JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

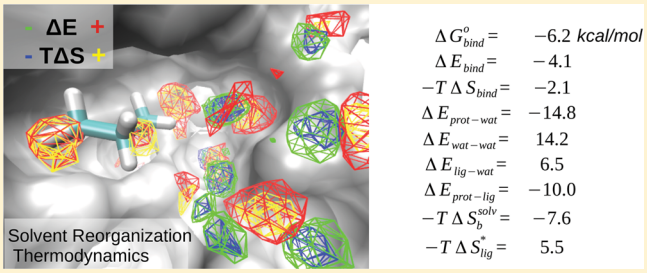
# Spatial Analysis and Quantification of the Thermodynamic Driving Forces in Protein–Ligand Binding: Binding Site Variability

E. Prabhu Raman* and Alexander D. MacKerell, Jr.

Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20 Penn Street HSF II, Baltimore, Maryland 21201, United States

Supporting Information

**ABSTRACT:** The thermodynamic driving forces behind small molecule–protein binding are still not well-understood, including the variability of those forces associated with different types of ligands in different binding pockets. To better understand these phenomena we calculate spatially resolved thermodynamic contributions of the different molecular degrees of freedom for the binding of propane and methanol to multiple pockets on the proteins Factor Xa and p38 MAP kinase. Binding thermodynamics are computed using a statistical thermodynamics based end-point method applied on a canonical ensemble comprising the protein–ligand complexes and the corresponding free states in an explicit solvent environment. Energetic and entropic contributions of water and ligand degrees of freedom computed from the configurational ensemble provide an unprecedented level of detail into the mechanisms of binding. Direct protein–ligand interaction energies play a significant role in both nonpolar and polar binding, which is comparable to water reorganization energy. Loss of interactions with water upon binding strongly compensates these contributions leading to relatively small binding enthalpies. For both solutes, the entropy of water reorganization is found to favor binding in agreement with the classical view of the "hydrophobic effect". Depending on the specifics of the binding pocket, both energy–entropy compensation and reinforcement mechanisms are observed. It is notable to have the ability to visualize the spatial distribution of the thermodynamic contributions to binding at atomic resolution showing significant differences in the thermodynamic contributions of water to the binding of propane versus methanol.




## 1. INTRODUCTION

The noncovalent association of macromolecules with ligands plays an important role in biological functions. Although binding thermodynamics has been rigorously formulated,[1] the quantification of the different component driving forces is largely lacking. Of the forces driving noncovalent ligand binding, hydrophobic association forms an important class of interactions, which involves the binding of a nonpolar ligand to a nonpolar binding pocket. The classical view of the so-called "hydrophobic effect" is that water molecules are more structured at binding interfaces than in bulk, resulting in an entropic driving force to cause the association of the solutes to minimize the solvent exposed surface area. However, more recent explanations[2] have noted the complicating roles of size, polarity, and surface topography of the associating species that make the binding process context specific. Some computational studies have indicated that the enthalpic and entropic contributions of the associating species and water may be quite different for protein–ligand association than what the entropy dominated classical view in the context of small hydrophobic solutes suggests.[3]

Inhomogeneous solvation theory (IST)[4,5] and related methods[6] have been used to calculate the thermodynamic properties of water at ligand binding interfaces in proteins[7] and to relate those properties to experimentally measured ligand binding affinities. A study that combined isothermal titration calorimetry (ITC), X-ray crystallography, and IST based computational analysis[8] measured an enthalpy dominated hydrophobic binding that could be explained by the calculated energy changes of local water molecules. However, analysis of individual local water molecules performed in these studies, while being both qualitatively and quantitatively informative, does not rigorously compute binding thermodynamics. Taken together, these studies show that the mechanism of binding in the context of protein–ligand association is still not completely understood.

In an attempt to decipher the molecular mechanism behind the binding of ligands to proteins and derive general principles, McCammon and co-workers investigated the binding thermodynamics of idealized spherical ligands to hemispherical model cavities with zero or unit charge.[9,10] They obtained the thermodynamic signature of a neutral probe binding to a neutral cavity which involved compensating favorable enthalpic and unfavorable entropic components. Favorable binding was due to the water reorganization energy contribution being

Received: November 25, 2014
Published: January 27, 2015





DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608–2621

larger in magnitude than the entropic component. Introducing a unit charge on the probe and/or the cavity resulted in large changes in the thermodynamic signature, compared to relatively smaller changes in binding affinity. While this manuscript was in preparation, Michel et al. reported on the use of the Grid Cell Theory method[11] to study the thermodynamic signature of idealized ligands binding to model cavities.[12] Their tests with a range of cavities with differing polarity and geometry resulted in significantly varying thermodynamic signatures highlighting the lack of a singular explanation for hydrophobic association. However, it is not clear if studies involving such idealized host−guest systems are representative of association between real ligands and real hydrophobic or polar cavities present in proteins.

In this work, we undertake analyses of ligands binding to a relatively large number of protein pockets of different geometry and topography with the goal of understanding the contributions of the different thermodynamic driving forces. Two well-studied and pharmaceutically important proteins, Factor Xa (FXa) and P38 MAP kinase alpha (P38MK), are selected as examples. Molecular dynamics (MD) simulations of these proteins in a solution of propane and methanol, respectively, representing hydrophobic and neutral polar ligands, are performed to identify the binding sites of these molecules on the proteins. The use of this approach to select binding sites to which ligands bind spontaneously makes the set of investigated association reactions relevant and representative of protein−ligand binding in biochemical systems. Following the identification of the binding sites, additional MD simulations are undertaken to generate a solvated configurational ensemble of the protein−ligand complexed states. MD simulations are also performed to obtain the configurational ensemble of the ligand-free protein and the ligand in solution. The canonical ensembles generated in these end states are subjected to an end-point thermodynamic analysis detailed below, which calculates the energy, entropy, and free energy of binding. The analysis yields a detailed view of the energetic and entropic contributions of direct protein−ligand interactions and water reorganization upon binding, including the variability of the thermodynamic contributions in different binding sites. Further, the formulation decomposes the contributions spatially and allows for local analysis of the thermodynamics at an atomic level of detail.

The paper is organized as follows. In the Theory and Methods section, we describe the method for the calculation of binding thermodynamics and its different components. In the Results section, the method is first validated by reproducing experimental solvation free energies of 6 small molecules. Next, we describe the simulations and calculations involving the binding of ligands to the two proteins and quantify the different energetic and entropic contributions to binding. The Discussion section presents further analysis of our results and elaborates on the implications of our findings.

## 2. THEORY AND METHODS

The binding free energy of noncovalent association is expressed as the sum of enthalpic and entropic contributions. In the context of macromolecule−small molecule binding, where the pressure−volume work is negligible,[1] enthalpy can be replaced by energy to a good approximation. The standard free energy of binding is given as

$$\Delta G_b^o = \Delta E_b - T\Delta S_b \quad (1)$$

where $\Delta E_b$ is the energy change and $\Delta S_b$ is the entropy change during the binding process. This work follows an "end-point" approach to the calculation of binding thermodynamics where we analyze the configurational distribution of the solutes and solvent obtained in the canonical ensemble in the complexed and free states. The energetic and entropic differences between these end states are calculated to yield the free energy of binding. Figure 1 depicts the binding of a ligand to a solvent exposed protein. During this process, the protein (P, center) and the ligand (L, left) undergo conformational changes. L leaves the solvated environment and binds to P resulting in the complexed PL state (right) where water molecules are displaced and reorganized. Our goal is to understand the contributions of different degrees of freedom to binding. Particularly, it is of interest to learn about the magnitudes of the driving forces implicated in the binding of hydrophobic and polar molecules to diverse protein pockets in which water is an important player. Here, we apply a variation of Grid Inhomogeneous Solvation Theory (GIST[13]) to calculate thermodynamics associated with water, which is formally defined for a fixed solute conformation. In principle, it is possible to perform such an analysis for multiple conformations, but in this initial study, we restrict ourselves to using restrained protein geometries during the MD simulations. The ligands chosen are small such that they do not have rotatable bonds involving heavy atoms.

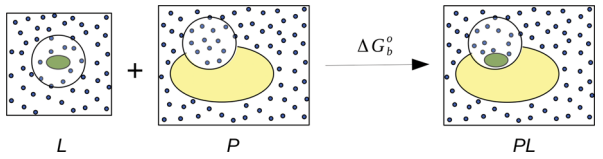



**Figure 1.** Schematic showing protein−ligand binding. The free ligand (L), free protein (P), and the protein−ligand complex (PL), all of which are solvated are shown. The protein is shown in yellow, the ligand in green, and water molecules as blue circles. The binding thermodynamics and free energy $\Delta G_b^o$ are obtained using an end-point method which calculates excess energy and entropy in a local spherical region which is also shown.

IST-based approaches calculate the energetic and entropic perturbation introduced in bulk water due to the presence of a solute. GIST is employed presently to calculate the excess energy and entropy in a spherical region surrounding the ligand (L), the solvated protein pocket (P), and the protein−ligand complex (PL), as schematically shown in Figure 1. The fact that solvent far away from the solute attains bulk-like character allows for restricting the calculation to a local surrounding region. The difference in the excess water energy and entropy between the PL, P, and L states results in the solvation free energy change $\Delta\Delta G_b^{solv}$ contributing to binding. To calculate the binding free energy $\Delta G_b^o$, we analyze the PL ensemble to calculate the energy and entropy of the ligand in the bound state with respect to an ideal gas state at the standard concentration of 1 M. The thermodynamic cost of transfer of the ligand from solution to an ideal gas state is accounted for through the calculation of L (left in Figure 1). The following text describes in detail the approach used to calculate the different contributions to the binding energy and entropy.

**2.1. Binding Energy.** The binding energy $\Delta E_b$ is defined as

$$\Delta E_b = \langle E\rangle_{PL} - \langle E\rangle_P - \langle E\rangle_L = \Delta E_P + \Delta E_L + E_{PL} + \Delta\Delta E_b^{solv} \quad (2)$$

where $\langle E\rangle_{PL}$, $\langle E\rangle_P$, and $\langle E\rangle_L$ are the ensemble averages of energies in the end states. The second equality is introduced so as to decouple the energy calculation of the solute from solvent. $\Delta E_P$ and $\Delta E_L$ correspond to the difference in internal or intramolecular energy between the complexed and free states for the protein and ligand, respectively. $E_{PL}$ is the protein−ligand interaction energy. $\Delta E_P$ and $\Delta E_L$ can be computed as the difference of internal energies recorded in the solvated end states


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

$$\Delta E_x = \langle E_x \rangle_{PL} - \langle E_x \rangle_x \quad (3)$$

where $x$ represents either the protein P, or the ligand L. The angular braces indicate ensemble averages over the bound (PL) and the unbound (P or L) states. However, in this study, we simplify this calculation by having strong restraints on protein heavy atoms and by selecting ligand molecules that involve no rotatable bonds linking heavy atoms. This causes the intramolecular conformational distributions of P and L to be very similar in the complexed and free states, making $\Delta E_P = \Delta E_L \approx 0$. The approximation assumes that energetic contributions due to hydrogen atom positions largely cancel between the end states. Thus, the binding energy to a good approximation becomes

$$\Delta E_b \approx E_{PL} + \Delta\Delta E_b^{solv} \quad (4)$$

$\Delta\Delta E_b^{solv}$ is the change in the solvation energy upon binding, which is the difference between the solvation energy of PL and that of P and L.

$$\Delta\Delta E_b^{solv} = \Delta E_{PL}^{solv} - \Delta E_P^{solv} - \Delta E_L^{solv} \quad (5)$$

The calculation of these molecular solvation energies is described in the Solvation Energy subsection below.

