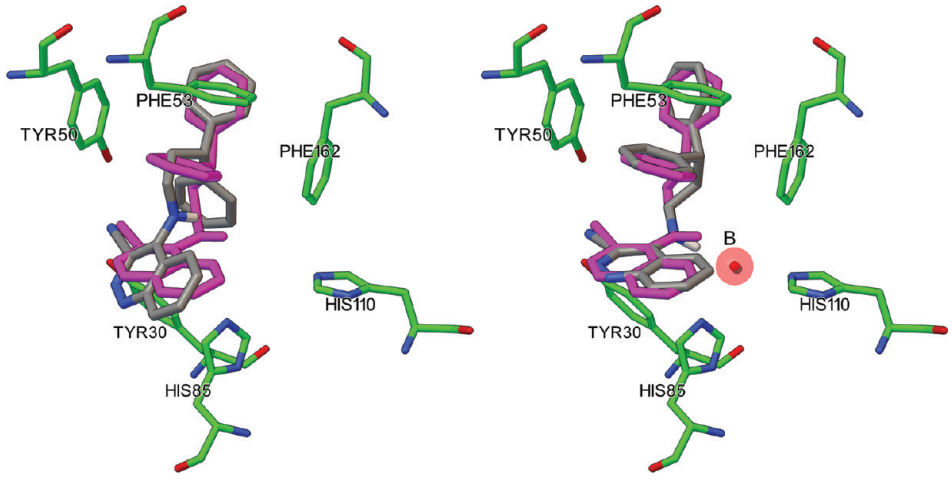


improving estimates of ligand binding energetics. Dockings of PARP inhibitors show the capabilities of the new hydration method to predict the position of weakly bound waters and rank them in accordance with experimental data, improving the docking accuracy. In particular, transient waters stabilized 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 (1pax), water A seems to be stabilized by weak van der Waals interactions with a nearby ligand methyl group, but because of the nature of the interaction itself the hydration force field is not able to predict its presence.

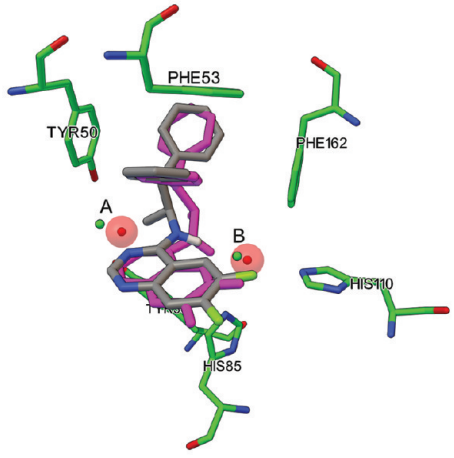
■ CONCLUSIONS

We have developed a force field 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 force field 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 force field separately calculates entropy and enthalpy for every water, enabling a fine-grained estimation of their contributions in the 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 force field are able to find experimentally determined waters and predict their stability using grid-based calculations. The new force field combines the two approaches of predicting water positions and uses them during docking. As demonstrated in case studies and cross-dockings, the fact that the force field 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%) as compared with the standard docking. Validation of the force field 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

dx.doi.org/10.1021/jm2005145 | J. Med. Chem. 2012, 55, 623−638

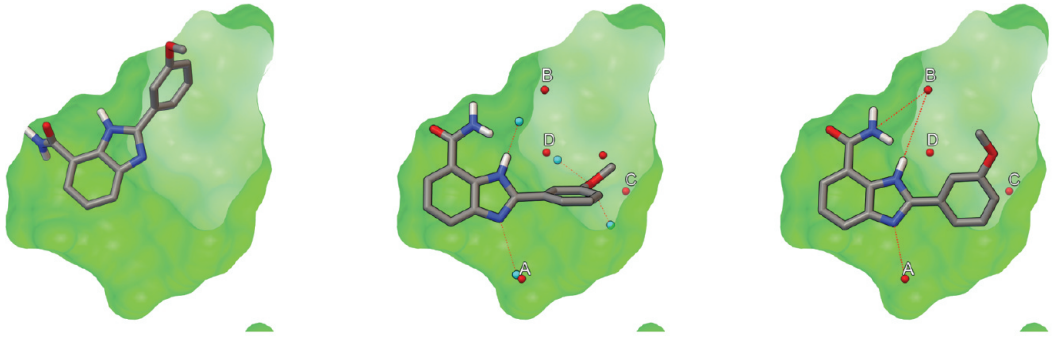


(a) *3std* standard (RMSD 4.18 Å) (b) *3std* hydrated (RMSD 1.84 Å)



(c) *5std* hydrated (RMSD 1.01 Å)

Figure 10. Comparison of dockings of **19** with the standard and hydrated methods (a and b) and water predictions obtained by docking **15** (c). The crystallographic poses are represented as purple sticks; protein residues are represented as green sticks; docked results are represented as gray sticks; and crystallographic (red) and predicted waters (cyan) are shown as spheres.



