JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

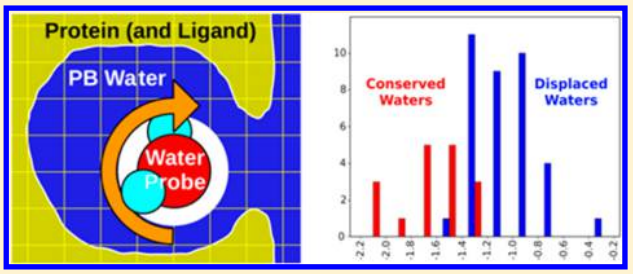
Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP

Alexander S. Bayden,[‡,#] Demetri T. Moustakas,[‡,§] Diane Joseph-McCarthy,[⊥] and Michelle L. Lamb*

Oncology and Infection Innovative Medicines Units, AstraZeneca R&D Boston, 35 Gatehouse Drive, Waltham, Massachusetts 02451, United States

Supporting Information

ABSTRACT: The SZMAP method computes binding free energies and the corresponding thermodynamic components for water molecules in the binding site of a protein structure [*SZMAP, 1.0.0*; OpenEye Scientific Software Inc.: Santa Fe, NM, USA, 2011]. In this work, the ability of SZMAP to predict water structure and thermodynamic stability is examined for the X-ray crystal structures of a series of protein−ligand complexes. SZMAP results correlate with higher-level replica exchange thermodynamic integration double decoupling calculations of the absolute free energy of bound waters in the test set complexes. In addition, SZMAP calculations show good agreement with experimental data in terms of water conservation (across multiple crystal structures) and B-factors over a subset of the test set. In particular, the SZMAP neutral entropy difference term calculated at crystallographic water positions within each of the complex structures correlates well with whether that crystallographic water is conserved or displaceable. Furthermore, the calculated entropy of the water probe relative to the continuum shows a significant degree of correlation with the B-factors associated with the oxygen atoms of the water molecules. Taken together, these results indicate that SZMAP is capable of quantitatively predicting water positions and their energetics and is potentially a useful tool for determining which waters to attempt to displace, maintain, or build in through water-mediated interactions when evolving a lead series during a drug discovery program.



INTRODUCTION

A variety of computational solvent mapping approaches−including SZMAP,[1] the focus of this work−are currently being developed to better assess and account for solvent−solute interactions which have been shown to contribute significantly to the free energy of protein−ligand and protein−protein binding.[2] In the process of binding to its protein target, the ligand must be partially or fully desolvated to enter the protein binding pocket and it must also displace any water that occupies the binding pocket. The net free energy of dehydration of the solute molecules (protein and ligand) can contribute favorably or unfavorably to protein−ligand binding, and thereby, solvation effects may complicate the elucidation of structure−activity relationships (SAR) of a lead series in drug discovery. In this capacity, bridging waters that mediate the interaction between the protein and a ligand can play a critical role in binding and in defining the SAR of the series. For example, a water molecule increases the binding affinity of methotrexate for human dihydrofolate reductase by forming hydrogen bonds with both the protein and the ligand and thus acting as a bridge promoting their interactions.[3] The ability to predict the importance of individual water molecules in protein binding sites therefore could be advantageous when evaluating new ligand ideas and SAR.

Several methods have been reported to assess the energy of bound waters. Of the empirical force field approaches, one of the first programs for examining the energetics of a water probe inside a protein binding site to predict water positions was GRID.[4] Similarly, the MCSS approach[5,6] is able to predict bound water positions with some degree of accuracy based on the enthalpic contributions estimated using a standard molecular mechanics force field.[7,8] In a recent article, Wallnoefer et al. showed that, while neglecting important water positions in Factor Xa at the start of a simulation results in serious structural distortions during molecular dynamics simulations, a GRID-derived water network can stabilize the simulations, as does a water set derived from an ensemble of all 73 X-ray structures of the protein.[9] Prediction of waters within a protein−protein interface was a recent topic for a CAPRI (critical assessment of predicted interactions) experiment, in which 20 research groups participated.[10]

Consolv[11] and WaterScore[12] are different approaches that take into account, among other empirical factors, crystallographic B-factors in determining which waters in an apo protein structure are likely to be displaceable. AcquaAlta[13] was developed to reproduce water molecules bridging polar interactions between ligands and proteins; it utilizes geometric criteria obtained from extensive searches of the Cambridge Structural Database[14] to place water molecules into protein

Received: December 16, 2014
Published: July 15, 2015

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

binding sites and uses *ab initio* calculations to estimate propensities of ligand hydration. HINT[15,16] has also been successfully applied to studying water in protein pockets.[17−19] The HINT force field estimates the Gibbs free energies of noncovalent interactions based on van der Waals (vdW) interactions and partial atomic partition coefficients assigned using rules obtained by analyzing a large set of partition coefficients of small organic molecules: hydrophobic−hydrophobic and Lewis base−Lewis acid interactions are favorable while hydrophobic−hydrophilic, Lewis acid−Lewis acid and Lewis base−Lewis base interactions are unfavorable. The WaterDock protocol[20] uses AutoDock Vina[21] to dock water molecules into protein binding sites. In addition, data-mining, heuristic, and machine learning techniques are combined to develop probabilistic water molecule classifiers capable of predicting which of the docked waters are likely to be conserved or displaced and, to an accuracy of 80% on the test set studied, whether they would be displaced by polar or nonpolar ligand groups.

For simulations and large-scale ligand docking, water is often modeled as a high-dielectric continuum field through either a Poisson−Boltzmann (PB)[22,23] or Generalized Born (GB) model.[24−27] Both of these implicit solvent models feature a surface area (SA) term which imparts an energetic penalty for the creation of cavities in the solvent, and the solvent field is comprised of infinitely small, polarizable particles, which interacts with electrostatic charges of solvated molecules, dampening the magnitude of protein−ligand electrostatic energies. In general, the advantage of implicit models over the explicit solvent models is lower computational cost, due to less extensive sampling of configurations of solvent molecules. In the PB/SA model, this interaction is approximated by numerically solving the PB equation; the GB model is a simplification to enable similar accuracy at a further reduced computational cost.

The 3D-RISM[28] method derives properties and thermodynamic parameters associated with water from the three-dimensional distribution of average densities obtained using the density functional theory of nonuniform polyatomic liquids.[29] As such, 3D-RISM achieves full atomistic sampling of solvent, including ions, through an integral approach. In one recent application to an HIV1 protease crystal structure, the RMSD for predicted water positions versus their crystallographic location was found to correlate with both the experimental B-factor and calculated entropy.[30] Other advantages of 3D-RISM are that it does not assume solvent molecules to be infinitely small, and as implemented in the MOE software,[31] it can utilize fairly standard force field parameters (e.g., AMBER[32] for the protein, GAFF[33] parameters for ligands, and modified SPC water parameters)[34] and therefore is expected to be applicable to many different systems.

Another class of methods derives thermodynamic properties of water by analyzing trajectories from all-atom, explicit solvent simulations. A number of recent methods,[35−37] of which WaterMap[38−41] may be the best known, are based on inhomogeneous solution theory and employ a cluster analysis of molecular dynamics trajectories to determine water binding sites. Other groups have taken a Monte Carlo simulation approach for sampling water configurations.[42,43] Barillari et al. used the replica exchange thermodynamic integration (RETI) double decoupling method to determine the absolute free energy of water binding to protein−ligand systems.[42] In this method, two simulations are performed. In one, the water molecule is decoupled from the bulk solvent and, in the other, from the protein ligand complex. A spherical hard wall potential is introduced later in the simulation to prevent water molecules of interest from wandering off and other solvent molecules from occupying their space. Decoupling is done by gradually switching off electrostatic interactions and then Lennard-Jones interactions for the water molecule being studied in the isothermal−isobaric ensemble. While this approach is thermodynamically rigorous, it is time-consuming.

SZMAP (solvent-Zap-mapping)[1] is a semiexplicit solvent mapping approach for studying the thermodynamics of water in protein−ligand complexes that combines the advantages of both explicit and implicit solvent modeling methods. SZMAP places an explicit water probe into a protein−ligand system. The water probe is a three-atom probe with partial charges, similar to the water model used in many all-atom force fields. The remaining solvent is represented as a Poisson−Boltzmann implicit solvent (Figure 1) modeled using the ZAP toolkit.[44]

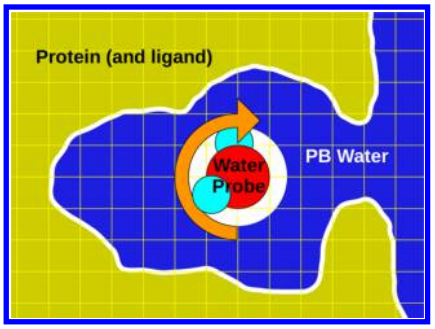


Figure 1. SZMAP places an explicit water probe into a protein−ligand system flooded by implicit water. It can sample water orientations (as represented by the orange arrow) either on a grid or at a user-specified set of coordinates.

SZMAP can sample water energies either at the points comprising a cubic grid placed over the protein, thereby building three-dimensional maps of the thermodynamic binding parameters of water within a binding site, or at user-specified locations such as the coordinates of a crystallographic water molecule or ligand atom. One obvious advantage of SZMAP over explicit solvent methods is its speed, since it does not require lengthy equilibration and extensive sampling to obtain reliable results. Few applications of SZMAP have been published to date; however, it has shown some promise in the assessment of G-protein coupled receptor druggability[45] and ligand binding free energy predictions[46] and, very recently, to the analysis of selectivity between the leishmanial homologues of GSK3β and cdc2-related protein kinase 3.[47]

In this work, we evaluate SZMAP by applying it to the training set described by Barillari et al.[42] to determine if the predicted free energies of binding for each water molecule correlate with the results of the higher-level Monte Carlo-based calculations reported in that study. We also investigate whether thermodynamic parameters computed by SZMAP correlate with experimental data for crystallographic waters: their positions, conservation across a series of protein−ligand complexes, and B-factors. The sensitivity of these calculations to water position and orientational sampling is explored. Finally, we evaluate the utility of SZMAP, and the descriptors and energy terms it calculates, for determining which waters in protein−ligand complexes to attempt to displace through ligand modification(s).