**2.2. Binding Entropy.** The binding entropy, $\Delta S_b$, is defined as the difference between the entropies of the complexed and free states.

$$\Delta S_b = S_{PL} - S_P - S_L \quad (6)$$

Here the terms on the right-hand side include the entropic contributions from water. The binding entropy can be exactly re-expressed as a sum of excess solute and solvent entropies

$$\Delta S_b = (S_{PL}^* - S_P^* - S_L^*) + \Delta\Delta S_b^{solv} \quad (7)$$

where the asterisk indicates entropy purely due to the solute degrees of freedom. Due to the strong restraints on the protein simulated in the bound and free states (see below), and the relatively rigid nature of the fragment-like ligands, the solute entropy can be approximated as the rigid body entropy loss of the ligand upon binding $\Delta S_L^*$, with the reference state being the ligand in 1 M aqueous solution.

$$\begin{aligned}(S_{PL}^* - S_P^* - S_L^*) &\approx \Delta S_L^* \\ &= \Delta S_L^{trans} + \Delta S_L^{rot}\end{aligned} \quad (8)$$

In practice, $\Delta S_L^*$ is computed by analyzing the positions and orientations of the ligand obtained from the simulation of the complexed state and computing the 1-body Shannon's entropy as is done for water molecules described below. The approximation assumes that entropic contributions due to hydrogen atom positions are nearly equal between the end states.

The solvation entropy $\Delta\Delta S_b^{solv}$ is given by

$$\begin{aligned}\Delta\Delta S_b^{solv} &= \Delta S_{PL}^{solv} - \Delta S_P^{solv} - \Delta S_L^{solv} \\ &= \int \Delta S_{PL}^{solv}(\mathbf{r_P}, \mathbf{r_L}) p(\mathbf{r_P}, \mathbf{r_L})\, d\mathbf{r_P} d\mathbf{r_L} \\ &\quad - \int \Delta S_P^{solv}(\mathbf{r_P}) p(\mathbf{r_P})\, d\mathbf{r_P} - \int \Delta S_L^{solv}(\mathbf{r_L}) p(\mathbf{r_L})\, d\mathbf{r_L}\end{aligned} \quad (9)$$

where $\Delta S_x^{solv}(\mathbf{r_x})$ indicates the solvation entropy of species $x$ in conformation $\mathbf{r_x}$, and $p(\mathbf{r_x})$ is the equilibrium probability density of that conformation. In other words, the molecular solvation entropy is expressed as a weighted average of the solvation entropies of the different conformations. The restrained protein and nearly rigid ligands cause $\Delta S_{PL}^{solv}$ and $\Delta S_L^{solv}$ to be independent of the protein and ligand coordinates $\mathbf{r_P}$ and $\mathbf{r_L}$, respectively

$$\int \Delta S_P^{solv}(\mathbf{r_P}) p(\mathbf{r_P})\, d\mathbf{r_P} \approx \Delta S_P^{solv}|_{\mathbf{r}^*_P} \quad (10)$$

$$\int \Delta S_L^{solv}(\mathbf{r_L}) p(\mathbf{r_L}) d\mathbf{r_L} \approx \Delta S_L^{solv}|_{\mathbf{r}^*_L} \quad (11)$$

where $\mathbf{r}_P^*$ and $\mathbf{r}_L^*$ represent the relatively rigid equilibrium geometries of the protein and ligand, respectively. In the remainder of this work, for simplicity we refer to these quantities as $\Delta S_P^{solv}$ and $\Delta S_L^{solv}$, omitting the subscripts indicating the restrained geometries. In the complexed state, the ligand is expected to assume a narrow range of conformations, which is also due to the use of a restrained protein in this study. This logic prompted us to adopt an approximation which allows to decouple the calculation of $\Delta S_{PL}^{solv}$ from $\mathbf{r_L}$.

$$\int \Delta S_{PL}^{solv}(\mathbf{r_P}, \mathbf{r_L}) p(\mathbf{r_P}, \mathbf{r_L}) d\mathbf{r_P} d\mathbf{r_L} \approx \Delta S_{PL}^{solv}|_{\mathbf{r}^*_P, \{\mathbf{r_L}\}} \quad (12)$$

Here $\{\mathbf{r_L}\}$ represents the family of ligand conformations in the complexed state. The approximation in the equation mainly arises due to the assumption that solvation free energy will not be significantly affected by the mobility of the ligand in the binding pocket. In principle, the accuracy of this calculation can be improved by performing a clustering analysis of the ligand bound states and computing a weighted average of $\Delta S_{PL}^{solv}(\mathbf{r_P}, \mathbf{r_L})$ over the finite ensemble of predominant complex states $\{\mathbf{r_P}, \mathbf{r_L}\}$. Such a treatment would also allow inclusion of protein flexibility. However, we do not undertake such an analysis here and reserve it for a future study. In the remainder of this work, for simplicity, the solvation entropy of the complexed state is referred to as $\Delta S_{PL}^{solv}$, omitting the subscript indicating the conformation(s).

Therefore, $\Delta S_b$ can be rewritten as

$$\Delta S_b = \Delta S_L^* + \Delta\Delta S_b^{solv} \quad (13)$$

The calculations of the molecular solvation entropies $\Delta S_{PL}^{solv}$, $\Delta S_P^{solv}$, and $\Delta S_L^{solv}$ are described in the Solvation Entropy subsection below.

**2.3. Solvation Energy.** The solvation energy $\Delta E_x^{solv}$, of a solute $x$ per the presented approach is calculated as the sum of the interaction energy of the solute with water and the energy change in bulk water caused due to the introduction of the solute. As detailed above, these calculations need to be performed for the free and the complexed states ($x$ can be L, P, or PL). The perturbation to bulk water due to the solute is expected to vanish at long distances, allowing the calculation of excess energy to be limited to a region local to the solute. Here, the calculation region $R$ is a sphere of radius 8 Å centered on the solute or the binding pocket. To facilitate regional analyses, our calculation follows the Grid Inhomogeneous Solvation Theory (GIST) approach.[13] The excess energy is integrated over a 3D grid spanning the system composed of cubic volume elements $v$ of edge length 0.5 Å

$$\begin{aligned}\Delta E_x^{solv} &= \rho_o \int_R g(\mathbf{r}) \left[ E_{sw}(\mathbf{r}) + \frac{1}{2} E_{ww}(\mathbf{r}) - E_{bulk} \right] d\mathbf{r} \\ &= \rho_o V_{vox} \sum_{v \in R} g(v) \left[ E_{sw}(v) + \frac{1}{2} E_{ww}(v) - E_{bulk} \right]\end{aligned} \quad (14)$$

where $\rho_o$ is the bulk density of water, $g(\mathbf{r})$ is the relative density of water ($g(\mathbf{r}) = ((\rho(\mathbf{r}))/(\rho_0))$), where $\rho(\mathbf{r})$ is the density of water at position $r$ within the region $R$. $E_{sw}(\mathbf{r})$ and $E_{ww}(\mathbf{r})$ are the ensemble averaged solute−water and water−water interaction energies as a function of position. The second equality discretizes the energy calculation by introducing per-voxel energies $E_{sw}(v)$, and $E_{ww}(v)$. The subtraction of the energy of water in bulk, $E_{bulk}$, results in the excess energy. $V_{vox}$ is the volume of a voxel element.

It is straightforward to decompose the solvation energy into solute−water and water−water terms.

$$E_{sw,x} = \rho_o V_{vox} \sum_{v \in R} g(v) E_{sw}(v) \quad (15)$$

This expression can be used to calculate various energetic terms: ligand-water energy in the L or PL systems as $E_{lw,L}$, $E_{lw,PL}$; protein−water energy in the P or PL systems as $E_{pw,P}$, $E_{pw,PL}$.

The excess water−water energy is given by

$$\Delta E_{ww,x} = \rho_o V_{vox} \sum_{v \in R} g(v) \left[ \frac{1}{2} E_{ww}(v) - E_{bulk} \right] \quad (16)$$

In all summations, only the voxels which have their center within 8 Å of the center of $R$ are included. This approach is followed to


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

calculate the excess energies from the explicit solvent L, P and PL simulations. Solute-water and water−water nonbonded interaction energies are calculated using periodic boundary conditions without a cutoff. Energy evaluations extend to one unit cell in each direction to account for long-range interactions.

The calculation of per-voxel contribution to the solvation energy is useful for spatial analysis and is computed using the following.

$$\Delta E_x^{\mathrm{solv}}(\nu) = \rho_o V_{\mathrm{vox}} g(\nu)\left[E_{\mathrm{sw}}(\nu) + \frac{1}{2}E_{\mathrm{ww}}(\nu) - E_{\mathrm{bulk}}\right] \quad (17)$$

The per-voxel contributions are visualized using the Visual Molecular Dynamics (VMD) package.[14]

**2.4. Solvation Entropy.** The solvation entropies of the P, L, or the PL systems, $\Delta S_x^{\mathrm{solv}}$, are also calculated using the GIST approach with some modifications as described below. The solvation entropy can be expressed as a sum of the entropy due to solute−water correlations, $\Delta S_{\mathrm{sw}}^{\mathrm{solv}}$, which is a 1-water term and water−water correlations $\Delta S_{\mathrm{ww}}^{\mathrm{solv}}$, involving pairs of water molecules.

$$\Delta S_x^{\mathrm{solv}} = \Delta S_{\mathrm{sw}}^{\mathrm{solv}} + \Delta S_{\mathrm{ww}}^{\mathrm{solv}} \quad (18)$$

$\Delta S_{\mathrm{ww}}^{\mathrm{solv}}$ is the most difficult contribution to converge with limited simulation data, which prompted us to employ approximations to estimate it. Huggins observed recently[15] that using a scaled value of $\Delta S_{\mathrm{sw}}^{\mathrm{solv}}$ for an estimate of $\Delta S_{\mathrm{ww}}^{\mathrm{solv}}$ resulted in good reproduction of experimental hydration free energies of 20 solutes of diverse chemical nature.

$$\Delta S_{\mathrm{ww}}^{\mathrm{solv}} = c\Delta S_{\mathrm{sw}}^{\mathrm{solv}} \quad (19)$$

A scaling factor of $c = -0.5$ was used in the above cited work, which is motivated by a similar scaling observed in the analogous energetic terms. The use of a negative factor assumes that increased solute−water correlations are offset by decreased water−water correlations. We adopted this approximation in our work. However, our initial tests with reproducing the hydration free energies of a set of small molecules showed a factor of −0.3 to result in better correlation with experimental hydration free energy data (see Results section).

The solute−water entropy $\Delta S_{\mathrm{sw}}^{\mathrm{solv}}$ can be exactly expressed as a sum of translational and rotational entropy.

$$\Delta S_{\mathrm{sw}}^{\mathrm{solv}} = \Delta S_{\mathrm{sw,tr}}^{\mathrm{solv}} + \Delta S_{\mathrm{sw,rot}}^{\mathrm{solv}} \quad (20)$$

The excess translational entropy is calculated as

$$\begin{aligned}\Delta S_{\mathrm{sw,tr}}^{\mathrm{solv}} &= -k_B\rho_o \int_R g(\mathbf{r})\ln[g(\mathbf{r})]\,\mathrm{d}\mathbf{r} \\ &= -k_B\rho_o V_{\mathrm{vox}} \sum_{\nu\in R} g(\nu)\ln[g(\nu)]\end{aligned} \quad (21)$$

where $g(\nu)$ and $g(\mathbf{r})$ are as defined above. Thus, in practice the translational component of the entropy can be computed by calculating the water density obtained from the simulation frames. Rotational entropy is calculated as

$$\begin{aligned}\Delta S_{\mathrm{sw,rot}}^{\mathrm{solv}} &= \rho_o \int_R g(\mathbf{r})S^{\mathrm{rot}}(\mathbf{r})\,\mathrm{d}\mathbf{r} \\ &= \rho_o V_{\mathrm{vox}} \sum_{\nu\in R} g(\nu)S^{\mathrm{rot}}(\nu)\end{aligned} \quad (22)$$

where $S^{\mathrm{rot}}(\mathbf{r})$ and $S^{\mathrm{rot}}(\nu)$ are the rotational entropies associated with position $\mathbf{r}$ and voxel $\nu$ within the region $R$, respectively. $S^{\mathrm{rot}}(\nu)$ is computed using a nearest neighbor estimator[13]

$$\begin{aligned}\int_\nu S^{\mathrm{rot}}(\mathbf{r})\mathbf{dr} &\approx S^{\mathrm{rot}}(\nu) \\ &= -k_B[-\gamma + \frac{1}{N_\nu}\sum_{i=1}^{N_\nu}\ln\frac{6\pi}{N_\nu(\Delta\omega_i)^3}]\end{aligned} \quad (23)$$

where $\gamma$ is Euler's constant. $N_\nu$ is the total number of different water molecules in the voxel $\nu$, which was limited to a maximum of 1000 to limit memory use. $\Delta\omega_i$ is the angular distance of a water molecule $i$ to its nearest neighbor geometry in the same voxel. Unlike our previous study,[16] we used the norm of the distance between quaternions to evaluate $\Delta\omega$. This metric was recently shown to possess superior convergence properties than the Eucledian distance between the Euler angles.[17] The first two sections in the Supporting Information describe the formulas used in the conversion of Euler angles to quaternions, and the convergence properties evaluated using randomly generated data. Thus, the total rotational entropy is obtained as an average over voxels in $R$ weighted by the voxel relative water density $g(\nu)$.