(a) standard (RMSD 4.6 Å) (b) hydrated (RMSD 2.08 Å) (c) crystallographic pose

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

the binding site and the interactions that 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, because of its hydrogen bond bivalency, a water molecule can also "invert" a receptor's hydrogen bond acceptor region into a donor. The force field is able to model this behavior, thereby sampling a broader bioisosteric 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.

dx.doi.org/10.1021/jm2005145 | *J. Med. Chem.* 2012, 55, 623–638

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	1efy	4/2
21	0.47	0.48	2.40	1pax	0/0
22	0.81	0.62	2.40	3pax	2/2
23	2.45	0.95	2.50	3l3m	4/4
24	0.38	0.51	2.40	2pax	2/2
25	1.19	1.03	2.50	3l3l	3/3
26	3.52	0.54	3.00	1wok	3/3[a]
27	1.14	0.85	2.30	2rd6	2/3

[a]Ambiguous/problematic water positions in the experimental 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 force field, which will also be addressed in future work. In particular, flexible W atoms or a soft potential might compensate for nonideal water placement and phosphate/sulfate 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 force field to explicitly hydrated target structures, increasing the desolvation potential accuracy and providing a model for waters displaced by nonpolar 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.

Because of 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]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

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

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +1-858-784-9702. Fax: +1-858-784-2860. E-mail: olson@scripps.edu.

■ ACKNOWLEDGMENTS

S.F. thanks Dr. Luca Bellucci for the inspiring discussion and 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 was published.[73] A brief comparison of the two methods is provided in the Supporting Information.

■ ABBREVIATIONS USED

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, *N*-methyl-2-phenylpirroline; PR, HIV-1 protease; SD, scytalone dehydratase; MMP, matrix metalloproteinase; PDE, phosphodiesterase; OA, oxygen acceptor probe; HD, hydrogen donor probe

■ REFERENCES

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

(2) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* **1995**, *3*, 581–590.

(3) Blum, A. P.; Lester, H. A.; Dougherty, D. A. 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.

(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.

(5) Chen, J. M.; Xu, S. L.; Wawrzak, Z.; Basarab, G. S.; Jordan, D. B. Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry* **1998**, *37*, 17735–17744.

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

(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.

(8) 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(ADPribose) polymerase ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624–633.

(9) Wissner, A.; Berger, D. M.; Boschelli, D. H.; Floyd, M. B.; Greenberger, L. M.; Gruber, B. C.; Johnson, B. D.; Mamuya, N.; Nilakantan, R.; Reich, M. F.; Shen, R.; Tsou, H. R.; Upeslacis, E.; Wang, Y. F.; 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.

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

(11) Mysinger, M. M.; Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **2010**, *50*, 1561–1573.