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

METHODS

System Preparation Protocol. We obtained from the Protein Data Bank the protein structures used by Barillari et al.[42] Their training set included five structures of HIV-1 protease, seven structures of neuraminidase (NA), seven structures of trypsin, five structures of Factor Xa (FXa), five structures of scytalone dehydratase (SD), and six structures of salmonella oligopeptide-binding protein (OppA), while their test set consisted of three structures of cyclin-dependent kinase 2 (CDK2). For each protein, structures were aligned to the highest-resolution example in the set with Maestro 9.2[48] using the positions of the protein α-carbons within 5 Å of the ligand. Missing side chains were built using Prime 3.0.[49,50] We utilized their training set as the "test set" for this study and only included the CDK2 structures when explicitly stated.

Structures were further prepared in Molecular Operating Environment (MOE), version 2010.1004.[31] Ligand atom and bond types were manually inspected. The Protonate3D protocol[51] was run with the default settings and GB/VI electrostatics to find the optimal protonation state and hydrogen geometry for each structure. All protonatable groups were titrated; HIS, ASN, and GLN residues were flipped as needed, and temperature (300 K), pH (7.0), and salt concentration (0.1M) values were defined according to the defaults.

For consistency with the Barillari et al.[42] publication, some structures required further manual preparation. Specifically, for the scytalone dehydratase (SD) structures, the protonation state described by Zheng and Bruice[52] was used, in which His 85 and His 110 were protonated on the Nδ and the Nε atom was unprotonated. For HIV-1 protease, in complexes 1EC1, 1EC0, 1EBW, and 1EBY, both catalytic aspartate residues were protonated, while in complex 1HPX, Asp 25 was protonated and Asp 25′ was not, based on the work of Trylska et al.[53] and Yamazaki et al.[54] It should be noted that there exists some controversy regarding the protonation states of residues ASP 25 and ASP 25′ in HIV-1 protease.[55] The protonation state we employed is consistent with the ligand screening the "water A" binding site and, thus, is more relevant for ligand-binding than for water-binding. Furthermore, we have found (data not shown) that although the SZMAP n_dTdS and SZMAP n_ddG values (as defined below) do change by 0.052 and 1.14 kcal/mol on average between the singly and doubly protonated ASP25/ASP25′, the protonation state does not qualitatively change the correlations with either the RETI ΔG_{abs} or water conservation.

Once protonation states were assigned, Amber99 partial charges, as implemented in MOE, were applied to the protein and waters and AM1-BCC charges[56] to the ligand atoms. Next, for each protein−ligand complex, all hydrogen atoms and heavy atoms within 4.5 Å of any ligand heavy atom were minimized subject to decreasing harmonic restraints. Seven cycles of 500 steps of minimization were performed, decreasing the harmonic restraint by 10-fold at each stage from 10 000 kcal/mol to zero.

For comparison to the RETI results (see Table 1), complex preparation up to but not including minimization is referred to as using the "standard" protocol; including minimization is called the "minimized" protocol. Two additional filtering protocols were also employed following the "standard" preparation: (i) removing waters with two or more interwater hydrogen bonds and (ii) removing waters with wsolv values <3 kcal/mol. (See definition of wsolv below and in Table 2.)

Table 1. Correlation (R^2, Spearman r) of SZMAP Free Energies with RETI Absolute Free Energies for Crystallographic Waters

	protocol			
	standard[a]	minimized[a]	filtering out ≥2 inter-water H-bonds	filtering out SZMAP wsolv <3 kcal/mol
number of waters in data set	54	54	40	46
SZMAP n_ddG	0.43, 0.63	0.21, 0.49	0.66[b], 0.63	0.59[c], 0.79
SZMAP v_ddG	0.15, 0.37	0.21, 0.45	0.10, 0.37	0.14. 0.35
SZMAP dG	0.06, 0.25	0.09, 0.28	0.00, 0.07	0.03, 0.16

[a]No minimization with the standard protocol, while with the minimized protocol the ligand binding pocket was subject to minimization with decreasing harmonic restraints as described in the Methods. [b]For 1000 random samples leaving out 14 waters, mean R^2 = 0.43, stdev = 0.07. [c]For 1000 random samples leaving out 8 waters, mean R^2 = 0.43, stdev = 0.05.

Table 2. SZMAP-Calculated Properties Considered in This Study

property name	short name	description
szmap_neut_diff_free_energy	n_ddG	water: neutral probe free energy difference
szmap_neut_diff_entropy	n_TdS	water: neutral probe entropy difference
szmap_vac_diff_free_energy	v_ddG	water: vacuum probe free energy difference
szmap_free_energy	dG	free energy of explicit water vs continuum
szmap_entropy	TdS	entropy of explicit water vs continuum
szmap_neutral_entropy	n_dTdS	entropy of neutral probe vs continuum
szmap_wsolv	wsolv	water probe desolvation penalty
szmap_probe_burial	burial	fractional amount water probe is buried

SZMAP Methodology. Using SZMAP 1.0,[1] we calculated the thermodynamic properties of a water probe placed either on a grid or at crystallographic water oxygen position (after all crystallographic waters were removed from the complex). At a given probe location, SZMAP samples many orientations of the probe water. For each orientation, it calculates the energy of the system (E_j), which consists of the following components: vdW interactions of the water probe with the protein and ligand ($E_{\text{vdw protein−water}}$), electrostatic interactions of the water probe with the ligand, the protein, and the polarizable PB continuum solvent ($E_{\text{PB protein−water}}$), and the protein and water electrostatic desolvation terms (E_{psolv} and E_{wsolv}).

$$E_j = E_{\text{vdw protein−water}} + E_{\text{PB protein−water}} + E_{\text{psolv}} + E_{\text{wsolv}} \quad (1)$$

The ensemble of water probe orientations and corresponding energies is then used to construct a partition function (Q),

$$Q = \sum_{j=1}^{N_{\text{orient}}} e^{-\frac{E_j - E_{\text{min}}}{kT}} \quad (2)$$

from which SZMAP derives the probability of the probe being in each of the sampled orientations (prob_j)

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

$$\text{prob}_j = \frac{e^{-\frac{E_j - E_{\min}}{kT}}}{Q} \quad (3)$$

The probabilities and partition function are used in the calculation of SZMAP thermodynamic properties such as the probe water orientation entropy difference vs continuum water ($T\Delta S$, eq 4), the probe free energy difference (ΔG, eq 5), and the enthalpy difference (ΔH, eq 6).

$$T\Delta S = (-kT \sum_{j=1}^{N_{\text{orient}}} \text{prob}_j \ln \text{prob}_j) - kT \ln N_{\text{orient}} \quad (4)$$

$$\Delta G = -(kT \ln Q - E_{\min} - kT \ln N_{\text{orient}}) \quad (5)$$

$$\Delta H = \Delta G + T\Delta S \quad (6)$$

All SZMAP calculated thermodynamic terms discussed in the manuscript are listed in Table 2, and other relevant terms are described in Table S1 in the Supporting Information.

Furthermore, SZMAP can calculate the thermodynamic properties of a water probe relative to two different reference states (Figure 2). One is a vacuum probe, which is a spherical

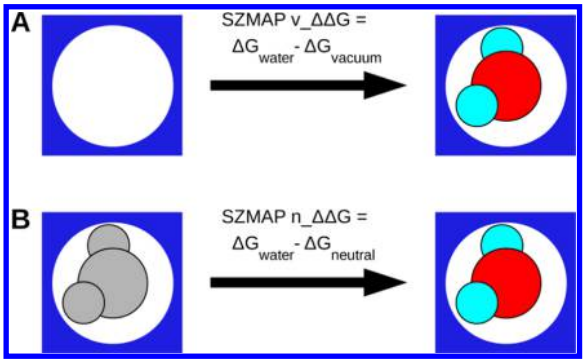


Figure 2. Reference states used in SZMAP. In (A), the vacuum reference state and, in (B), the neutral reference state, are depicted. The water probe is colored to reflect partial charges associated with its hydrogen and oxygen atoms, while the neutral reference state is uncolored to reflect that it is uncharged.

cavity that displaces the continuum solvent similar to the water probe, but lacking atoms, it makes no contributions to the vdW energy. This reference state was designed to allow a comparison within a system of a water molecule and empty space at the same location. The other reference state available in SZMAP is the "neutral water probe" which mimics an aliphatic group. The neutral water probe is constructed with the same three atomic centers as the water molecule, except that all partial charges are set to zero. Additionally, an identical set of orientations is sampled for the neutral water probe as for the water probe. Therefore, when the free energy of the neutral water probe is subtracted from that of the polar water probe, the vdW terms cancel, and the energy difference reported reflects the binding energy due to the partial charges of the water molecule.

SZMAP Grids: Water Positions. To evaluate the ability of SZMAP grids to predict crystallographic water positions, we selected two crystal structures from the training set, an HIV-1 protease structure (1EBW) and a scytalone dehydratase structure (3STD). Crystallographic waters in HIV-1 protease complex structures have been studied previously[57−59] and, of the HIV-1 protease structures in the training set, 1EBW has the largest number of waters near the ligand. SD was chosen because the hydrogen-bonding network of crystallographic waters in the active site of this protein plays an important role in inhibitor binding.[60−62] 3STD was selected out of all the SD structures in the training set because it is the highest resolution (1.65 Å).