For spatial analysis the per-voxel contribution to $\Delta S_x^{\mathrm{solv}}$ is computed as

$$\Delta S_x^{\mathrm{solv}}(\nu) = -(1+c)k_B\rho_o V_{\mathrm{vox}} g(\nu)\left[\ln[g(\nu)] - \frac{1}{k_B}S^{\mathrm{rot}}(\nu)\right] \quad (24)$$

where the prefactor $(1 + c)$ is introduced to add the empirical estimate of the water−water entropy and the other terms are as defined above.

**2.5. Protein−Ligand Interaction Energy.** The protein−ligand interaction energy $E_{\mathrm{pl}}$ is calculated from the simulation snapshots. Similar to the grid based approach adopted for the calculation of solute−water and water−water energies, the protein−ligand energies are also integrated over the grid spanning the binding site. However, the spatial decomposition of this contribution is not presented in this study as the binding site is a relatively small region.

**2.6. Ligand Entropy.** The ligand entropy calculation involves translational and rotational components spanning six degrees of freedom as for a single water molecule. Equations 21−23 are thus employed with appropriate modifications for the calculation of ligand entropy, $\Delta S_L^*$. The bulk concentration of the ligand used is the standard 1 M concentration, equivalent to $\rho_0 = (1/1660)$ Å$^{-3}$. For propane, the positions of the three carbon atoms were used to calculate the Euler angles. For methanol, the atoms used were the carbon, oxygen, and the alcohol hydrogen. Symmetry corrections were not applied during the entropy calculation.

**2.7. Binding Free Energy and Its Components.** The subsections above gave expressions for total binding energy and entropy and solvation energy and entropy. However, to get mechanistic insights into binding it is of interest to decompose these terms into contributions arising from the different degrees of freedom, which is given as follows.

$$\begin{aligned}\Delta G_b^o &= \Delta E_b - T\Delta S_b \\ &= \Delta\Delta E_{\mathrm{ww}} + \Delta E_{\mathrm{pw}} + \Delta E_{\mathrm{lw}} + E_{\mathrm{pl}} - T\Delta\Delta S_b^{\mathrm{solv}} - T\Delta S^*{}_L\end{aligned} \quad (25)$$

Here, $\Delta\Delta E_{\mathrm{ww}}$ is the water reorganization energy, and is obtained from the analysis of P, L, and PL states

$$\Delta\Delta E_{\mathrm{ww}} = \Delta E_{\mathrm{ww,PL}} - \Delta E_{\mathrm{ww,P}} - \Delta E_{\mathrm{ww,L}} \quad (26)$$

The change in protein−water interaction energy is given as

$$\Delta E_{\mathrm{pw}} = E_{\mathrm{pw,PL}} - E_{\mathrm{pw,P}} \quad (27)$$

The change in ligand−water interaction energy is given as

$$\Delta E_{\mathrm{lw}} = E_{\mathrm{lw,PL}} - E_{\mathrm{lw,L}} \quad (28)$$

**2.8. System Setup and Molecular Dynamics Simulations.** The crystal structures of Factor Xa (PDB ID: 1FJS[18]) and P38 MAP kinase (PDB ID: 1OUY[19]) in complex with ligands were downloaded from the Protein Data Bank (PDB).[20,21] To maintain a charge neutral system, for both proteins, solvent exposed charged residues were neutralized. This was done instead of the standard practice of adding neutralizing counterions in order to facilitate convergence of the calculations. For Factor Xa (FXa), Lys62, Lys204, and Lys236 were deprotonated. To neutralize the P38 MAP kinase (P38MK) system, Glu12, Glu22, Glu97, Glu160, Asp145, Asp161, Asp177, and Asp315 were protonated, an approach previously used for free energy perturbation and linear interaction energy studies of ligand−protein interactions.[22,23] The ligands were removed from the crystal structure,


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

Table 1. Thermodynamic Properties of Ligand Solvation[a]

| molecule | $n_{wat}$ | $E_{lw}$ | $\Delta E_{ww}$ | $-T\Delta S_{lw}$ | $-T\Delta S_{ww}$ | $\Delta G^{solv}_{calc}$ | $\Delta G^{solv}_{FEP}$ | $\Delta G^{solv}_{exp}$ |
|---|---|---|---|---|---|---|---|---|
| propane | 69.47 | −8.08 | 5.83 | 6.22 | −1.87 | 2.10 | 1.60 | 1.96 |
| methanol | 71.24 | −18.72 | 9.08 | 4.72 | −1.42 | −6.34 | −6.16 | −5.10 |

[a] $n_{wat}$ is the number of water molecules recorded in the analysis region of the 8 Å radius sphere. Units are kcal/mol.

while maintaining crystallographic water molecules. The Reduce software[24] was used to assign His protonation states and adjust the side-chain conformations of Asn and Gln residues. Following a short restrained minimization in vacuo with the steepest descent method to remove bad contacts, the proteins were aligned on the basis of their principal axes and immersed in rectangular water boxes with dimensions 16 Å longer than the distance between protein extrema in each direction. For simulating the free ligand systems, all ligands were placed in cubic water boxes of side 30 Å. MD simulations were performed using the GROMACS package v5.0.0.[25] The CHARMM36 force field[26−28] was used for the protein and the CHARMM General Force Field (CGenFF)[29] for the ligands. The TIP3P model[30] modified for the CHARMM force field[31] was used to represent water. All simulations were performed under periodic boundary conditions. van der Waals (vdW) interactions were switched off smoothly in the range 8−10 Å, and the particle mesh Ewald method was used to treat long-range electrostatics with a real space cut off of 10 Å. Long-range dispersion correction to the energy and pressure was applied. The LINCS algorithm was used to constrain all bonds involving a hydrogen atom. All solvated systems were energy minimized for 500 steps using the steepest descent method under periodic boundary conditions. The systems were then equilibrated in the NVT ensemble at 298 K using the velocity rescaling method over 100 ps. Following equilibration, production simulations were performed in the NPT ensemble, where the temperature of 298 K was maintained using the Nose−Hoover thermostat and pressure of 1 atm maintained using the Parrinello−Rahman barostat. The first 1 ns of production in each series of simulations was not considered for analysis. In all systems simulated, all protein non-hydrogen atoms were restrained with a force constant of 24 kcal/mol Å$^{-2}$. The same restraints were applied to ligand heavy atoms in the free ligand simulations, but not in the complexed state simulations. In the latter case, the ligand was restrained to the binding site center by a flat bottom spherical boundary potential, with a radius of 4 Å.

**2.9. Identification of Ligand Binding Sites on the Proteins.** The study of binding thermodynamics undertaken here required an initial step of identifying valid binding sites on the protein surfaces. The approach adopted here closely follows our previous work in the context of the Site Identification by Ligand Competitive Saturation (SILCS) method,[32,33] modified to map the binding sites of single solutes as opposed to multiple solutes. Another modification necessary to keep the subsequent calculations consistent with binding site detection was the use of the same protein restraints as used in all other simulations. The proteins were immersed in an aqueous solution of 1 M propane or methanol. Ten independent 40 ns long MD simulations separately for propane and methanol were performed involving the binding and unbinding of ligands with the protein. In simulations involving propane, an artificial repulsive term was added to prevent propane−propane aggregation in the simulation box.[32] The simulation snapshots output every 10 ps were analyzed to build 3D histograms with resolution 0.5 Å of the center of mass (CoM) coordinates of the ligand. In order to identify binding sites from the histograms, we applied a clustering algorithm described previously.[34] Briefly, the clustering algorithm begins by assigning an arbitrarily chosen grid center point to the first cluster, and thereafter, each grid element is either assigned to an existing cluster if its center is located closer than 5 Å to any cluster or else to a newly created cluster. After the inclusion of each element in a cluster, the cluster centers are recomputed as the mean of the coordinates of the members. Following the initial assignment, an iterative loop is run, which would redo the cluster assignment on the basis of the distance from the existing cluster centers. The iteration is terminated once no more updates of the cluster assignment occur. Only voxels with significant occupancy were selected for clustering in order to restrict the number of favorable binding sites to be less than 20. A cutoff of 80× bulk density was used except for propane histogram maps in P38MK, where it was 200×.

## 3. RESULTS

Our goal is to provide a thorough thermodynamic characterization of the binding of fragment-sized ligands to diverse protein pockets. This is achieved by first generating the configurational ensemble of the solvated systems in the end states of the thermodynamic transition, and second, using it to calculate excess energy and entropy. As shown schematically in Figure 1, three series of MD simulations are performed that involve the free ligand (L), the free protein (P), and the protein−ligand complex (PL). The first calculation (L) results in the hydration thermodynamics and free energy of the ligand. The second calculation results in the local hydration thermodynamics of the free protein pocket (P). The third system calculates the thermodynamics of the complexed PL state, which includes contributions from hydration and the ligand with respect to an ideal gas state of the ligand. As we describe below, detailed insight into binding thermodynamics is obtained by calculating the differences in the energetic and entropic terms obtained from these three ensembles and associated calculations.

**3.1. Bulk Water Properties.** The excess solvation energy and entropy calculated in this study are referenced to bulk water. Therefore, it is important to obtain properties of bulk water accurately using the same force field and simulation parameters as used in the remainder of this study. MD simulations of cubic water boxes with edge length 30 Å were performed per the protocol described above to obtain 10 trajectories, each 20 ns long, with frames output every 400 fs for analysis. Excess energy and entropy were calculated for a spherical region of radius 8 Å centered in the simulation box. The entire simulation data (200 ns) was used to calculate the average energy of a water molecule in bulk $E_{bulk}$ yielding −9.821 96 kcal/mol. The bulk density $\rho_0$ from the same simulation data was calculated to be 0.033 97 Å$^{-3}$. To test convergence properties, only the first 120 000 frames were used from the simulations. Figure S1a in the Supporting Information shows the convergence of excess energy $\Delta E_{ww}$, the translational entropy $T\Delta S^{solv}_{sw,tr}$, and rotational entropy $T\Delta S^{solv}_{sw,rot}$ which attain values sufficiently close to zero. Deviation from zero for these quantities is less than $RT$ with 20 000 frames and reaches less than $(1/2)RT$ by 50 000 frames. On the basis of this test, we conclude that our approach is capable of yielding reasonably converged thermodynamic properties for spherical volumes of radius 8 Å.