(12) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

(13) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.
(14) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, *49*, 1023−1033.
(15) Huang, N.; Shoichet, B. K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51*, 4862−4865.
(16) Roberts, B. C.; Mancera, R. L. Ligand-protein docking with water molecules. *J. Chem. Inf. Model.* **2008**, *48*, 397−408.
(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.
(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. *J. Med. Chem.* **1993**, *36*, 148−156.
(19) Amadasi, A.; Surface, J. A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust classification of "relevant" water molecules in putative protein binding sites. *J. Med. Chem.* **2008**, *51*, 1063−1067.
(20) García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: A novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *J. Mol. Model.* **2003**, *9*, 172−182.
(21) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445−464.
(22) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577−2587.
(23) Young, T.; Abel, T.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein−ligand binding. *PNAS* **2007**, *104*, 808−813.
(24) de Beer, S. B. A.; Vermeulen, N. P. E; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Top. Med. Chem.* **2010**, *10*, 55−66.
(25) Villacanas, O.; Madurga, S.; Giralt, E.; Belda, I. Explicit Treatment of Water Molecules in Protein-Ligand Docking. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 145−154.
(26) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(27) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, *50*, 726−741.
(28) Minke, W. E.; Diller, D. J.; Hol, W. G.; Verlinde, C. L. 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.
(29) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145−1152.
(30) Goodsell, D.; Olson, A. Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct., Funct., Bioinf.* **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. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(32) Wiberg, K.; Marquez, M.; Castejon, H. Lone pairs in carbonyl-compounds and ethers. *J. Org. Chem.* **1994**, *59*, 6817−6822.
(33) Taylor, R.; Kennard, O.; Versichel, W. Geometry of the n-h=/=c hydrogen-bond 0.1. lone-pair directionality. *J. Am. Chem. Soc.* **1983**, *105*, 5761−5766.
(34) 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.
(35) 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.
(36) Morris, G.; Goodsell, D.; Huey, R.; Olson, A. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293−304.
(37) Mehler, E.; Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Eng.* **1991**, *4*, 903.
(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(39) The Protein Data Bank; http://www.pdb.org/ (accessed: 08/09/2011).
(40) Morris, G. M.; Goodsell, D. S.; Pique, M. E.; Lindstrom, W.; Huey, R.; Forli, S.; Hart, W. E.; Hallyday, S.; Belew, R.; Olson, A. J. *AutoDock 4.2 User Guide*; http://autodock.scripps.edu/faqs-help/manual/autodock-4-2-user-guide/AutoDock4.2_UserGuide.pdf (accessed August 9, 2011).
(41) Illingworth, C. J. R.; Morris, G. M.; Parkes, K. E. B.; Snell, C. R.; Reynolds, C. A. Assessing the role of polarization in docking. *J. Phys. Chem. A* **2008**, *112*, 12157−12163.
(42) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discovery* **2010**, *5*, 597−607.
(43) Chang, M.; Belew, R.; Carroll, K.; Olson, A.; Goodsell, D. Empirical entropic contributions in computational docking: Evaluation in APS reductase complexes. *J. Comput. Chem.* **2008**, *29*, 1753−1761.
(44) Kiso, Y.; Matsumoto, H.; Mizumoto, S.; Kimura, T.; Fujiwara, Y.; Akaji, K. Small dipeptidebased HIV protease inhibitors containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic. *Pept. Sci.* **1999**, *51*, 59−68.
(45) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van Der Oost, J.; Smit, A. B.; Sixma, T. K. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **2001**, *411*, 269−276.
(46) Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, 670.
(47) Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **2007**, *6*, 211−219.
(48) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358−364.
(49) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305−320.
(50) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-ligand docking against non-native protein conformers. *J. Chem. Inf. Model.* **2008**, *48*, 2214−2225.
(51) Tuccinardi, T.; Botta, M.; Giordano, A.; Martinelli, A. Protein kinases: Docking and homology modeling reliability. *J. Chem. Inf. Model.* **2010**, *50*, 1432−1441.
(52) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632−5633.
(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.
(54) Celie, P. H. N.; van Rossum-Fikkert, S. E.; van Dijk, W. J.; Brejc, K.; Smit, A. B.; Sixma, T. K. Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **2004**, *41*, 907−914.
(55) Talley, T. T.; Harel, M.; Hibbs, R. E.; Radic, Z.; Tomizawa, M.; Casida, J. E.; 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.

(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.
(57) Beers, W. H.; Reich, E. Structure and activity of acetylcholine. *Nature* **1970**, *228*, 917−922.
(58) Glennon, R. A.; Dukat, M. Central nicotinic receptor ligands and pharmacophores. *Pharm. Acta Helv.* **2000**, *74*, 103−114.
(59) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature* **2009**, *458*, 534−537.
(60) Martínez, L.; Andreani, R.; Martínez, J. M. Convergent algorithms for protein structural alignment. *BMC Bioinf.* **2007**, *8*, 306.
(61) Brik, A.; Wong, C.-H. HIV-1 protease: Mechanism and drug discovery. *Org. Biomol. Chem.* **2003**, *1*, 5−14.
(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. *J. Med. Chem.* **1997**, *40*, 898−902.
(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. *Chem. Biol.* **1996**, *3*, 301−314.
(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. *J. Med. Chem.* **1995**, *38*, 1884−1891.
(65) A 0.55 Å rmsd on the protein backbone.
(66) Valent, B.; Farrall, L.; Chumley, F. G. Magnaporthe grisea genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics* **1991**, *127*, 87−101.
(67) Delaney, C. A.; Wang, L. Z.; Kyle, S.; White, A. W.; Calvert, A. H.; Curtin, N. J.; Durkacz, B. W.; Hostomsky, Z.; Newell, D. R. 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.
(68) Yap, T. A.; Sandhu, S. K.; Carden, C. P.; de Bono, J. S. Poly(ADP-ribose) polymerase (PARP) inhibitors: Exploiting a synthetic lethal strategy in the clinic. *CA Cancer J. Clin.* **2011**, *61*, 31−49.
(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. *J. Med. Chem.* **2000**, *43*, 4084−4097.
(70) A 0.9 Å rmsd on the protein backbone.
(71) Minor changes to the source code can further improve efficiency by excluding *W* atoms from the ligand intramolecular interaction calculation, removing almost entirely the overhead of simulating waters.
(72) Sanner, M. F. Python: A programming language for software integration and development. *J. Mol. Graphics Modell.* **1999**, *17*, 57−61.
(73) Lie, M. A.; Thomsen, R.; Pedersen, C. N. S.; Schiøtt, B.; Christensen, M. H. Molecular docking with ligand attached water molecules. *J. Chem. Inf. Model.* **2011**, *51*, 909−917.