Following the standard protein preparation protocol as described above, crystallographic waters were removed from PDB structures 1EBW and 3STD, and SZMAP free energy maps were calculated on a grid for regions in the binding sites within an 8 Å box around the ligand, using a grid spacing of 0.5 Å and 50 probe orientations per point. Crystallographic water positions were then compared to the SZMAP map minima.

Crystallographic Waters: At Coordinates Calculations. Our test set, as described above, includes 54 crystallographic waters from a total of 34 protein-ligand crystal structures of six different proteins. For each of these waters, SZMAP free energy terms were calculated at crystallographic water positions, by sampling 1000 orientations of the water probe at each position. This extensive sampling (1000 compared to 50 at each grid point) was feasible since many fewer positions were explored than for the grid-based site mapping described above. Thermodynamic values were determined for the water probe, the neutral probe, and the vacuum probe at each position.

Crystallographic Waters: Assessing Hydrogen Bond Networks. For each of the 54 waters in the test set, we calculated the number of potential interwater hydrogen bonds, defined by a distance of less than 3.8 Å between crystallographic positions of water oxygen atoms. We have also calculated the value of the SZMAP wsolv term, which is defined as the desolvation penalty that occurs as the charged water is brought closer to the neutral (all charges removed) protein surface. It is calculated as the sum of the difference between the charged water probe interaction with the neutral protein and the interaction of the same probe with bulk implicit solvent. SZMAP wsolv approaches zero for waters that are far from the protein−ligand system, and it increases in value for waters in contact with the solute atoms. The less contact a water probe has with the implicit solvent, the greater the value of SZMAP wsolv.

Crystallographic Waters: "Cloud Sampling" Sensitivity Analysis. To assess how sensitive SZMAP thermodynamic quantities are to small translations of the water probe, we have conducted a sampling procedure we refer to as "cloud sampling". For each of the 54 waters in the test set, we generated 200 new water positions within 0.5 Å of the original crystallographic position by randomly perturbing the Cartesian coordinates of the oxygen atom. This is illustrated in Figure 3 for the key "water A" in HIV protease. SZMAP was run on all of the perturbed water positions, and the distribution of SZMAP property values was analyzed to reveal the sensitivity of the calculations to small changes in atomic coordinates.

Crystallographic Waters: Conservation and B-factor. We compared SZMAP results with several types of experimental data from the test set. An assessment of whether each of these 54 waters is conserved or displaceable by ligands had been derived previously from a larger series of 441 crystal structures.[42] We investigated whether any of the SZMAP thermodynamic terms listed in Table 2 and Table S1, Supporting Information, could be used to predict water conservation.

We also examined the correlation of various thermodynamic parameters calculated using SZMAP to the B-factors of the water oxygens in the Factor Xa and HIV-1 protease structures

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

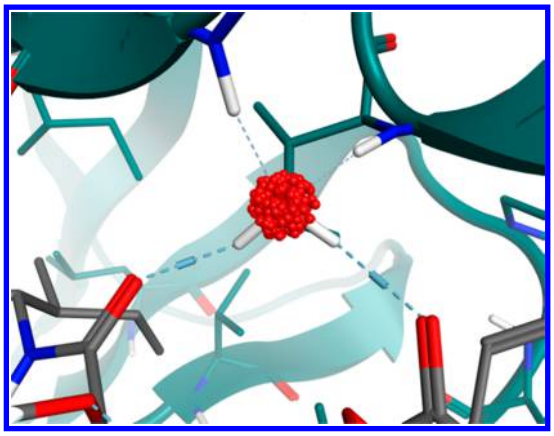


Figure 3. A cloud of 200 randomly sampled points within 0.5 Å of the crystallographic oxygen position of "water A" in the HIV-1 protease crystal structure 1EBW.

in the test set. Of the ten structures, we excluded the one for each protease with the least number of waters in a radius within 10 Å of their ligands (1KSN, n = 20; 1HPX, n = 6). We then calculated the SZMAP properties for all crystallographic waters with oxygen atoms in this radius for the remaining structures. We excluded waters for which SZMAP reported a steric clash with the protein (szmap_clash = CLASH) that was not resolved within the 1000 orientations sampled, as these would be high energy based on vdW interactions alone. At most, five waters were flagged for exclusion from a given structure, and on average, two. After this processing, the remaining four HIV-1 protease structures had between 33 and 47 waters, and the remaining four FXa structures had between 19 and 41 waters for analysis. Prior to any comparisons, the B-factors for each atom were normalized by dividing by the average heavy atom B-factor for that structure.

RESULTS AND DISCUSSION

To assess the accuracy of SZMAP, we compared water free energies calculated with SZMAP to those calculated with RETI, the Monte Carlo simulation-based method described above. We also compared SZMAP maps and various calculated thermodynamic terms to experimentally determined water positions, conservation, and B-factors. The test set compiled by Barillari et al.[42] is fairly diverse; some waters participate in hydrogen bonding with several other crystallographic water molecules, like "water K" from OppA (Figure 4), while others make contact with only the protein and the ligand, such as HIV-1 protease "water A" (Figure 3). We anticipated that modeling water networks might be challenging for SZMAP due to the use of the implicit solvent model.

The two methods we compare are quite distinct. SZMAP derives thermodynamic properties of the water probe by evaluating its partition function, while the calculations of Barillari et al.[42] use finite difference thermodynamic integration. While SZMAP calculates the rotational entropy of water, it does not account for the translational entropy. RETI computes both the rotational and translation entropy of water; as water orientations are explored, all rotational and translation degrees of freedom are sampled. Vibrational entropy is also likely important but is neglected by both methods. In SZMAP, the protein conformation is held fixed, while Barillari et al.[42] allowed for protein motion in their simulations, incorporating side chain angle and dihedral sampling as well as backbone Cα

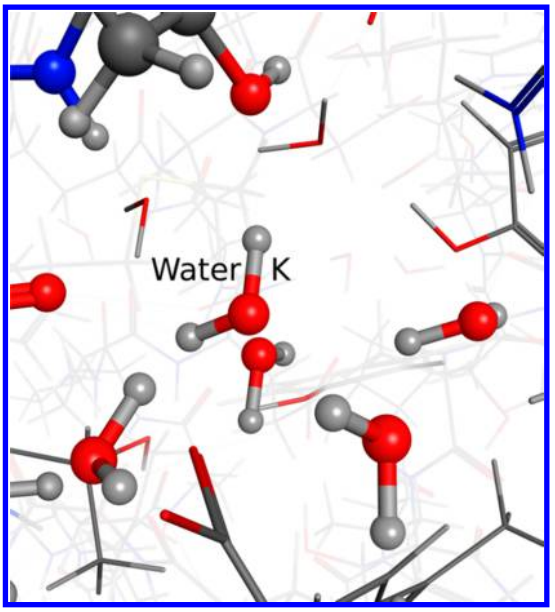


Figure 4. "Water K" from the structure of OppA (1B3L) can potentially make hydrogen bonds with four other crystallographic waters, depending on their orientations.

moves. In the RETI simulations, all water atoms are treated explicitly. In SZMAP, only atoms of the water probe are treated explicitly, while the remaining water molecules are modeled implicitly. The more rigorous RETI results, based on explicit simulations, are expected to be quantitatively more accurate than the SZMAP calculations that employ the additional approximations described above in order to realize significantly shortened computing times. Thus, the RETI calculations were used as a standard for evaluating the SZMAP results as discussed below.

Use of SZMAP Grids to Explore the Environment of Crystal Waters. A primary application of SZMAP, as the name suggests, is to map preferred water positions within a protein binding site. SZMAP can sample a grid of points on a protein−ligand complex. Isocontours of the calculated free energies of solvation and other thermodynamic terms can then be rendered to visualize how the values vary across the binding site. The SZMAP neutral probe free energy difference (n_ddG) measures the degree to which a normally, charged water probe is preferred over a water-shaped probe without charges. This quantity is negative (favorable) in hydrophilic regions of the binding site and positive (unfavorable) in hydrophobic areas while regions with n_ddG energies close to 0 reflect a preference for neither polar nor nonpolar molecules. A contour level of −0.5 kcal/mol was selected to identify those positions where water is significantly more favorable than a nonpolar probe. When we plotted a SZMAP n_ddG map contoured at less than −0.5 kcal/mol for an HIV-1 protease structure (1EBW) and an SD structure (3STD) as shown in Figure 5, more than half of the crystallographic waters within 5 Å of the ligand are located within the contours (14/24 and 3/3, respectively, see in particular key waters "A" and "H"). These mapping calculations run on the order of hours, depending on the number of grid points considered and processors employed (see the Supporting Information).