**3.2. Solvation Thermodynamics of Ligands.** Using the same protocol as for the pure water simulation described above, each ligand molecule was placed at the center of a cubic water box with edge length 30 Å. Ten independent MD simulations, each 5 ns long, were performed, during which the heavy atoms of the ligand were strongly restrained (see the Theory and Methods section). Thermodynamic properties were calculated


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

for the ligand and the surrounding water contained in the 8 Å radius sphere centered in the box. Snapshots from the simulations were analyzed to calculate the solvation energy $\Delta E_{\mathrm{L}}^{\mathrm{solv}}$, which is composed of ligand−water energy $E_{\mathrm{lw}}$, and water reorganization energy $\Delta E_{\mathrm{ww}}$. The first order excess entropy of water, $\Delta S_{\mathrm{sw}}^{\mathrm{solv}}$, was calculated using eq 20. Table 1 shows the energy terms and $\Delta S_{\mathrm{sw}}^{\mathrm{solv}}$ along with the calculated solvation free energy $\Delta G_{\mathrm{calc}}^{\mathrm{solv}}$ for propane and methanol. As detailed in the Theory and Methods section, we adopted a simple approximation to estimate the water−water entropy $\Delta S_{\mathrm{ww}}^{\mathrm{solv}}$ by scaling the solute-water entropy by a constant $c$ (eq 19). Previous work[15] has used a value of $c = -0.5$. However, our work differs by its use of discretization that is involved in GIST. It has been noted that the discretization adopted during the calculation of histogram based $\Delta S_{\mathrm{sw,tr}}^{\mathrm{solv}}$ has a tendency to underestimate this quantity. Thus, it is expected that the value of $c$ would be sensitive to implementation details including the histogram bin size, and other parameters, such as the snapshot output frequency. A value of $c = -0.3$ resulted in the best reproduction of experimental hydration free energies. For methanol, the deviation from experimental value is >1 kcal/mol. However, we note that the agreement is much better when we compare our values with those computed using free energy perturbation (FEP) using the same force field,[35] also shown in Table 1. In addition, we performed the simulations and analysis for four other ligands (benzene, methane, acetamide, formamide) spanning a hydration free energy range of more than 10 kcal/mol, to validate our choice of $c$. Table S1 in the Supporting Information shows good reproduction of experimental/FEP data across the data set and significant improvement in the prediction of $\Delta G_{\mathrm{exp}}^{\mathrm{solv}}$ when compared to $c = -0.5$. Across the 4 ligands for which FEP data were computed, the average absolute unsigned error is 0.46 kcal/mol. Supporting Information Figure S1b−g shows good convergence of the excess energy $\Delta E_{\mathrm{L}}^{\mathrm{solv}}$, and excess solvent entropy $\Delta S_{\mathrm{L}}^{\mathrm{solv}}$ for all the systems investigated. The well-reproduced experimental or computational $\Delta G_{\mathrm{L}}^{\mathrm{solv}}$ shows that our approach is capable of capturing solvation thermodynamics within the accuracy of the force field used. Therefore, the same protocol, including a value of $c = -0.3$ was adopted for calculations in the P and PL states described below.

**3.3. Calculation of Binding Thermodynamics.** As a first step toward the investigation of binding thermodynamics, MD simulations of the proteins FXa and P38MK were performed in 1 M solutions of propane or methanol, as described in Theory and Methods section. A cumulative sampling time of 400 ns for each of the four systems identified protein sites to which the ligands bind spontaneously. Simulation trajectories were analyzed to build discretized 3D probability maps of ligand position, which were clustered as described above to identify centers of the binding pockets. Figure 2 shows the surfaces of the proteins FXa and P38MK, overlaid with the 3D ligand occupancy maps. The identified cluster centers are shown as spheres of the green and red colors for propane and methanol, respectively. Following the detection of binding sites, they were ranked on the basis of the highest occupancy voxel of the cluster, and the top ten propane and methanol sites for both proteins were chosen for further calculations. Figure 2 also shows the resultant binding affinity computed for each pocket using the end-point method.

To generate the configurational ensemble for the free proteins P, we obtained 5 MD trajectories, each 5 ns long of the solvated proteins. For the calculation of local thermodynamics of each binding pocket, the same 25 ns cumulative sampling data was used. In the case of the complexed states, PL, separate simulations were needed to generate the configurational ensembles for each binding site. For each pocket the protein−solvent system used in the P state was retained, and the ligand was placed at the pocket center. There were 5 × 5 ns MD trajectories generated, with the ligand confined within 4 Å of the binding pocket center using a spherical flat bottom restraint. During all simulations, snapshots were output every 400 fs, resulting in 62 500 frames for analysis. In the complexed state, a spherical volume of 2 Å radius centered at the pocket center was designated as the binding site. Thus, the ligand was considered unbound when its CoM was more than 2 Å away from the pocket center, and such conformations were not considered a part of the PL ensemble. The spherical flat bottom potential was applied with a larger radius than that of the designated binding site so as to allow the ligand and water to rearrange in the binding site. The section in the Supporting Information titled "Effect of Binding Site Size" and Supporting Information Figure S2 show that the results do not change significantly by reducing or increasing the binding site radius to 1 or 3 Å, respectively. This also serves to validate the approximation used in eq 12, as it shows weak dependence of $\Delta G_{\mathrm{b}}^{\circ}$ with the ligand position $r_{\mathrm{L}}$. For FXa and P38MK, 10 binding sites each of propane and methanol were investigated, resulting in a total of 40 pockets. Figure S3 in the Supporting Information shows the excellent convergence properties of computed thermodynamic quantities from the PL states of 12 pockets as representative examples. Calculations for other pockets and from the P states showed similarly-well converged properties (data not shown).

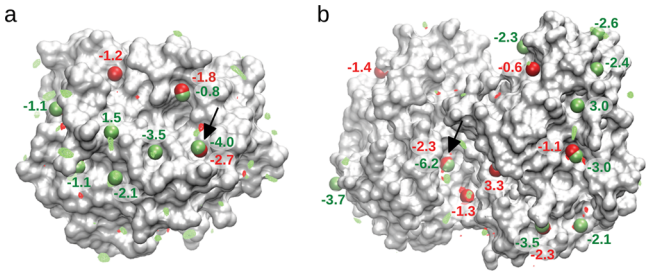



**Figure 2.** Proteins Factor Xa (FXa) (a) and P38 MAP kinase (P38MK) (b) shown in surface representation. Overlaid are the 3D ligand occupancy maps of propane (methanol) center of mass (CoM) coordinates as green (red) wireframes at 40× (100×) bulk density for FXa (P38MK). Methanol densities are shown at a cutoff of 40× bulk density for both FXa and P38MK. The top 10 centers of affinity for each ligand identified by clustering the occupancy maps are depicted as spheres with the same color as the maps. Arrows point to the known ligand binding pockets FP2/FM5, and PP1/PM4 for panels a and b, respectively, which are discussed in the text. Also indicated are the binding free energies of each pocket calculated by the end-point method.

*3.3.1. Propane Binding.* Table 2 displays the thermodynamic contributions of different molecular degrees of freedom to the binding of propane to pockets in FXa and P38MK. The table also reports averages and standard deviations computed over all pockets. The difference in the number of waters in the P and PL states $\Delta n_{\mathrm{wat}}$ shows that propane binding displaces 3−4 water molecules from the binding pockets. The water reorganization energy $\Delta\Delta E_{\mathrm{ww}}$ and direct protein−ligand interaction energy $E_{\mathrm{pl}}$ favor binding. On the other hand, the change in protein−water interaction energy $\Delta E_{\mathrm{pw}}$ and the


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

**Table 2. Thermodynamic Contributions of the Different Degrees of Freedom Implicated in the Binding of Propane to Pockets in Factor Xa (F) and P38 MAP Kinase (P)[a]**

| site | $\Delta n_{wat}$ | $\Delta\Delta E_{ww}$ | $\Delta E_{pw}$ | $\Delta E_{lw}$ | $E_{pl}$ | $-T\Delta\Delta S_b^{solv}$ | $-T\Delta S_L^*$ | $\Delta E_b$ | $-T\Delta S_b$ | $\Delta G_b^o$ |
|---|---|---|---|---|---|---|---|---|---|---|
| FP1 | −3.10 | −9.38 | 8.67 | 4.62 | −6.80 | −5.01 | 4.39 | −2.88 | −0.62 | −3.51 |
| FP2 | −3.78 | −12.57 | 16.35 | 4.77 | −8.89 | −8.19 | 4.49 | −0.34 | −3.70 | −4.04 |
| FP3 | −3.96 | −11.21 | 15.83 | 4.70 | −7.56 | −7.16 | 4.27 | 1.76 | −2.89 | −1.13 |
| FP4 | −3.92 | −8.15 | 8.79 | 3.61 | −5.59 | −4.53 | 3.81 | −1.34 | −0.72 | −2.07 |
| FP5 | −4.34 | −8.10 | 12.06 | 3.64 | −5.15 | −5.99 | 3.77 | 2.44 | −2.23 | 0.21 |
| FP6 | −3.35 | −12.00 | 15.83 | 3.94 | −7.04 | −6.28 | 4.95 | 0.73 | −1.33 | −0.60 |
| FP7 | −3.82 | −2.82 | 4.89 | 3.56 | −5.52 | −4.40 | 3.89 | 0.12 | −0.51 | −0.40 |
| FP8 | −3.81 | −7.42 | 10.54 | 3.84 | −6.24 | −5.59 | 3.78 | 0.72 | −1.80 | −1.08 |
| FP9 | −4.29 | −7.19 | 12.75 | 2.66 | −4.65 | −5.69 | 3.67 | 3.56 | −2.03 | 1.54 |
| FP10 | −3.33 | −10.95 | 13.35 | 3.60 | −6.87 | −4.39 | 4.50 | −0.87 | 0.11 | −0.76 |
| PP1 | −3.23 | −14.80 | 14.24 | 6.48 | −10.04 | −7.62 | 5.51 | −4.13 | −2.12 | −6.24 |
| PP2 | −2.78 | −15.48 | 15.44 | 5.47 | −9.08 | −5.44 | 5.55 | −3.65 | 0.10 | −3.55 |
| PP3 | −4.37 | −5.77 | 10.99 | 4.47 | −7.26 | −7.34 | 4.43 | 2.43 | −2.91 | −0.48 |
| PP4 | −4.50 | −10.05 | 14.05 | 5.27 | −7.54 | −8.66 | 3.92 | 1.73 | −4.75 | −3.02 |
| PP5 | −3.67 | −6.10 | 7.36 | 3.63 | −6.04 | −5.24 | 4.09 | −1.15 | −1.15 | −2.29 |
| PP6 | −4.06 | −5.84 | 5.37 | 3.28 | −5.47 | −4.74 | 3.72 | −2.66 | −1.02 | −3.68 |
| PP7 | −3.12 | −14.29 | 18.54 | 5.58 | −9.54 | −8.74 | 6.01 | 0.28 | −2.72 | −2.44 |
| PP8 | −3.21 | −3.20 | 9.60 | 4.86 | −6.51 | −6.11 | 4.35 | 4.75 | −1.76 | 2.99 |
| PP9 | −4.18 | −4.46 | 7.31 | 4.13 | −6.76 | −6.44 | 4.13 | 0.22 | −2.31 | −2.10 |
| PP10 | −4.36 | −7.22 | 9.80 | 3.85 | −5.75 | −7.10 | 3.79 | 0.68 | −3.30 | −2.62 |
| av | −3.76 | −8.85 | 11.59 | 4.30 | −6.91 | −6.23 | 4.35 | 0.12 | −1.88 | −1.76 |
| std | 0.51 | 3.75 | 3.87 | 0.92 | 1.51 | 1.39 | 0.67 | 2.34 | 1.28 | 2.08 |

[a]Units are kcal/mol. The first letter (F or P) in the site nomenclature identifies the protein, and the second letter identifies the ligand; propane (P), methanol (M); e.g., FP1 denotes propane binding pocket 1 in Factor Xa. $\Delta n_{wat}$ is the number of water molecules displaced upon ligand binding.

**Table 3. Thermodynamic Contributions of the Different Degrees of Freedom Implicated in the Binding of Methanol to Pockets in Factor Xa (F) and P38 MAP Kinase (P)[a]**

| site | $\Delta n_{wat}$ | $\Delta\Delta E_{ww}$ | $\Delta E_{pw}$ | $\Delta E_{lw}$ | $E_{pl}$ | $-T\Delta\Delta S_b^{solv}$ | $-T\Delta S_L^*$ | $\Delta E_b$ | $-T\Delta S_b$ | $\Delta G_b^o$ |
|---|---|---|---|---|---|---|---|---|---|---|
| FM1 | −1.62 | −16.12 | 17.95 | 16.05 | −20.39 | −5.39 | 7.43 | −2.50 | 2.04 | −0.46 |
| FM2 | −1.66 | −10.73 | 8.51 | 8.63 | −9.08 | −4.64 | 5.99 | −2.67 | 1.35 | −1.31 |
| FM3 | −1.71 | −12.28 | 11.22 | 11.34 | −12.68 | −4.64 | 6.24 | −2.40 | 1.60 | −0.80 |
| FM4 | −1.80 | −11.55 | 11.26 | 10.45 | −10.25 | −5.17 | 5.10 | −0.10 | −0.07 | −0.17 |
| FM5 | −2.13 | −9.39 | 10.23 | 6.14 | −7.82 | −6.47 | 4.67 | −0.84 | −1.81 | −2.65 |
| FM6 | −2.02 | −11.92 | 13.33 | 12.45 | −14.02 | −5.25 | 6.45 | −0.16 | 1.20 | 1.04 |
| FM7 | −1.99 | −14.53 | 15.74 | 14.50 | −15.12 | −4.76 | 5.50 | 0.59 | 0.73 | 1.33 |
| FM8 | −1.68 | −8.67 | 8.02 | 9.35 | −10.95 | −5.05 | 6.13 | −2.26 | 1.08 | −1.17 |
| FM9 | −1.88 | −5.95 | 7.82 | 5.63 | −7.69 | −6.40 | 4.80 | −0.18 | −1.59 | −1.77 |
| FM10 | −1.95 | −12.93 | 12.80 | 11.58 | −13.35 | −5.12 | 5.72 | −1.89 | 0.60 | −1.30 |
| PM1 | −1.70 | −16.71 | 14.59 | 13.58 | −14.22 | −5.22 | 5.66 | −2.75 | 0.44 | −2.31 |
| PM2 | −1.95 | −9.81 | 10.25 | 9.80 | −11.01 | −5.32 | 4.75 | −0.78 | −0.58 | −1.35 |
| PM3 | −1.73 | −13.27 | 15.98 | 10.19 | −10.10 | −6.28 | 6.82 | 2.80 | 0.54 | 3.34 |
| PM4 | −1.62 | −10.88 | 8.18 | 9.33 | −8.17 | −5.60 | 4.88 | −1.54 | −0.72 | −2.26 |
| PM5 | −1.58 | −15.15 | 19.01 | 16.82 | −21.23 | −7.29 | 7.72 | −0.55 | 0.42 | −0.13 |
| PM6 | −1.89 | −10.35 | 10.48 | 11.26 | −12.95 | −5.20 | 5.39 | −1.55 | 0.19 | −1.36 |
| PM7 | −2.34 | −13.14 | 15.32 | 12.15 | −13.48 | −6.03 | 5.63 | 0.85 | −0.40 | 0.44 |
| PM8 | −2.09 | −5.42 | 5.41 | 5.71 | −6.46 | −5.22 | 5.41 | −0.76 | 0.19 | −0.57 |
| PM9 | −2.71 | −9.84 | 10.99 | 8.59 | −8.23 | −7.21 | 4.62 | 1.51 | −2.59 | −1.07 |
| PM10 | −2.01 | −10.64 | 13.60 | 8.80 | −10.72 | −5.51 | 6.12 | 1.04 | 0.61 | 1.65 |
| av | −1.90 | −11.46 | 12.03 | 10.62 | −11.90 | −5.59 | 5.75 | −0.71 | 0.16 | −0.55 |
| std | 0.28 | 2.98 | 3.64 | 3.11 | 3.93 | 0.78 | 0.88 | 1.53 | 1.17 | 1.49 |

[a]Units are kcal/mol. The first letter (F or P) in the site nomenclature identifies the protein, and the second letter identifies the ligand; propane (P), methanol (M); e.g., FM1 denotes methanol binding pocket 1 in Factor Xa. $\Delta n_{wat}$ is the number of water molecules displaced upon ligand binding.

ligand−water interaction energy $\Delta E_{lw}$ oppose binding. Although the magnitude of $\Delta\Delta E_{ww}$ is on average more favorable than $E_{pl}$ by 1.9 kcal/mol, the contributions are comparable, a finding which is different from the analysis by McCammon and co-workers using model ligands and cavities.[10] The entropy of water reorganization ($-T\Delta\Delta S_b^{solv}$) favors binding. This contribution spans a range of more than 4 kcal/mol (−8.74 to −4.39) and does not correlate with the number of water molecules displaced ($R^2$ = 0.03). Expectedly, the loss in translational and rotational entropy of the ligand



DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

upon binding ($-T\Delta S_t^*$) opposes binding. The computed standard binding free energies, $\Delta G_b^o$, of most sites are negative as expected, with the exception of 3 sites (see Discussion).

*3.3.2. Binding of Methanol and Comparison to Propane Binding.* Table 3 presents the magnitude of the different thermodynamic contributions to binding of methanol to the pockets in FXa and P38MK. The first notable difference between the binding of propane and methanol is that the former displaces on average 3.76 ± 0.51 water molecules upon binding, whereas the latter displaces only 1.90 ± 0.28. One possible reason for this could be the slightly larger size of propane causing it to bind to larger pockets, leading to the displacement of more number of water molecules. However, Figure 2 shows many pockets that bind both propane and methanol. Therefore, pocket geometry alone may not determine the number of water molecules displaced. The number of waters displaced by methanol ranges from 1.58 to 2.71, whereas the range for propane is 2.78−4.50, which indicates a larger number of water molecules displaced by the latter irrespective of the pocket considered. The more plausible reason could be the polar interaction of methanol with water allowing for favorable water−solute interactions leading to approximately an additional water remaining in the binding pocket. This is confirmed when we discuss binding of both ligands to pocket PP1 below.

Next, we compare differences in the energetic and entropic contributions to the binding of the two solutes. There is a significant difference in the change in ligand−water interactions between propane and methanol. In both cases $\Delta E_{lw}$ opposes binding; i.e., the ligand has more favorable interactions with water in the free state than in the complexed state. The polar methanol molecule typically loses more stabilizing interaction energy with water upon binding (10.62 ± 3.11 kcal/mol) than propane (4.30 ± 0.92 kcal/mol). This gain in energy is somewhat compensated by a more favorable protein−ligand interaction energy for methanol (−11.90 ± 3.93 kcal/mol) than propane (−6.91 ± 1.51 kcal/mol). The sum of these two terms ($\Delta E_{lw} + E_{pl}$) is more favorable for propane at −2.62 ± 0.77 kcal/mol, than methanol at −1.28 ± 1.37 kcal/mol. But the relative unfavorability of ligand associated interactions of methanol are compensated by more favorable water−water energy $\Delta\Delta E_{ww}$ and protein−water energy $\Delta E_{pw}$ resulting in a lack of significant total binding energy difference $\Delta E_b$ between methanol (−0.71 ± 1.53 kcal/mol) and propane (0.12 ± 2.34 kcal/mol). Among entropic contributions, propane binding leads to a slightly larger water entropy gain ($-T\Delta\Delta S_b^{solv}$ = −6.13 ± 1.36 kcal/mol) than methanol binding (−5.56 ± 0.71 kcal/mol). Again, this relatively small difference in entropy change is not reflected in the number of displaced waters, which is different by a factor of 2. Methanol loses more rigid body entropy upon binding than propane, on average by 1.32 kcal/mol. This can be rationalized by the fact that methanol, with its hydrogen bonding ability, can form more specific interactions with the protein and neighboring waters causing it to be more conformationally restrained. The ligand entropy and water entropy together lead to a more favorable entropy of propane binding (−1.88 ± 1.28 kcal/mol) than methanol binding (0.16 ± 1.17 kcal/mol). The average binding affinity of propane (−1.76 ± 2.08 kcal/mol) is more favorable than methanol (−0.55 ± 1.49 kcal/mol).

**3.4. Spatial Analysis of Binding Thermodynamics.** To obtain further mechanistic insight into the binding processes, we compute spatially resolved contributions of the thermodynamic quantities, which are enabled by the grid based formulation adopted in this work. To illustrate this ability and provide an overview of the method, we analyze the binding of propane and methanol to the binding pocket PP1/PM4 in P38MK, which binds both ligands. This pocket has the highest affinity for propane (−6.24 kcal/mol), and is shown by an arrow in Figure 2b. Pocket PP1 is located in the vicinity of Thr106, which directly interacts with a difluoro phenyl group of the ligand in the 1OUY cocrystal structure. Thus, the analysis of this pocket PP1/PM4 is relevant to pharmaceutically important protein−ligand binding. This pocket also has the second highest affinity for methanol (pocket PM4; $\Delta G_b^o$ = −2.26 kcal/mol). The centers of sites PP1 and PM4 nearly coincide with each other, even though they were obtained from different simulations, facilitating the use of the site for detailed analysis of both methanol and propane binding.

Spatial analysis of the water energy distribution in methanol binding is first presented for the PP1/PM4 site in Figure 3,

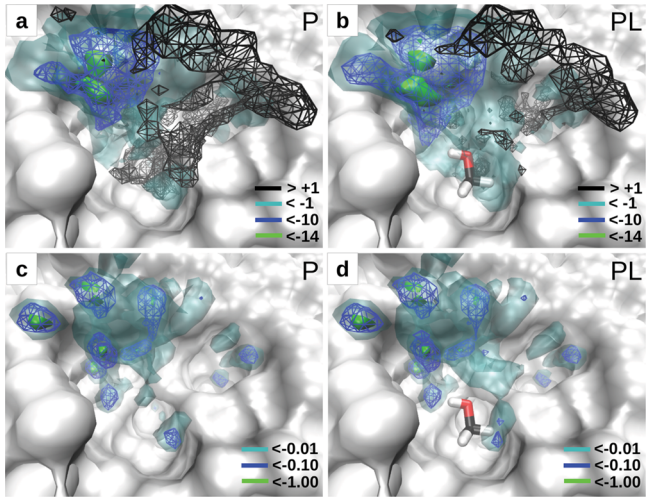



**Figure 3.** Visual analysis of the spatial distribution of per voxel un-normalized excess energy $u\Delta E_x^{solv}(\nu)$, and normalized excess energetic contribution $\Delta E_x^{solv}(\nu)$ for methanol binding to pocket PP1/PM4 in p38 MAP kinase. Panels a and b show $u\Delta E_x^{solv}(\nu)$ for states P and PL, respectively. Both unfavorable and favorable energy contours are drawn, with the cutoffs for each contour indicated at the bottom right corner of each panel. Panels c and d show $\Delta E_x^{solv}(\nu)$, which is normalized by local density for systems P and PL, respectively. Protein atoms blocking the view of the pocket were removed. Units of energy are kcal/mol.

followed by more detailed analysis which reveals intricate details of the binding mechanisms of both ligands to this pocket. In panels a and b, un-normalized voxel energy $u\Delta E_x^{solv}(\nu)$ is plotted at different contour cutoffs, which is given as

$$u\Delta E_x^{solv}(\nu) = E_{sw}(\nu) + \frac{1}{2}E_{ww}(\nu) - E_{bulk} \quad (29)$$

The expression for $u\Delta E_x^{solv}(\nu)$ differs from that of $\Delta E_x^{solv}(\nu)$ (eq 17) simply by the absence of the water density prefactor. The solute water energy $E_{sw}$ is computed as a sum of protein−water energy $E_{pw}$ and ligand−water energy $E_{lw}$, the latter of which is absent in the P state. Panel a shows many regions with favorable (<−1 kcal/mol; transparent cyan) and unfavorable excess energies (>1 kcal/mol; black wireframes). Panel b shows a representative snapshot of methanol to mark the location of


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

the binding site. By comparing panels a and b it is evident that, upon ligand binding, most of the unfavorable regions in the immediate vicinity of the ligand are converted to favorable regions. Additionally, in both panels blue wireframes and green solid contours show regions that have significantly more favorable average energy than bulk, which show subtle changes upon methanol binding. Panels c and d show $\Delta E_x^{solv}(\nu)$ (eq 17), which represents the actual contribution to hydration energy in each voxel as it is weighted by voxel water density. Notably, the unfavorable regions observed in panels a and b are absent in panels c and d because the corresponding voxels have very low water occupancies. A comparison of panels c and d reveals additional favorable water energetic contributions adjacent to the oxygen of methanol. While the solid green contours removed from the ligand highlight the most favorable water binding sites, these regions do not show significant differences between the P and the PL states. A more detailed analysis of the differential contributions between the two states which determine the driving forces behind binding is presented below. Figure 3 thus highlights the ability of the method to identify the spatial distribution of the contributions of individual terms to the thermodynamic properties. An integral over the entire region R irrespective of occupancy enables our approach to carry out a rigorous calculation of binding thermodynamics. Such spatial analysis can also be performed for solvent entropy and free energy, as shown below.