Correlation of SZMAP Free Energies with Absolute $\Delta G_{Binding}$. We computed the SZMAP free energy of solvation relative to a neutral probe (n_ddG) and a vacuum probe

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

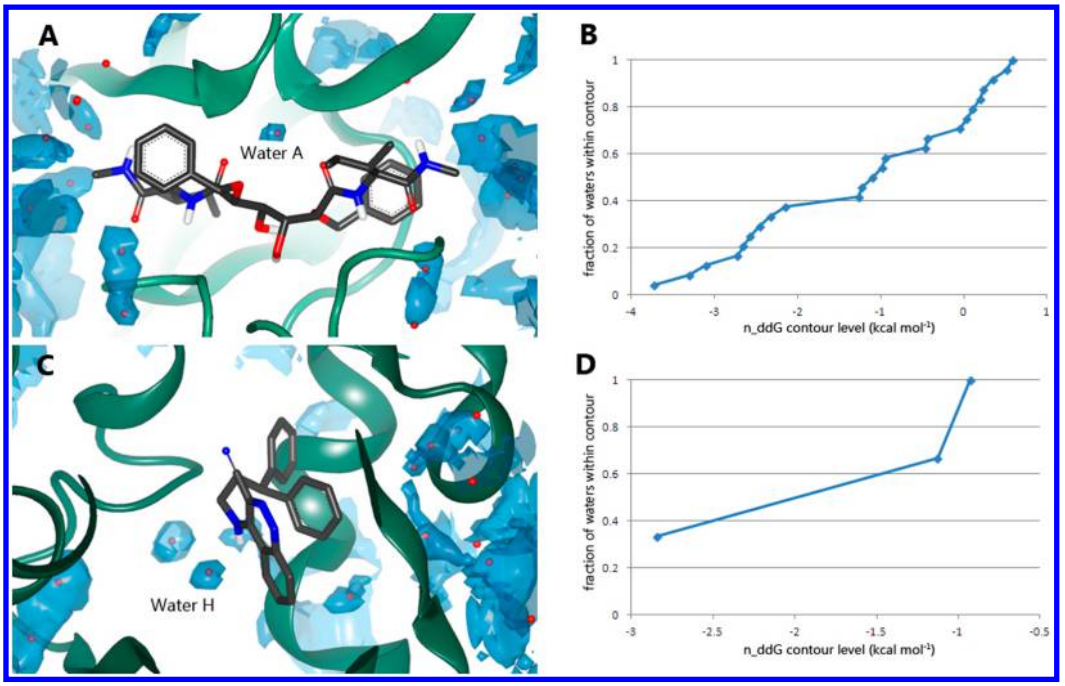


Figure 5. In the crystal structure of (A) HIV-1 protease (1EBW) and (C) of scytalone dehydratase (3STD), the positions of crystallographic waters largely correspond to regions where the SZMAP neutral probe free energy difference is less than −0.5 kcal/mol, as depicted by blue contours. Plots of the cumulative number of waters contained within the different contour levels are shown for 1EBW (B) and 3STD (D), illustrating that the majority of waters neighboring the ligand fall within the −0.5 kcal/mol contour level.

(v_ddG), respectively, and absolute (dG) for test set complexes prepared using the "standard" protocol and "minimized" protocol and filtering protocols outlined in the Methods section (Table 1). Correlations (R^2 and Spearman r[63]) were computed by comparing the SZMAP results to the RETI absolute free energy of water binding. We also performed a random leave-N-out analysis (see Table 1 notes) to ensure that the two filtering schemes were not simply improving the correlations by reducing the numbers of waters evaluated.

In addition to calculating the free energy relative to implicit solvent, SZMAP can also calculate the free energy of binding relative to neutral probe reference state (n_ddG) and relative to a vacuum reference state (v_ddG), as described in the Methods section. For the standard protocol, the correlation between SZMAP n_ddG and the RETI absolute free energy of binding is R^2 = 0.43 for all waters in the training set (Table 1 and Figure 6A.) The corresponding mean-difference analysis, as described by Bland and Altman,[64] further indicates that there is agreement between these two methods (Figure S1, Supporting Information). Minimization of the complexes prior to SZMAP calculations reduces this correlation to R^2 = 0.21. Minimization causes the water molecules to deviate from their crystallographic positions (the average RMSD was 0.44 Å, with individual RMSDs ranging from 0 to 1.15 Å), and thus SZMAP samples positions which differ significantly from those sampled by Barillari et al.[42] Conceptually, the most analogous SZMAP-calculated thermodynamic quantity to the simulations-based RETI free energy is the vacuum probe free energy difference (v_ddG). However, this value did not correlate well with the RETI absolute free energy of binding, R^2 = 0.15 (Figure 6B). The v_ddG term is essentially the n_ddG with a vdW component added; as such, this difference between n_ddG and v_ddG likely explains the poorer performance of v_ddG (R^2 0.15 vs 0.43). More specifically, the vdW component cancels out in the n_ddG value, since the value for the identical neutral probe orientation is subtracted from that for the water probe. The performance of v_ddG may be attributable to van der Waals clashes of the water probe with the protein−ligand environment, since the SZMAP calculations are performed at a fixed position. In addition, RETI uses the Lennard-Jones potential to calculate the vdW component, while SZMAP uses the MMFF94 "Buffered-14-7" form[65] of the vdW potential with parameters from SZYBKI.[66]

For the "standard" protocol, there was little correlation between SZMAP dG and the RETI ΔG_{abs} of binding, with R^2 = 0.06 for all 54 waters in the test set (Figure 6C.) Energy minimization of the complex structures prior to the SZMAP calculations did not improve the correlation substantially. These results may be expected though, because for RETI the reference state is a vacuum bubble at the position of water of interest, while for SZMAP the reference state is implicit solvent at the same position. The SZMAP dG is essentially the correction an explicit water probe introduces relative to the implicit solvent model.

Interestingly, we also found a relationship between the RETI-calculated absolute free energy of water binding with the displacement of waters from their crystallographic position following the minimization protocol (Figure 6D, R^2 = 0.33). Low energy water molecules that make multiple hydrogen bonds to the protein−ligand system are fairly well-constrained by the hydrogen bonding network and do not move very far from their initial positions. On the other hand, relatively high energy (unfavorable) water molecules that make fewer hydrogen bonds are free to move significantly during the optimization to improve the geometry of their hydrogen bonds.

One important potential limitation of the SZMAP method is in its treatment of networks of water molecules. There is a growing body of literature in which accurate modeling of networks of water molecules within a binding site have been shown to be critical for understanding SAR.[9,16,35,36,46,67,68] In a

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

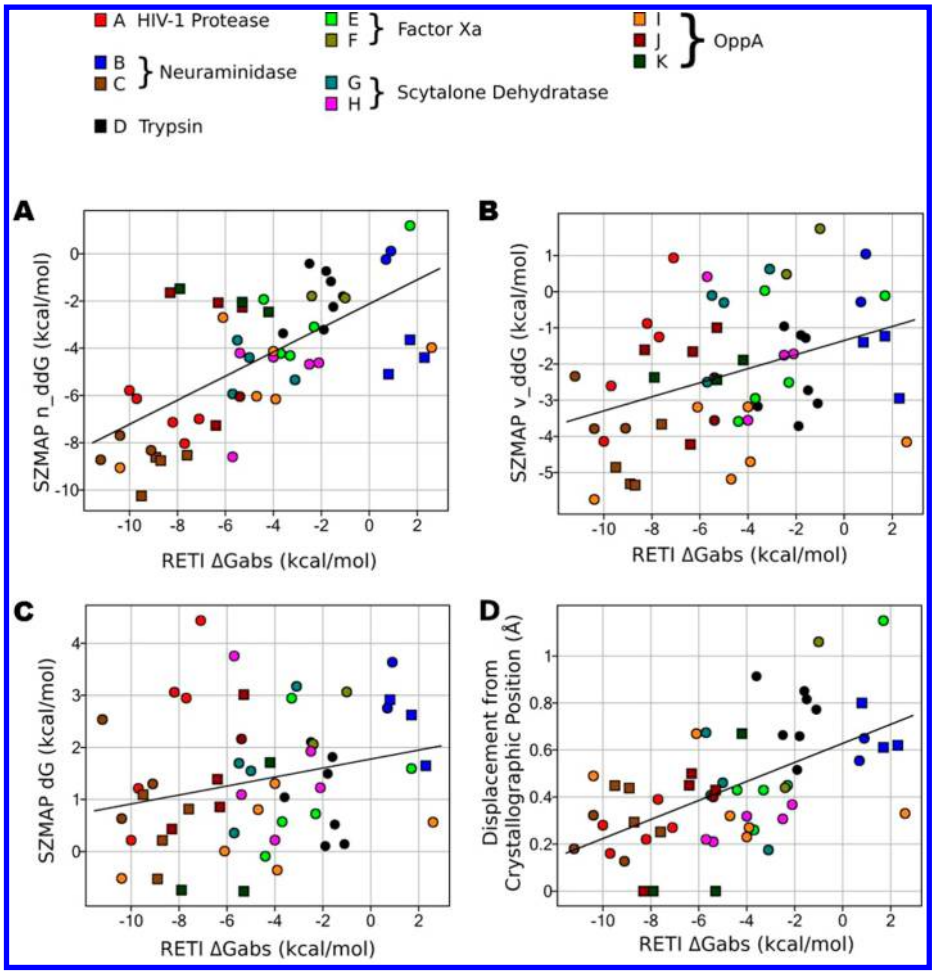


Figure 6. Comparison of (A) SZMAP neutral probe free energy difference (n_ddG), (B) SZMAP vacuum probe free energy difference (v_ddG), and (C) SZMAP water probe free energy (dG) using the standard preparation protocol without minimization and (D) the displacement from crystallographic position during minimization to the absolute free energies of binding found by Barillari et al.[42] using RETI for the test set water molecules. Waters that make hydrogen bonds with <2 other waters are shown as circles, while those that make two or more hydrogen bonds with other waters are shown as squares.

recent study of binding site waters by Snyder et al.,[69] the authors demonstrate that ligand modifications induce significant changes to a network of waters in the binding site, resulting in significant changes to the thermodynamic binding parameters. A key outstanding question remaining in that study is whether the rearrangement of those crystal water networks was driven by protein−water interactions or water−water interactions.