Figure 4 highlights the contributions of water probability, energy and entropy spatial distributions in the P and PL states, with the results for propane (panels a–c) stacked over those for methanol (panels d–f). In all panels, data from the P simulation is depicted in cyan, and that from PL simulation in red. The center of the binding pocket is demarcated by a small black sphere in all panels, except b, and e where a representative ligand conformation is shown instead. Panels a and d in Figure 4 show the binding pocket overlaid with water density contours at 2× bulk density from the P and PL simulations for propane and methanol, respectively. Panel a shows three regions of water density in P that are absent in the PL state. These represent regions where water molecules are displaced due to the binding of propane, and are in agreement with the nearly 3 water molecules displaced upon binding of propane (Table 2). One of the displacement sites, namely w0, coincides with the pocket center. The other two flanking sites namely w1 and w2, are indicated in panel a by arrows along with site w0. Examination of panel d shows that flanking sites w1 and w2 are not emptied completely upon binding of methanol.

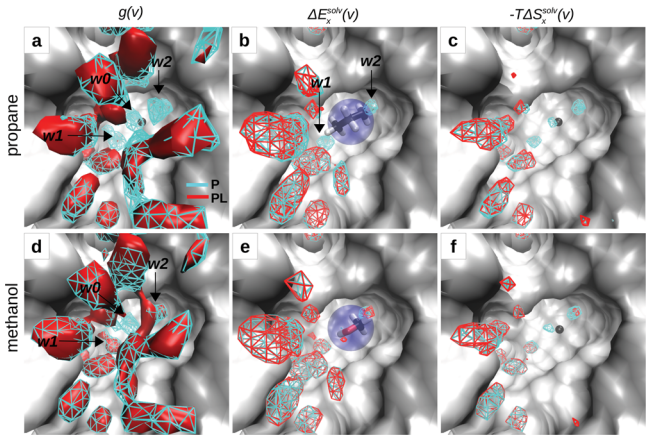



**Figure 4.** Visual analysis of the spatial distribution of the probability, energy and entropy of water in the binding of propane and methanol to pocket PP1/PM4 in p38 MAP kinase. Panels a–c show data for propane binding and d–f for methanol. Panels a,d show water density at 2× bulk as a function of voxel position $g(\nu)$ for the free protein (P, cyan) and the complex (PL, red surface). Panels b and e show excess solvation energy as a function of voxel $\Delta E_P^{solv}(\nu)$ for P (cyan), and $\Delta E_{PL}^{solv}(\nu)$ for PL (red) states at a cutoff of < −0.05 kcal/mol. Panels b and e also show a representative ligand conformation and a transparent blue sphere marking the extent of the binding site. Panels c and f show the excess entropic contribution $-T\Delta S_P^{solv}(\nu)$ for P (cyan), and $-T\Delta S_{PL}^{solv}(\nu)$ for PL (red) at a cutoff of >0.05 kcal/mol. The pocket center is demarcated by a small black sphere, and the proximal water sites are indicated by arrows. Contour legend is shown in panel a.

The spatial decomposition of the thermodynamic contributions was next analyzed. Since the ligand remains localized to the binding site, $E_{pl}$ and $\Delta S_L^*(\nu)$ are not included in the visual presentation. Equations 17 and 24 are used to calculate the per voxel contributions of the solvent excess energetic and entropic terms, respectively, for the states P and PL. Figure 4b,e shows the per voxel excess energy contours $\Delta E_P^{solv}(\nu)$ (cyan) and $\Delta E_{PL}^{solv}(\nu)$ (red) plotted at < −0.05 kcal/mol. In panels b and e the lack of favorable density in the P state at the center of the pocket, which corresponds to w0, indicates that water is not energetically favored at that position in either state. Second, the flanking water sites w1 and w2 prefer water energetically, and the displacement of these waters upon propane binding (panel b, PL state) is expected to result in an energetic penalty. However, this is not the case for methanol; energy contours of both the P and PL states in panel e show the presence of favorable energetic contributions for waters w1 and w2. Panels c and f show the per voxel excess entropy of the P ($-T\Delta S_P^{solv}(\nu)$; cyan) and PL ($-T\Delta S_{PL}^{solv}(\nu)$; red) states, contoured at >0.05 kcal/mol, such that the plots identify entropically unfavorable regions. For propane, all three displaced waters would contribute favorably to binding, as they are all displaced upon binding. On the other hand, for methanol only the central water site is emptied completely, and therefore the favorable entropic contribution to binding would be lesser, consistent with the $-T\Delta S_{PL}^{solv}(\nu)$ values of −7.62 kcal/mol for propane (Table 2, site PP1) and −5.60 kcal/mol for methanol (Table 3, site PM4), the difference being larger than the 1 kcal/mol difference in the solvation entropy of the ligands. These qualitative inferences drawn from panels a–f are further verified as detailed in the analyses below.

An informative way of visualizing the thermodynamic contributions is by creating "difference" maps, which display the per voxel differences of the excess energy and entropy values between the bound and unbound states.

$$\Delta\Delta A_{PL-P}^{solv}(\nu) = \Delta A_{PL}^{solv}(\nu) - \Delta A_{P}^{solv}(\nu) \quad (30)$$

Here $A$ indicates either energy $E$, entropy $S$, or the free energy $G$. In addition, to quantify the difference observations in overall terms, we computed a cumulative sum of the three thermodynamic quantities as a function of distance $R$ from the pocket center.

$$\Delta\Delta A_{PL-P}^{solv}(R) = \sum_{\nu \in R} \Delta\Delta A_{PL-P}^{solv}(\nu) \quad (31)$$

Here, $A$ indicates either energy $E$, entropy $S$, or the free energy $G$. Such analysis allows for detailed dissection of the differential contributions of propane versus methanol binding to this


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608–2621

common pocket. The spatial difference maps are shown in Figure 5, again with the results for propane (panels a–c) stacked over those for methanol (panels d–f).

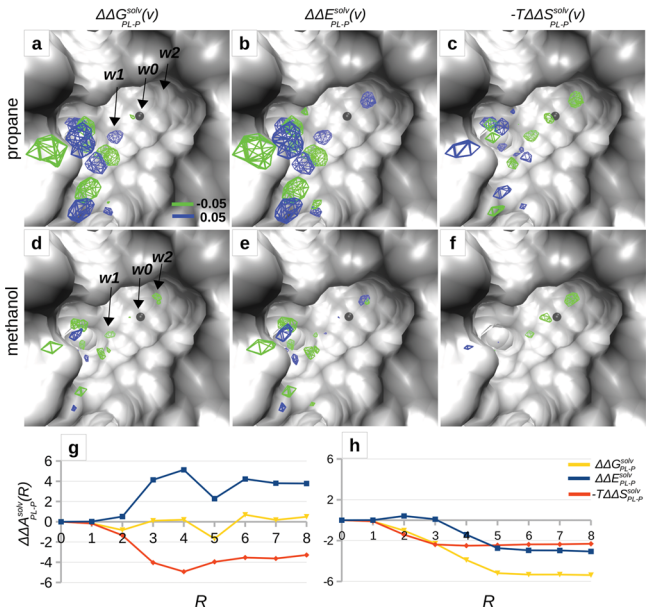



**Figure 5.** Visual analysis of the differences in the thermodynamic contributions of water to the binding of propane and methanol to pocket PP1/PM4 in p38 MAP kinase. Panels a–c and g show data for propane binding and d–f and h for methanol. Panels a and d show the difference contours $\Delta\Delta G^{solv}_{PL-P}(v)$ plotted at a favorable cutoff of < −0.05 kcal/mol (green) and an unfavorable cutoff of >0.05 kcal/mol (blue). Panels b and e show $\Delta\Delta E^{solv}_{PL-P}(v)$ plotted at a favorable cutoff of < −0.05 kcal/mol (green) and an unfavorable cutoff of >0.05 kcal/mol (blue). Panels c and f show $-T\Delta\Delta S^{solv}_{PL-P}(v)$ at a favorable cutoff of <−0.05 kcal/mol (green) and an unfavorable cutoff of >0.05 kcal/mol (blue). Panels g and h show the cumulative sum of the difference in solvation thermodynamic quantities $\Delta\Delta E^{solv}_{PL-P}(R)$, $-T\Delta\Delta S^{solv}_{PL-P}(R)$, and $\Delta\Delta G^{solv}_{PL-P}(R)$, as a function of distance $R$ from the binding site center. The pocket center in panels a-f is demarcated by a small black sphere, and the proximal water sites are indicated by arrows.

The free energy difference spatial distributions, $\Delta\Delta G^{solv}_{PL-P}(v)$, are shown as contours at favorable (<−0.05 kcal/mol; green) and unfavorable (>0.05 kcal/mol; blue) cutoffs in panels a and d of Figure 5. Panels b and e present the $\Delta\Delta E^{solv}_{PL-P}(v)$ favorable (<−0.05 kcal/mol; green) and unfavorable (>0.05 kcal/mol; blue) values, while the entropic terms $-T\Delta\Delta S^{solv}_{PL-P}(v)$ at < −0.05 kcal/mol in green and >0.05 kcal/mol in blue are in panels c and f. Cumulative energy $\Delta\Delta E^{solv}_{PL-P}(R)$, entropy $-T\Delta\Delta S^{solv}_{PL-P}(R)$, and free energy $\Delta\Delta G^{solv}_{PL-P}(R)$ contributions as a function of distance from pocket center are in panels g and h, for propane and methanol, respectively. With propane there is a favorable $\Delta\Delta G^{solv}_{PL-P}$ region near the center of the pocket corresponding to w0, with an unfavorable region corresponding to w1. The favorable contribution of w0 is associated with a favorable entropy contribution (Panel c). With w1, a favorable entropic contribution is more than compensated by an unfavorable $\Delta\Delta E^{solv}_{PL-P}$ contribution (panel b) leading to the unfavorable free energy contribution. Energy-entropy compensation also occurs with w2 upon propane binding, with the balance yielding no significant $\Delta\Delta G^{solv}_{PL-P}$ for this region. Moving away from the pocket center shows favorable $\Delta\Delta G^{solv}_{PL-P}$ gions in the vicinity of 5Å followed by unfavorable regions. This alternating pattern is also observed for entropic contributions which compensate the energetic regions (compare panels b and c). The reason for the alternating profile is likely due to the shifting of water molecules in PL, compared to the P state. The energy-entropy compensating behavior is evident in the cumulative $\Delta\Delta A^{solv}_{PL-P}(R)$ plot in panel g for propane. The energy contribution is unfavorable, compensated by favorable entropic contribution with the magnitude of both terms increasing out to 4 Å. A reversal in the trend occurs at 5 Å, associated with favorable free energy region beyond w1, followed by relatively small changes. Thus, the solvent thermodynamic contributions to the binding of propane are dominated by energy-entropy compensation.

The thermodynamic contributions of water to methanol binding are strikingly different than from propane. As seen in Figure 5, panel d, there are no significant $\Delta\Delta G^{solv}_{PL-P}$ contributions in the center of the pocket corresponding to w0, while favorable regions are observed corresponding to both w1 and w2. The favorable $\Delta\Delta G^{solv}_{PL-P}$ for the w1 and w2 regions is due to small unfavorable $\Delta\Delta E^{solv}_{PL-P}$ contributions (panel e) combined with significantly favorable $\Delta\Delta S^{solv}_{PL-P}$ contributions (panel f). The less unfavorable $\Delta\Delta E^{solv}_{PL-P}$ is consistent with partial displacement of these waters by methanol, as opposed to the full displacement that occurs upon propane binding. This overall effect is seen in the cumulative plot out to 3 Å in panel h. Favorable energy difference regions are present moving further from the binding pocket, which are not compensated by any entropic loss, leading to the favorable $\Delta\Delta G^{solv}_{PL-P}$ regions (panel e) and the cumulative more favorable $G$ (panel h). Thus, in contrast to the energy−entropy compensation with propane binding, energy−entropy reinforcement is occurring with methanol binding. Potentially contributing to this are favorable water−methanol interactions, also shown in Figure 3, that lead to only partial desolvation of the binding site, thereby leading to a less unfavorable $\Delta\Delta E^{solv}_{PL-P}$ contribution. We note here that the presented analysis highlights the spatial distribution differential of the solvent thermodynamic contributions between the PL and P states. In terms of the total energy and entropy, after including contributions from the L state and $\Delta S^{*}_{L}$, the binding of both ligands to this pocket happens via the re-enforcement mechanism (Tables 2 and 3).

The results summarized above and shown in Figures 4 and 5 represent an unprecedented picture of the thermodynamic forces that drive binding. It is notable how binding of these two small molecules to the same pocket happens through very different mechanisms, with the most notable being the differential contributions of water regions in the binding site in the presence and absence of the ligand. We anticipate that the present approach will be of utility for providing atomic level thermodynamic insights into binding mechanisms in a wide range of systems.

## 4. DISCUSSION

Our goal in this study was to provide thermodynamic insights in atomic detail into the binding of small molecules to proteins. While the importance of modeling water in the context of noncovalent binding has been long recognized,[36,37] rigorous estimation of thermodynamics associated with water reorganization upon binding remains a difficult task. Implicit solvent models have had successes in simplifying this and have provided avenues for constructing statistical thermodynamic approaches to model binding.[38] However, implicit solvents do not capture the molecular nature of water, and the directionality of hydrogen bonds, which become important at


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

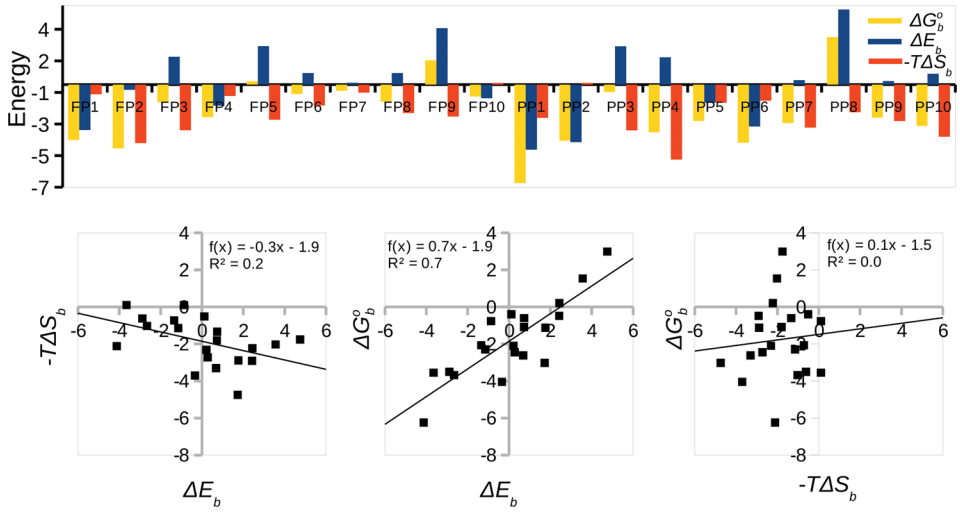



**Figure 6.** Thermodynamic signature of propane binding to different pockets on Factor Xa and P38 MAP kinase. The top panel shows the standard binding free energy, binding energy, and entropic contributions in yellow, blue, and orange, respectively. The nomenclature of binding sites indicates the protein, ligand, and pocket ID; e.g., FP1 indicates the protein Factor Xa (F), the ligand propane (P), and binding pocket 1. The bottom panels show correlation plots between the 3 thermodynamic quantities, with trend lines and associated parameters. The units are kcal/mol.

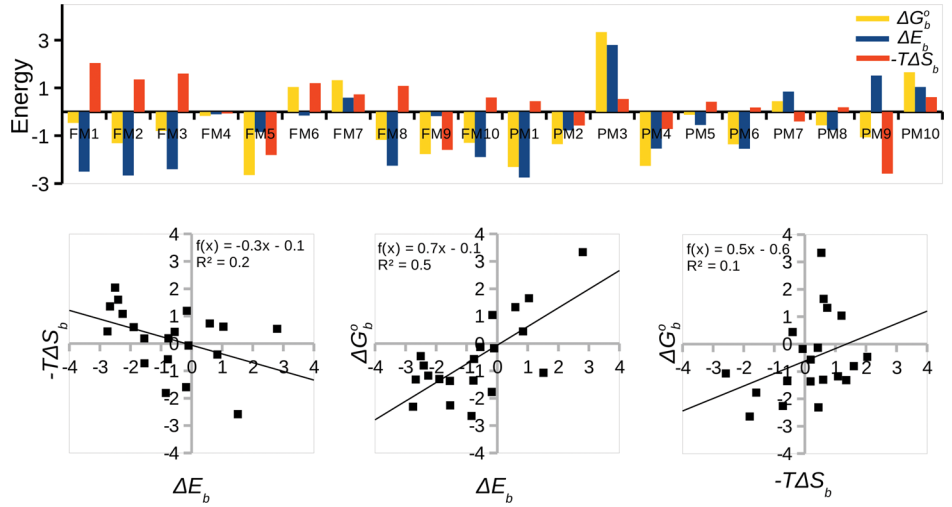



**Figure 7.** Thermodynamic signature of methanol binding to different pockets on Factor Xa and P38 MAP kinase. The top panel shows the standard binding free energy, binding energy, and entropic contributions in yellow, blue, and orange, respectively. The nomenclature of binding sites indicates the protein, ligand, and pocket ID; e.g., FM1 indicates the protein Factor Xa (F), the ligand methanol (M), and binding pocket 1. The bottom panels show correlation plots between the 3 thermodynamic quantities, with trend lines and associated parameters. The units are kcal/mol.

the length scales at which protein−ligand binding occurs. Thermodynamic pathway methods such as double decoupling free energy perturbation[1] or physical pathway methods[39] that include explicit water models have been shown to be powerful tools to calculate absolute binding free energies. However, these methods do not provide atomic scale thermodynamic insights into binding, e.g., the individual enthalpic contributions of different degrees of freedom, or the entropic contributions of the solute and solvent.

Two previous studies have taken important steps toward quantitatively understanding the driving forces implicated in binding using idealized binding sites and model ligands with solvent modeled explicitly.[10,12] In the present study we extend those efforts by identifying a set of binding events that are representative of real protein−ligand binding. This was achieved by employing MD simulations of two selected proteins in a 1 M solution of ligands. The 400 ns long cumulative sampling time resulted in the identification of a large number of sites that show spontaneous binding of the ligands. This approach thus ensures that the sites we selected for analysis represent true binding sites that participate in biomolecular interactions. Indeed the known ligand binding sites on these proteins show some of the highest ligand binding affinities calculated by our method; site PP1 in P38MK is a known ligand binding site,[19] and FP2 in FXa is the site S4 occupied by a large number of known ligands[18] (Figure 2). Interestingly, site S1 in FXa did not show any occupancy of propane or methanol. One possible reason for this could be this pocket being negatively charged such that favorable interactions between the ligands and the binding site cannot overcome the cost of desolvating that site. Another reason could be the strong restraints used here, which may prevent the necessary protein conformational transitions needed for binding in this relatively buried pocket.

The presented method for computing binding thermodynamics followed an end-point strategy, where configurational


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

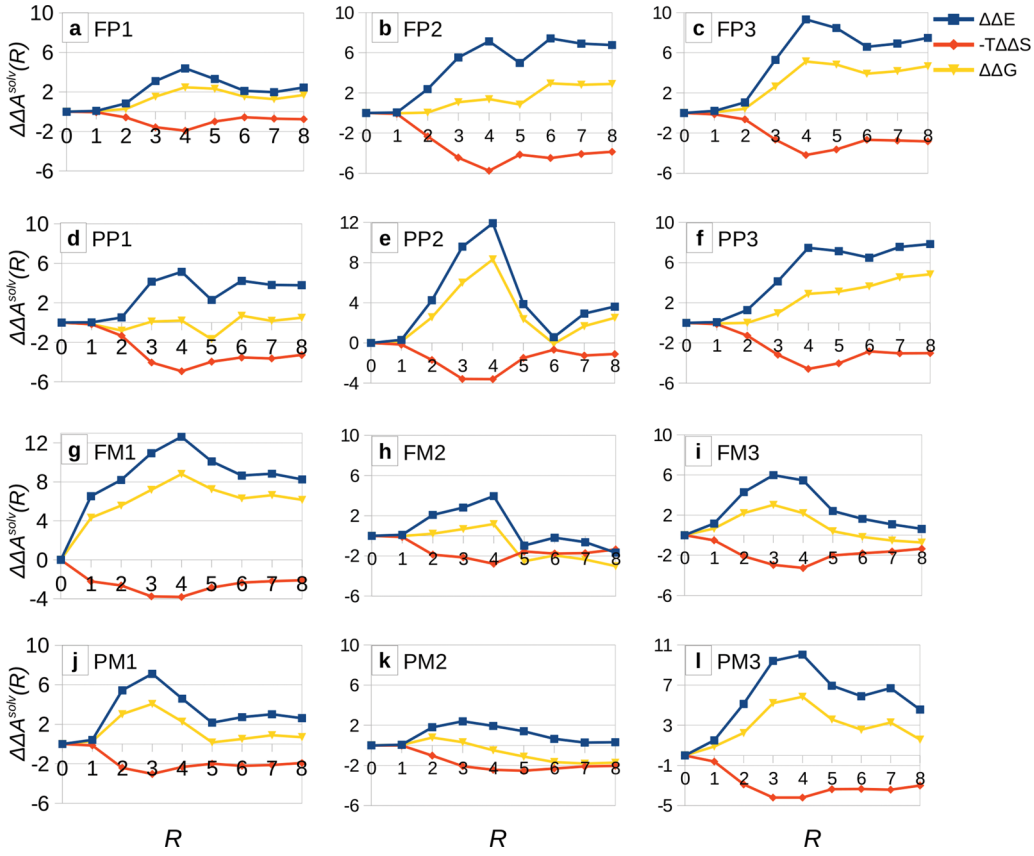



**Figure 8.** Cumulative sum of the difference in solvation thermodynamic quantities $\Delta\Delta E^{solv}_{PL-P}(R)$, $-T\Delta\Delta S^{solv}_{PL-P}(R)$, and $\Delta\Delta G^{solv}_{PL-P}(R)$, as a function of distance $R$ from the binding site center. A total of 12 pockets are shown.

ensembles in the bound and the free states were generated using MD simulations. These ensembles were analyzed using the Grid Inhomogeneous Solvation Theory (GIST[13]), with certain modifications, some of which were enabled by recent work on improved estimation of solvent entropy.[15,17] Our implementation for solvation thermodynamics was validated by reproducing the experimental hydration free energies of 6 solutes that spanned a range larger than 10 kcal/mol, to a reasonable degree of accuracy. Importantly, this work presents an approach to calculate standard binding free energies that account for explicit solvent reorganization. The solvent entropy estimation method used here is strictly defined for a single solute conformation. Therefore, to enable this computation, in each end state the protein conformation was restrained. Second, the choice of ligands was such that they lacked significant intramolecular degrees of freedom. In principle, eqs 3 and 9 provide generalizations for flexible unrestrained systems, which when used to analyze ensembles of conformations obtained from a long single or multiple simulations can be used to estimate excess solvent energy and entropy. Such a treatment would do away with approximations used in eqs 10−12. Information theoretic approaches[40,41] could be employed for the solute entropy calculation, which in the present treatment is approximated using eq 8. However, it is worth re-emphasizing that the choice of protein restraints and fragment sized ligands retains rigor in the calculations presented here and allows for well-converged estimates of thermodynamic quantities.