SZMAP uses a single water molecule probe complemented with a continuum solvent field (PB) to compute the solvation energy at all grid points in the calculation volume. One implication of this approach is that, for networks of multiple water molecules contained in a protein site, SZMAP calculates the energies of each individual water interacting with the protein but omits any calculation of explicit water−water interactions. This omission could be significant if those water−water interactions are not well-modeled by the PB continuum solvent model. However, the significance of this limitation is not entirely clear, as the magnitude of the protein−water interaction energy potentially dominates the solvation energy at the protein surface, which is precisely why models of bulk water behavior often fail when applied to waters located in confined protein pockets. The results of SZMAP grid calculations demonstrating that crystal water positions tend to fall in the minima of the SZMAP energy landscape imply either that the PB continuum solvent model sufficiently models water−water interactions or that water−protein interactions predominantly determine the structure of water networks. Additionally, the n_ddG energies computed by the at-coordinates calculations of the crystallographic water positions correlate with the desolvation free energies computed by RETI, which further supports the conclusion that the PB model reasonably approximates the water−water interactions.

Another way to assess whether water networks are important for a given water molecule is to count the number of potential hydrogen bonding partners from the set of crystallographic waters. We quantified this by counting the number of water oxygen atoms within 3.8 Å of the oxygen position of the water of interest. Of the 54 key waters examined in this study, 28, or 52%, did not make a hydrogen bond to another water. A recently published statistical analysis of 2.3 million water molecules found that close to 40% of captured (buried) waters with clear electron density made no hydrogen bond with a neighboring water.[70] In Figure 6, waters capable of making hydrogen bonds with two or more other crystallographic waters are plotted as squares. These tend to lie further from the regression line between SZMAP n_ddG and RETI ΔGabs than other waters. Eliminating them from the regression, which may

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

be justified since they may require explicit modeling of the water network, improves the R^2 from 0.43 to 0.66.

For the test set overall ("minimized"), the SZMAP wsolv term correlates with the number of hydrogen bonds a water molecule can potentially make with other crystallographic waters (Figure 7). As described in the Methods, wsolv

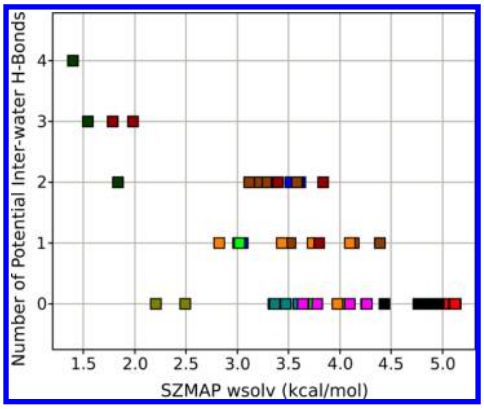


Figure 7. Number of potential interwater hydrogen bonds vs SZMAP wsolv calculated for the test set using the standard protocol. The color scheme is the same as in Figure 6.

represents the desolvation penalty that occurs as a charged water is brought closer to the neutral (all charges removed) protein surface. The more crystallographic waters that are near the probe water, the more open the site, and the lower is the value of SZMAP wsolv (the less the desolvation penalty). A notable exception is the Factor Xa "water F". This water is accessible to bulk solvent, and thus, the SZMAP desolvation penalty is relatively low (wsolv = 2.2 and 2.5 kcal/mol, respectively, for the two "water F" molecules from different Factor Xa structures represented as tan squares in Figure 7), even though it does not make hydrogen bonds with any other crystallographic waters.

Wsolv may in fact be a better method for identifying and filtering out crystallographic waters that participate in water networks prior to evaluating SZMAP ddG correlations with the RETI ΔG_{abs}. Wsolv values provide a continuous scale that is therefore more physically realistic than binning by the number of interwater hydrogen bonds. Discarding waters with wsolv less than 3 kcal/mol increases R^2 for SZMAP n_ddG vs ΔGabs from 0.43 to 0.59, which is similar to the value of 0.66 that resulted after filtering out waters making two or more interwater hydrogen bonds. The small degradation in the correlation may be because wsolv treats waters that interact only with bulk solvent similarly to waters which interact with a network of waters within the site. We performed a random leave-N-out analysis (see Table 1 notes) to ensure that the two filtering schemes were not simply improving the correlations by reducing the numbers of waters evaluated. Although with the random leave-N-out analysis the correlation of SZMAP n_ddG varied slightly depending on the subsets of waters chosen, both filtering protocols performed better overall than random selection.

Cloud Sampling Sensitivity Analysis. To assess how sensitive SZMAP is to small changes in probe position, we have conducted "cloud sampling", sampling orientations at 200 random points within 0.5 Å of the crystallographic position of each water oxygen atom. This 0.5 Å distance was chosen because it is the default solvent-mapping grid resolution. Varying the grid position slightly may eliminate unfavorable vdW interactions with the protein−ligand system, potentially improving the correlation of SZMAP v_ddG with RETI absolute free energy of binding. Figure 8A demonstrates that v_ddG can vary greatly with small displacements; the smallest and the greatest range of v_ddG values for a given water molecule were 4.3 and 45.8 kcal/mol, respectively. Using the cloud minimum average or Boltzmann-averaged values of v_ddG, however, rather than values at the crystallographic positions (cloud centers) did not improve correlations with the RETI ΔGabs (Table S2, Supporting Information). Because the vdW terms of neutral and water probes cancel out in the calculation of n_ddG, it is less sensitive to small changes in sampled position (Figure 8B), with the smallest and the greatest range between the minimum and the maximum n_ddG for a single water being 2.9 and 13.7 kcal/mol. Cloud sampling of the scytalone dehydratase structure 5STD "water H" produced the largest ranges in n_ddG and v_ddG values. This is caused by the particularly short distance between the crystallographic position of "water H" and the ligand (*N*-(1,1-diphenylpropan-2-yl)-6,7-difluoroquinazolin-4-amine). The distances from the oxygen of "water H" to the hydrogen and nitrogen of the ligand aliphatic amine are 1.55 and 2.53 Å, respectively. Several of the cloud points were sampled at significantly shorter distances and resulted in steric clashes. As may be expected, the value of n_dTdS, the entropy of the water probe relative to the neutral probe reference state, shows significantly less variability with small changes in probe position than the free energy term overall and, thus by inference, the enthalpy.

Conservation of Crystal Waters. In their work, Barillari et al. analyzed over 400 crystal structures and showed that waters with very favorable absolute free energies of binding (large negative values) are more likely to be conserved (Figure 9A).[42] As discussed above, the SZMAP n_ddG best reproduced the RETI absolute free energy of binding and, therefore, as expected, yielded similar results for prediction of water conservation (Figure 9B).

An examination of other terms calculated by SZMAP (see Table S1, Supporting Information) revealed that the neutral probe entropy difference (n_dTdS) term actually correlates best with water conservation. Conserved waters tend to have a neutral probe entropy difference less than −1.4 kcal/mol (Figure 9C). Out of the test set of 54 waters, only four violated this rule (93% accuracy). Crystallographic water positions within the binding site with a negative value for this entropy term would experience a penalty if the water at that position was displaced by a small aliphatic probe. The n_dTdS value represents the loss of orientational entropy between the water probe and neutral probe and, as such, indicates the degree to which solvent−solute electrostatic interactions restrict the rotational degrees of freedom of the water probe. That is, the term describes the change in solvation free energy due to rotational entropy loss caused by charge interactions. As might therefore be expected, the n_dTdS term was negative for all waters in the data set and ranged from −0.3 kcal/mol for "water F" in FXa 1KSN structure to −2.3 kcal/mol for "water C" in neuraminidase 1F8B structure. To ensure that these results were not overly influenced by the degree of rotational sampling employed in the calculation of n_dTdS, the number of orientations was varied from 10 to 10 000. As can be seen in Figure 10, the predicted n_dTdS values stabilize at around 360

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

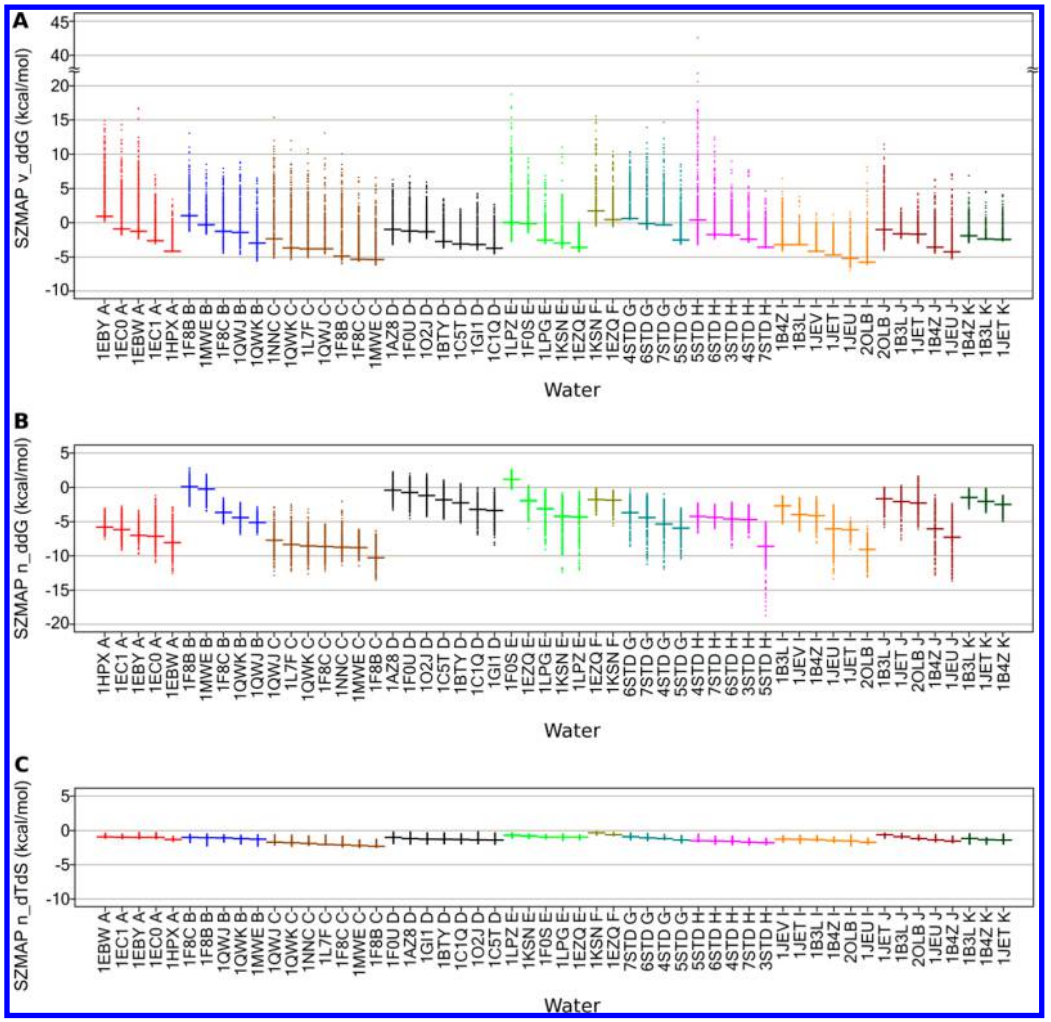


Figure 8. Spread of values obtained by cloud sampling for (A) SZMAP vacuum probe free energy difference (v_ddG), (B) SZMAP neutral probe free energy difference (n_ddG), and (C) SZMAP neutral probe entropy difference. Values for positions in the cloud are shown as dots, and values at crystallographic positions are shown as dashes. Within each plot, each crystallographic water position (A–K) is sorted from high energy to low. The color scheme is the same as in Figure 6.

orientations and thus the sampling utilized for the calculations described above (specifying 1000 orientations) is sufficient.

The four exceptions to the proposed n_dTdS < −1.4 kcal/mol conservation model are all due to OppA structures. "Water J" in 1B4Z is displaceable, but predicted to be conserved, with n_dTdS = −1.5, and although the "water I" in structures 1JET, 1JEV, and 1B3L each had n_dTdS values of −1.3 kcal/mol, they are considered conserved. This may be due to the peptidic nature of all of the OppA ligands. For the other proteins, a greater diversity in ligand structure was observed over the set of structures used to classify waters as conserved vs displaceable. Perhaps if a greater chemical diversity was explored by the OppA ligands, "water I" would be displaceable.

Another important consideration is that all four of waters that were classified incorrectly had values of n_dTdS close to the −1.4 kcal/mol threshold. The water of interest in the three additional CDK2 structures, "Water L", was considered by Barillari et al.[42] to be displaceable. This water was predicted correctly to be displaceable for structures 1H01 and 1V1K using n_dTdS, but not for the 1H08 structure. In 1H08, the n_dTdS value for the water was −1.53 kcal/mol, therefore on the conserved side. Including the CDK2 structures reduced overall prediction accuracy to 91%.

Thus, in general, the neutral probe entropy difference, n_dTdS, appears to be capable of discriminating between crystallographic waters that are conserved and those that are not across a large series of protein–ligand complexes. Like the RETI-based prediction method of Barillari et al.,[42] it uses a physics-based model and makes predictions about displacing waters from the *holo* protein binding sites, but the SZMAP method is faster and demonstrates better discrimination. In addition, the SZMAP predictions are based on a single physics-based descriptor.

Relating SZMAP-Calculated Properties and Experimental B-factors. The crystallographic B-factor applied to the X-ray scattering term for each atom (or for groups of atoms) in a structure describes the degree to which the electron density is localized (a low B-factor) or diffused (a high B-factor), and it is related to atomic motion as well as the structural heterogeneity between the unit cells in the crystal. The B-factors associated with crystallographic waters have previously been shown to correlate with the ability to displace the waters through ligand binding.[11,12,71] The relationship between all SZMAP-computed descriptors (Table S1, Supporting Information) and B-factors for the crystallographic waters was examined for the HIV-1 protease and FXa structures in the test set.

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552–1565

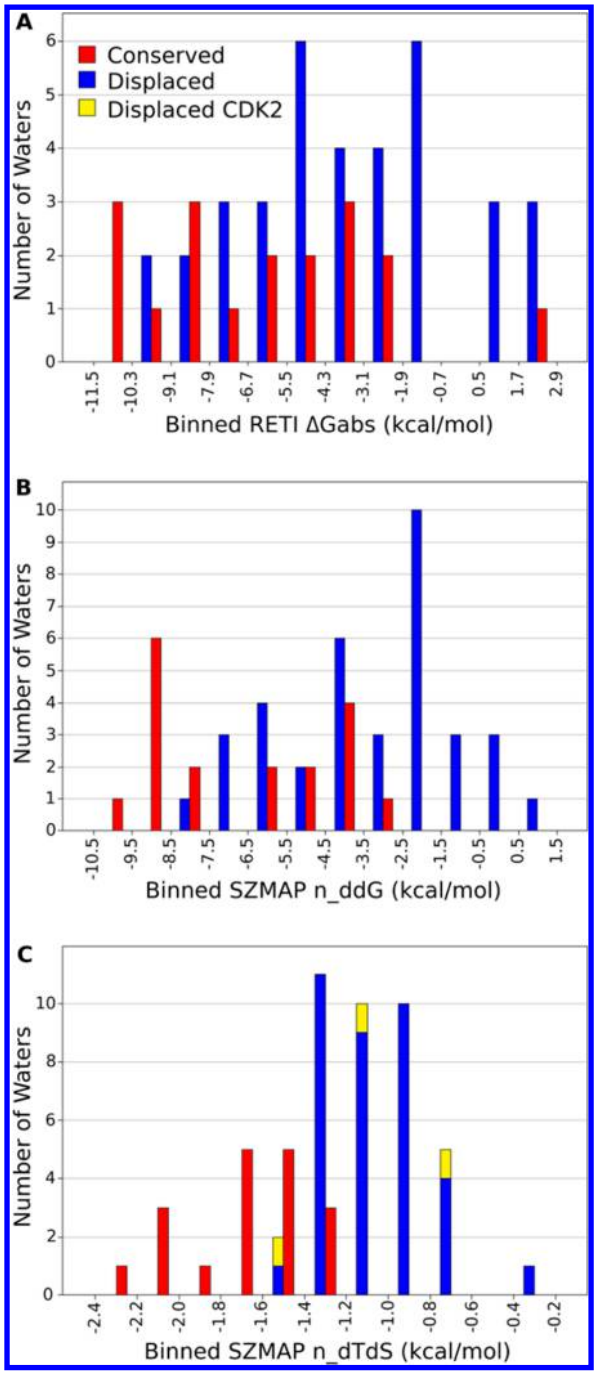


Figure 9. Dependence of water conservation on (A) the absolute free energy of water binding calculated by Barillari et al.[42], (B) SZMAP neutral probe free energy difference (n_ddG), and (C) SZMAP neutral probe entropy difference (n_dTdS), which appears to be the descriptor that correlates best with water conservation.

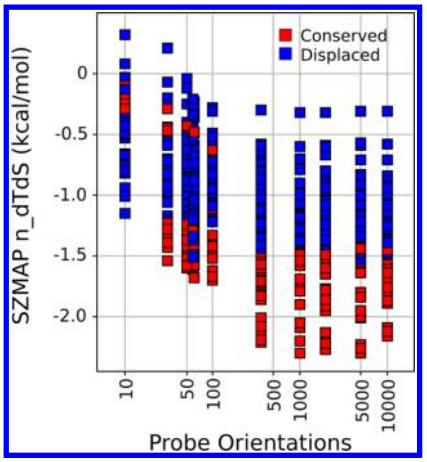


Figure 10. Orientational sampling sensitivity analysis.

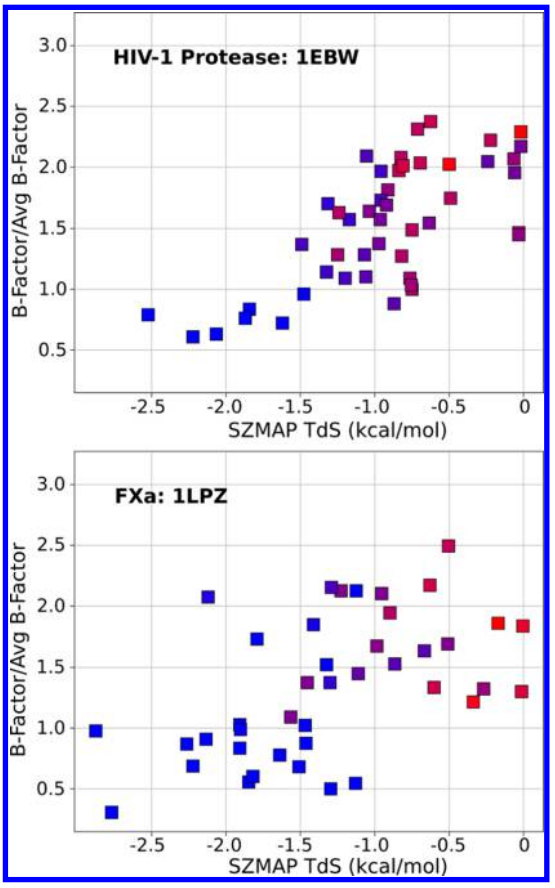


Figure 11. Plots of normalized B-factors of nonclashing crystallographic waters within 10 Å of ligands vs SZMAP entropy difference for HIV-1 protease crystal structure 1EBW and FXa crystal structure 1LPZ. Points are colored according to the SZMAP burial term for each water. Buried waters are blue, and exposed waters are red.