The thermodynamic signatures of propane and methanol binding to the two proteins are presented in Figures 6 and 7, respectively. The top panel in Figure 6 shows that, for propane, about 50% of binding processes resulted from energy−entropy re-enforcement and the remaining due to energy−entropy compensation, with energy contributing unfavorably in the latter. Total entropy was always found to contribute favorably to propane binding. The top panel in Figure 7 shows that favorable methanol binding in a majority of cases involves energy−entropy compensation, with entropy being unfavorable. Only two cases show a prominent reinforcement mechanism, and both of them happen to be known ligand binding sites: site FM5, which is the S4 pocket in FXa, and site PM4, which is the above-discussed ligand binding site in P38MK (Figure 2). Thus, while the binding of propane or methanol shows trends over all the sites studied, important variations at certain sites occur, emphasizing the need for the analysis of individual binding sites.

The bottom panels in Figures 6 and 7 show correlation plots between the total binding energy, entropy, and free energy. The energy−entropy plots for both ligands show a compensating trend, but the association is weak. For propane, $\Delta G^{o}_{b}$ correlates strongly with $\Delta E_{b}$, and shows no correlation with entropy. These findings also hold true for methanol as described below, but the correlation of $\Delta E_{b}$ with $\Delta G^{o}_{b}$ becomes weaker. We note that $\Delta G^{o}_{b}$ does not show strong correlation with $\Delta n_{wat}$, any of the energetic terms $\Delta\Delta E_{ww}$, $\Delta E_{pw}$, $\Delta E_{lw}$, $E_{pl}$, or the entropic terms $-T\Delta\Delta S^{solv}_{b}$ or $-T\Delta S^{*}_{L}$, considered individually ($R^2$ < 0.3).

An important observation is that in all pockets and with both ligands the solvation entropy contribution is always favorable (Tables 2 and 3). Unfavorable total entropy in cases where it


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

occurs results due to the unfavorable overcompensation by ligand entropy loss. A comparison of Tables 2 and 3 shows that, on average, propane binding is entropically more favorable by about 2 kcal/mol than methanol. The bigger part of this difference arises from the larger rigid body entropy loss of methanol, by about 1.3 kcal/mol, and the remainder can be explained by the small difference in total solvation entropy of binding $\Delta\Delta S_b^{solv}$, of about 0.6 kcal/mol. This small difference in solvation entropy can be explained by the nearly 1 kcal/mol more favorable entropic contribution from the L state when propane binds, i.e., due to the release of water molecules interacting with the free ligand. Thus, generally speaking the data do not show a significant difference in the complexed state solvent entropy when comparing propane and methanol binding.

Figures 6 and 7 also show a qualitative consistency between the rank of the binding sites identified by MD simulations immersed in 1 M ligand solution and the binding free energies $\Delta G_b^o$ calculated by the end-point method. Sites FP1−FP10 are sorted and ranked in descending order on the basis of the occupancy of the highest occupancy voxel in each cluster. Similarly, sites PP1−PP10, FM1−FM10, and PM1−PM10 are also sorted. Thus, top ranked binding sites are predicted by the MD data to have high affinity for the fragment. This trend is qualitatively followed by the end-point based $\Delta G_b^o$ for propane as is clear from Figure 6. A similar trend exists for methanol binding, but not as consistent as for propane. Methanol binding is also found to be weaker than propane according to computed $\Delta G_b^o$, which is also observed in higher occupancies associated with propane binding (data not shown).

Out of the 40 pockets investigated, 8 pockets show an unfavorable $\Delta G_b^o$. In two cases $\Delta G_b^o$ is approximately 3 kcal/mol, and $\Delta G_b^o > 1$ kcal/mol in 4 other cases. This difference between the thermodynamic end-point and MD approaches may be due to the interactions between ligands in the latter. Repulsive interactions are present between propane molecules to avoid solute aggregation and methanol molecules may interact with each other. These ligand−ligand interactions are absent in the thermodynamic end-point approach.

Also presented are detailed spatial analyses of binding thermodynamics of both ligands in pocket PP1/PM4 of P38MK. The two ligands bind via different mechanisms: propane displaced 3 binding site water molecules resulting in a significant entropic contribution, compensated by lost energetic interactions made by those water molecules; methanol displaced lesser number of waters, which causes a relatively lower entropy gain, but also lacks an energetic penalty. To expand this analysis to a larger number of pockets the solvent contributions $\Delta\Delta E_{PL-P}^{solv}(R)$, $-T\Delta\Delta S_{PL-P}^{solv}(R)$, and $\Delta\Delta G_{PL-P}^{solv}(R)$ computed for pockets FP1−3, PP1−3, FM1−3, and PM1−3 are shown in Figure 8. The solvent entropic contribution as a function of distance $R$ from the center shows a minimum at approximately 3−4 Å. Ligand binding thus causes an increase in entropy within this distance from the center, which is due to the direct displacement of structured waters. In nearly all cases, there is a "correction" to this number as the entropy reduces from $R > 4$ Å to make the final entropic contribution less favorable. This can be rationalized as more ordering of waters that form interactions with the ligand. However, in most cases this correction is not large enough to fully compensate the entropy gained from the directly displaced waters. The energetic contribution follows a similar trend, where there is a steep rise due to the loss of binding site waters up to about 4 Å, followed by a correction at larger distances. The correction is relatively small in magnitude for propane, but significant for methanol. This is consistent with the potentially more favorable interactions that water would make with a polar ligand in the bound state. An exception to this trend is pocket PM4 shown in Figures 4 and 5. Another observation from Figure 8 is that, at R = 8 Å, most pockets show relatively plateaued thermodynamic properties. For some pockets (e.g., PP2, PM3), it appears that a larger analysis region would be warranted.

The presented approach offers attractive possibilities to develop and validate new approximations for predicting protein−ligand binding thermodynamics. For example, calculations of the individual contributions of energy and entropy of the different interacting species performed on experimentally characterized protein−ligand complexes can serve as target data for optimizing empirical parameters used in scoring functions and implicit solvent models.

Finally, we note that, in addition to providing a detailed thermodynamic analysis of binding, the method holds potential for applications in drug design. Spatial analysis of solvent thermodynamic contributions could serve as valuable information to propose ligand modifications that may lead to enthalpic or entropic stabilization. More importantly, as also suggested previously,[13] differential binding thermodynamics between congeneric pairs in a lead optimization series could be calculated by accounting for not only solvent associated terms but also ligand associated terms as performed in this work. The approach presented here offers avenues to incorporate the full thermodynamic cycle in the calculation of binding equilibria in contrast to prior applications incorporating only solvent associated terms calculated for the unbound protein alone.[6,42] However, it must be noted that the inclusion of protein−ligand flexibility would require further methodological developments as discussed above.

## 5. CONCLUSIONS

Presented is a method for the calculation of binding thermodynamics while employing an explicit solvent model. The method is used to calculate and analyze thermodynamic contributions arising from different degrees of freedom implicated in the binding of fragment sized ligands to diverse protein pockets. The ability to spatially resolve the thermodynamic contributions allowing for an atomic detail understanding of those terms is notable. Binding, using methanol and propane with the proteins Factor Xa and p38 MAP kinase as models, always led to a gain in solvent entropy, which was compensated by rigid body entropy loss of the ligand. On average, propane binding was associated with favorable entropy, whereas methanol binding is associated with favorable energy. Direct protein−ligand interaction energy was seen to be nearly as important as water reorganization energy in favoring binding, which were compensated by a loss of solute−water interactions. However, significant variations in the thermodynamic contributions to binding occur in the different sites, with results showing both energy−entropy re-enforcement and compensation to drive binding as well as a diversity of mechanisms for the binding of the studied nonpolar and polar ligands. Further, the ability of the presented method to compute thermodynamic contributions of specific degrees of freedom can provide target data that may potentially be used to develop improved approximations for modeling protein−ligand binding.


DOI: 10.1021/ja512054f
J. Am. Chem. Soc. 2015, 137, 2608−2621

## ■ ASSOCIATED CONTENT

### Supporting Information

Additional calculation details and additional data presented in table and figure forms. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

prabhu@outerbanks.umaryland.edu


### Notes

The authors declare the following competing financial interest(s): A.D.M. Jr. is Co-Founder and CSO of SilcBio LLC.

## ■ ACKNOWLEDGMENTS

This work was supported by NIH Grants CA107331, GM070855, and GM051501. The authors acknowledge computer time and resources from the Computer Aided Drug Design (CADD) Center at the University of Maryland, Baltimore.

## ■ REFERENCES


(1) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(2) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521−533.
(3) Rossky, Y.-K. C. . P. J. *Nature* **1998**, *392*, 696−699.
(4) Lazaridis, T. *J. Phys. Chem. B* **1998**, *102*, 3531−3541.
(5) Lazaridis, T. *J. Phys. Chem. B* **1998**, *102*, 3542−3550.
(6) Abel, R.; Young, T. K.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(7) Lazaridis, T. *Phys. Chem. Chem. Phys.* **2006**, *9*, 573−581.
(8) Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Hroux, A.; Sherman, W.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17889−17894.
(9) Setny, P.; Baron, R.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 2866−2871.
(10) Baron, R.; Setny, P.; Andrew McCammon, J. *J. Am. Chem. Soc.* **2010**, *132*, 12091−12097.
(11) Gerogiokas, G.; Calabro, G.; Henchman, R. H.; Southey, M. W. Y.; Law, R. J.; Michel, J. *J. Chem. Theory Comput.* **2014**, *10*, 35−48.
(12) Michel, J.; Henchman, R. H.; Gerogiokas, G.; Southey, M.; Mazanetz, M. P.; Law, R. J. *J. Chem. Theory Comput.* **2014**, *10*, 4055−4068.
(13) Nguyen, C. N.; Young, T. K.; Gilson, M. K. *J. Chem. Phys.* **2012**, *137*, 044101.
(14) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph* **1996**, *14*, 33−38.
(15) Huggins, D. J. *J. Chem. Theory Comput.* **2014**, *10*, 3617−3625.
(16) Raman, E. P.; Mackerell, A. D., Jr. *J. Chem. Phys.* **2013**, *139*, 055105.
(17) Huggins, D. J. *J. Comput. Chem.* **2014**, *35*, 377−385.
(18) Adler, M.; Davey, D. D.; Phillips, G. B.; Kim, S.-H.; Jancarik, J.; Rumennik, G.; Light, D. R.; Whitlow, M. *Biochemistry* **2000**, *39*, 12534−12542.
(19) Fitzgerald, C. E.; Patel, S. B.; Becker, J. W.; Cameron, P. M.; Zaller, D.; Pikounis, V. B.; O'Keefe1, S. J.; Scapin, G. *Nat. Struct. Biol.* **2003**, *10*, 764−769.
(20) Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E. F., Jr.; M, B.; Rodgers, J.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Arch. Biochem. Biophys.* **1978**, *185*, 584−591.
(21) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(22) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2010**, *6*, 3850−3856.
(23) Boukharta, L.; Gutirrez-de Tern, H.; qvist, J. *PLoS Comput. Biol.* **2014**, *10*, e1003585.
(24) Word, J.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735−1747.
(25) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(26) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wirkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(27) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(28) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(29) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2010**, *31*, 671−690.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(31) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198−2202.
(32) Guvench, O.; Mackerell, A. D., Jr. *PLOS Comput. Biol.* **2009**.
(33) Raman, E. P.; Yu, W.; Lakkaraju, S. K.; Mackerell, A. D., Jr. *J. Chem. Inf. Model.* **2013**, *53*, 3384−3398.
(34) Raman, E. P.; Vanommeslaeghe, K.; Mackerell, A. D., Jr. *J. Chem. Theory Comput.* **2012**, *8*, 3513.
(35) Lakkaraju, S. K.; Raman, E. P.; Yu, W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2014**, *10*, 2281−2290.
(36) Ben-Naim, A. *Proc. Indian Acad. Sci.* **1987**, *98*, 357−377.
(37) Poornima, C.; Dean, P. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521−531.
(38) Chang, C.; Potter, M. J.; Gilson, M. K. *J. Phys. Chem. B* **2003**, *107*, 1048−1055.
(39) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(40) King, B. M.; Silver, N. W.; Tidor, B. *J. Phys. Chem. B* **2012**, *116*, 2891−2904.
(41) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. *J. Comput. Chem.* **2008**, *29*, 1605−1614.
(42) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. *J. Chem. Theory Comput.* **2014**, *10*, 2769−2780.




DOI: 10.1021/ja512054f
*J. Am. Chem. Soc.* 2015, 137, 2608−2621