The B-factors for the crystallographic waters in the test set were found to correlate well with the SZMAP calculated entropy (TdS) for the water positions. Results for an HIV-1 protease and a FXa structure (the protein structures with the largest number of waters available for analysis) are shown in Figure 11. More buried waters, shown in blue, tended to have lower B-factors and lower entropy terms, as expected since the B-factor can be thought of as a measure of the relative vibrational motion of different parts of an X-ray structure. The same trend was seen for all structures examined from the test set (additional examples are shown in Figure S2, Supporting Information); TdS is near zero for waters that are not rotationally restricted, and it is negative for waters that are prevented from rotating by interactions with the protein and/or the ligand. As we expected, crystallographic waters that were severely restricted according to their low SZMAP entropy difference values also demonstrated smaller B-factors than those for which the SZMAP entropy difference was close to zero.

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

Similarly, the B-factors also correlate with the SZMAP burial term (Figures 12 and S3, Supporting Information), although

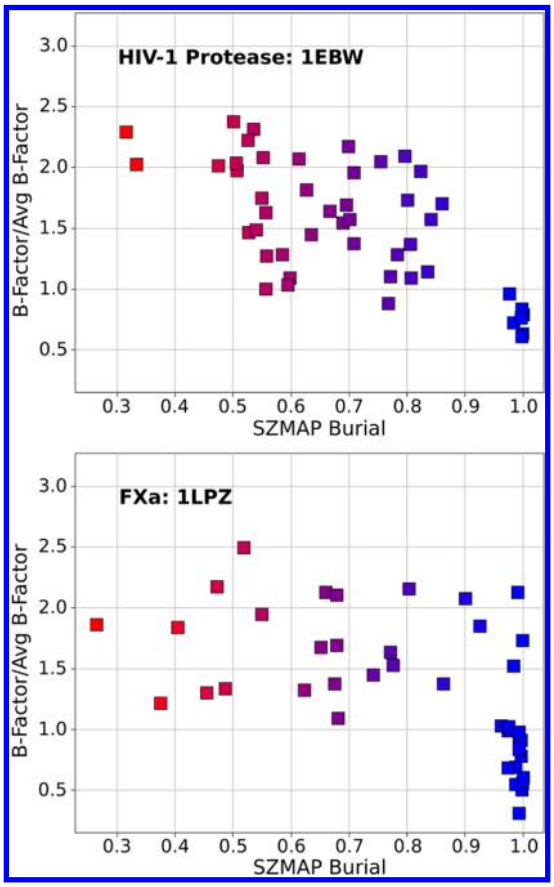


Figure 12. Plots of normalized B-factors of nonclashing crystallographic waters within 10 Å of ligands vs SZMAP burial for HIV-1 protease crystal structure 1EBW and FXa crystal structure 1LPZ. Points are colored according to the SZMAP burial term for each water. Buried waters are blue, and exposed waters are red.

not quite as well as with the entropy term. The burial term is a measure of the surface curvature. It is the fractional amount of the water probe that is buried; waters with a SZMAP burial term of zero are distant from the protein−ligand system, while those with a value of one are completely buried. Waters with higher burial terms (near one) tended to have lower B-factors. These results are similar to those of Parthasarathy and Murthy,[72] who found that buried protein residues have low B-factors while exposed ones have high B-factors. The entropy might be expected to correlate better with the B-factors than the burial, since one can imagine at least two types of fully buried waters, those that can satisfy optimal hydrogen bonds in only one orientation and those that can make similar interactions with the protein/ligand system in a number of different orientations. The latter would be expected to have a more favorable entropy term (a more negative TdS) and potentially a higher B-factor, and thus, entropy provides a greater granularity than the burial term.

The correlations observed between B-factors and entropy (TdS) and to a lesser extent burial term are impressive given the expected uncertainty in B-factors. The uncertainty in B-factors for a 1.5 Å resolution protein structure can be up to 15%.[73] Furthermore, the fact that the correlation is greater with the entropy than the burial term difference indicates that our results are in agreement with the Beuming et al.[41] study which reveals that all buried waters are not necessarily more stable; at least two categories of buried waters exist, again consistent with our findings described in the paragraph above.

CONCLUSIONS

We have evaluated the SZMAP program for its ability to predict, in a computationally efficient manner, solvation free energies and the effects of solvation on ligand binding, by comparing SZMAP results to both the results of the more rigorous and time-consuming RETI calculations of Barillari et al.,[42] as well as to experimental data. SZMAP makes many approximations that the RETI method does not: treating the protein as a rigid body and the solvent as a high-dielectric continuum field, as well as assuming that the freedom of translational motion does not have significant influence on water binding energy. Despite these assumptions, the SZMAP neutral probe free energy difference (n_ddG) showed reasonable correlation with the results of the RETI calculations, especially for waters that could not make more than one hydrogen bond with other crystallographic waters. Sampling many points around water crystallographic positions did not improve the correlation of the SZMAP n_ddG term with the RETI absolute free energy of binding. However, it did serve to illustrate the overall sensitivity of the free energies to water position and highlight that the n_ddG term displays substantially lower sensitivity to subtle changes in water position due to the cancellation of vdW terms relative to the vacuum-referenced free energy (v_ddG). Sensitivity to the number of orientations sampled, for calculations at a limited number of probe or crystallographic water positions, can be overcome by increasing sampling: 1000 orientations would not be practical for full grid-sampling calculations, however.

SZMAP results also correlated with experimental data in potentially useful ways. SZMAP predicts that waters with lower B-factors have less entropy and tend to be more buried, and crystallographic waters tend to occupy hydrophilic positions on a SZMAP n_ddG contour map. In addition, SZMAP was able to predict water conservation across a series of protein−ligand complexes. The SZMAP neutral probe entropy difference (n_dTdS) is a single physics-based descriptor that provides, using a simple threshold-value model, accurate predictions of water conservation in active sites of protein structures. Thus, we anticipate using SZMAP to attempt to predict optimal directions for derivatizing ligands during lead optimization. Having access to accurate and rapid information about which waters to displace and which to retain during lead optimization could enhance and make more efficient the drug discovery process.

ASSOCIATED CONTENT

Supporting Information

A summary of additional SZMAP-calculated properties considered in this study; correlations of SZMAP free energies within a "cloud" of points near each crystallographic water with RETI ΔG_{abs}; mean difference analysis associated with the data in Figure 6; plots of normalized B-factors vs SZMAP entropy difference or SZMAP burial for additional HIV-1 protease and Factor Xa crystal structures; details regarding SZMAP calculation timings and versions. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/ci500746d.

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

AUTHOR INFORMATION

Corresponding Author

*E-mail: michelle.lamb@astrazeneca.com.

Present Addresses

[#]A.S.B.: CMD Bioscience, Inc., 5 Science Park, New Haven, CT 06511.

[§]D.T.M.: Alkermes, 852 Winter Street, Waltham, MA 02451.

[⊥]D.J.-M.: EnBiotix, Inc., 700 Main Street, Cambridge, MA 02139.

Author Contributions

[‡]A.S.B. and D.T.M. contributed equally to this manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank Anthony Nicholls, Mike Word, and Matt Geballe of OpenEye for stimulating discussions and access to early versions of the SZMAP software and Professor Jonathan Essex, University of Southampton, for his insights into solvation thermodynamics. We would also like to thank Chris Williams from the Chemical Computing Group, as well as Ryszard Czerminski from AstraZeneca, for providing SVL scripts and assistance with the MOE software.

REFERENCES

(1) *SZMAP, 1.0.0*; OpenEye Scientific Software Inc.: Santa Fe, NM, USA, 2011.

(2) Fernández, A. Epistructural Tension Promotes Protein Associations. *Phys. Rev. Lett.* **2012**, *108*, 188102.

(3) Meiering, E. M.; Wagner, G. Detection of Long-Lived Bound Water Molecules in Complexes of Human Dihydrofolate Reductase with Methotrexate and NADPH. *J. Mol. Biol.* **1995**, *247*, 294−308.

(4) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(5) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Bioinf.* **1991**, *11*, 29−34.

(6) Evensen, E.; Joseph-McCarthy, D.; Weiss, G. A.; Schreiber, S. L.; Karplus, M. Ligand Design by a Combinatorial Approach Based on Modeling and Experiment: Application to HLA-DR4. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 395−418.

(7) Bitetti-Putzer, R.; Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Functional Group Placement in Protein Binding Sites: A Comparison of GRID and MCSS. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 935−960.

(8) Joseph-McCarthy, D.; Fedorov, A. A.; Almo, S. C. Comparison of Experimental and Computational Functional Group Mapping of an RNase A Structure: Implications for Computer-Aided Drug Design. *Protein Eng., Des. Sel.* **1996**, *9*, 773−780.

(9) Wallnoefer, H. G.; Liedl, K. R.; Fox, T. A GRID-Derived Water Network Stabilizes Molecular Dynamics Computer Simulations of a Protease. *J. Chem. Inf. Model.* **2011**, *51*, 2860−2867.

(10) Lensink, M. F.; Moal, I. H.; Bates, P. A.; Kastritis, P. L.; Melquiond, A. S. J.; Karaca, E.; Schmitz, C.; van Dijk, M.; Bonvin, A. M. J. J.; Eisenstein, M.; Jiménez-García, B.; Grosdidier, S.; Solernou, A.; Pérez-Cano, L.; Pallara, C.; Fernández-Recio, J.; Xu, J.; Muthu, P.; Praneeth Kilambi, K.; Gray, J. J.; Grudinin, S.; Derevyanko, G.; Mitchell, J. C.; Wieting, J.; Kanamori, E.; Tsuchiya, Y.; Murakami, Y.; Sarmiento, J.; Standley, D. M.; Shirota, M.; Kinoshita, K.; Nakamura, H.; Chavent, M.; Ritchie, D. W.; Park, H.; Ko, J.; Lee, H.; Seok, C.; Shen, Y.; Kozakov, D.; Vajda, S.; Kundrotas, P. J.; Vakser, I. A.; Pierce, B. G.; Hwang, H.; Vreven, T.; Weng, Z.; Buch, I.; Farkash, E.; Wolfson, H. J.; Zacharias, M.; Qin, S.; Zhou, H.; Huang, S.; Zou, X.; Wojdyla, J. A.; Kleanthous, C.; Wodak, S. J. Blind Prediction of Interfacial Water Positions in CAPRI. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 620−632.

(11) 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.

(12) Garcia-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: A Novel Method for Distinguishing between Bound and Displaceable Water Molecules in the Crystal Structure of the Binding Site of Protein-Ligand Complexes. *J. Mol. Model.* **2003**, *9*, 172−182.

(13) Rossato, G.; Ernst, B.; Vedani, A.; Smieško, M. AcquaAlta: A Directional Approach to the Solvation of Ligand-Protein Complexes. *J. Chem. Inf. Model.* **2011**, *51*, 1867−1881.

(14) Allen, F. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.

(15) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: Is LogPo/W More than the Sum of its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651−661.

(16) Kellogg, G. E.; Chen, D. L. The Importance of being Exhaustive. Optimization of Bridging Structural Water Molecules and Water Networks in Models of Biological Systems. *Chem. Biodiversity* **2004**, *1*, 98−105.

(17) Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the Energetics of Water-Protein and Water-Ligand Interactions with the "Natural" HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules. *J. Mol. Biol.* **2006**, *358*, 289−309.

(18) 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.

(19) Bayden, A.; Fornabaio, M.; Scarsdale, J.; Kellogg, G. Web Application for Studying the Free Energy of Binding and Protonation States of Protein−ligand Complexes Based on HINT. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 621−632.

(20) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS One* **2012**, *7*, e32036.

(21) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(22) Grochowski, P.; Trylska, J. Continuum Molecular Electrostatics, Salt Effects, and Counterion Binding–A Review of the Poisson-Boltzmann Theory and its Modifications. *Biopolymers* **2008**, *89*, 93−113.

(23) Baker, N. A. Improving Implicit Solvent Simulations: A Poisson-Centric View. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137−143.

(24) Kleinjung, J.; Fraternali, F. Design and Application of Implicit Solvent Models in Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2014**, *25*, 126−134.

(25) Friesner, R. A.; Abel, R.; Goldfeld, D. A.; Miller, E. B.; Murrett, C. S. Computational Methods for High Resolution Prediction and Refinement of Protein Structures. *Curr. Opin. Struct. Biol.* **2013**, *23*, 177−184.

(26) Chen, J.; Brooks, C. L., III; Khandogin, J. Recent Advances in Implicit Solvent-Based Methods for Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140−148.

(27) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.

(28) Beglov, D.; Roux, B. An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.

(29) Chandler, D.; McCoy, J. D.; Singer, S. J. Density Functional Theory of Nonuniform Polyatomic Systems. I. General Formulation. *J. Chem. Phys.* **1986**, *85*, 5971−5976.

(30) Sindhikara, D. J.; Hirata, F. Analysis of Biomolecular Solvation Sites by 3D-RISM Theory. *J. Phys. Chem. B* **2013**, *117*, 6718−6723.

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

(31) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: Montreal, QC, Canada, 2010.
(32) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(33) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(34) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6*, 607−624.
(35) Huggins, D. J. Application of Inhomogeneous Fluid Solvation Theory to Model the Distribution and Thermodynamics of Water Molecules Around Biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15106−15117.
(36) Huggins, D. J. Benchmarking the Thermodynamic Analysis of Water Molecules Around a Model Beta Sheet. *J. Comput. Chem.* **2012**, *33*, 1383−1392.
(37) Cui, G.; Swails, J. M.; Manas, E. S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 5539−5549.
(38) Pearlstein, R. A.; Sherman, W.; Abel, R. Contributions of Water Transfer Energy to Protein-Ligand Association and Dissociation Barriers: Watermap Analysis of a Series of p38α MAP Kinase Inhibitors. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 1509−1526.
(39) Yang, Y.; Lightstone, F. C.; Wong, S. E. Approaches to Efficiently Estimate Solvation and Explicit Water Energetics in Ligand Binding: The use of WaterMap. *Expert Opin. Drug Discovery* **2013**, *8*, 277−287.
(40) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(41) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic Analysis of Water Molecules at the Surface of Proteins and Applications to Binding Site Prediction and Characterization. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 871−883.
(42) 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.
(43) 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.
(44) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson-Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608−640.
(45) Mason, J. S.; Bortolato, A.; Congreve, M.; Marshall, F. H. New Insights from Structural Biology into the Druggability of G Protein-Coupled Receptors. *Trends Pharmacol. Sci.* **2012**, *33*, 249−260.
(46) Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water Network Perturbation in Ligand Binding: Adenosine A2A Antagonists as a Case Study. *J. Chem. Inf. Model.* **2013**, *53*, 1700−1713.
(47) Efstathiou, A.; Gaboriaud-Kolar, N.; Smirlis, D.; Myrianthopoulos, V.; Vougogiannopoulou, K.; Alexandratos, A.; Kritsanida, M.; Mikros, E.; Soteriadou, K.; Skaltsounis, A. An Inhibitor-Driven Study for Enhancing the Selectivity of Indirubin Derivatives Towards Leishmanial Glycogen Synthase Kinase-3 Over Leishmanial Cdc2-Related Protein Kinase 3. *Parasites Vectors* **2014**, *7*, 234.
(48) *Maestro, 9.2*; Schrödinger LLC: New York, NY, 2011.
(49) *Prime, 3.0*; Schrödinger LLC: New York, NY, 2011.
(50) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320*, 597−608.
(51) Labute, P. Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 187−205.
(52) Zheng, Y.; Bruice, T. C. Role of a Critical Water in Scytalone Dehydratase-Catalyzed Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4158−4163.
(53) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. Thermodynamic Linkage between the Binding of Protons and Inhibitors to HIV-1 Protease. *Protein Sci.* **1999**, *8*, 180−195.
(54) Yamazaki, T.; Nicholson, L. K.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Torchia, D. A. NMR and X-Ray Evidence that the HIV Protease Catalytic Aspartyl Groups are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor. *J. Am. Chem. Soc.* **1994**, *116*, 10791−10792.
(55) Wittayanarakul, K.; Hannongbua, S.; Feig, M. Accurate Prediction of Protonation State as a Prerequisite for Reliable MM-PB(GB)SA Binding Free Energy Calculations of HIV-1 Protease Inhibitors. *J. Comput. Chem.* **2008**, *29*, 673−685.
(56) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(57) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 Protease: A Major Success of Structure-Assisted Drug Design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.
(58) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.
(59) Lu, Y.; Yang, C.; Wang, S. Binding Free Energy Contributions of Interfacial Waters in HIV-1 Protease/Inhibitor Complexes. *J. Am. Chem. Soc.* **2006**, *128*, 11830−11839.
(60) 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.
(61) Wawrzak, Z.; Sandalova, T.; Steffens, J. J.; Basarab, G. S.; Lundqvist, T.; Lindqvist, Y.; Jordan, D. B. High-Resolution Structures of Scytalone Dehydratase-Inhibitor Complexes Crystallized at Physiological pH. *Proteins: Struct., Funct., Bioinf.* **1999**, *35*, 425−439.
(62) Jennings, L. D.; Wawrzak, Z.; Amorose, D.; Schwartz, R. S.; Jordan, D. B. A New Potent Inhibitor of Fungal Melanin Biosynthesis Identified through Combinatorial Chemistry. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2509−2514.
(63) Spearman, C. The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **1904**, *15*, 72−101; *Reprinted. Int. J. Epidemiol.* **2010**, *39*, 1137−1150.
(64) Bland, J. M.; Altman, D. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet* **1986**, *327*, 307−310.
(65) Halgren, T. A. Merck Molecular Force Field. II. MMFF94 Van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(66) *SZYBKI, 1.8.0.2*; OpenEye Scientific Software Inc.: Santa Fe, NM, USA, 2015.
(67) Biela, A.; Sielaff, F.; Terwesten, F.; Heine, A.; Steinmetzer, T.; Klebe, G. Ligand Binding Stepwise Disrupts Water Network in Thrombin: Enthalpic and Entropic Changes Reveal Classical Hydrophobic Effect. *J. Med. Chem.* **2012**, *55*, 6094−6110.
(68) Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M. Water Networks Contribute to Enthalpy/Entropy Compensation in Protein-Ligand Binding. *J. Am. Chem. Soc.* **2013**, *135*, 15579−15584.
(69) Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Héroux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the Hydrophobic Effect in the Biomolecular Recognition of Arylsulfonamides by Carbonic Anhydrase. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17889−17894.
(70) Nittinger, E.; Schneider, N.; Lange, G.; Rarey, M. Evidence of Water Molecules—A Statistical Evaluation of Water Molecules Based on Electron Density. *J. Chem. Inf. Model.* **2015**, *55*, 771−783.

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565

(71) Karplus, P. A.; Faerman, C. Ordered Water in Macromolecular Structure. *Curr. Opin. Struct. Biol.* **1994**, *4*, 770−776.
(72) Parthasarathy, S.; Murthy, M. R. N. Analysis of Temperature Factor Distribution in High-Resolution Protein Structures. *Protein Sci.* **1997**, *6*, 2561−2567.
(73) Kuriyan, J.; Karplus, M.; Petsko, G. A. Estimation of Uncertainties in X-Ray Refinement Results by use of Perturbed Structures. *Proteins: Struct., Funct., Bioinf.* **1987**, *2*, 1−12.

DOI: 10.1021/ci500746d
J. Chem. Inf. Model. 2015, 55, 1552−1565