JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

3D-RISM-Dock: A New Fragment-Based Drug Design Protocol

Dragan Nikolić,[†] Nikolay Blinov,[†,‡] David Wishart,[§,∥] and Andriy Kovalenko*[,†,‡]

[†]National Institute for Nanotechnology, National Research Council of Canada, Edmonton, Alberta, Canada

[‡]Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada

[§]Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

[∥]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

Supporting Information

ABSTRACT: We explore a new approach in the rational design of specificity in molecular recognition of small molecules based on statistical-mechanical integral equation theory of molecular liquids in the form of the three-dimensional reference interaction site model with the Kovalenko−Hirata closure (3D-RISM-KH). The numerically stable iterative solution of conventional 3D-RISM equations includes the fragmental decomposition of flexible ligands, which are treated as distinct species in solvent mixtures of arbitrary complexity. The computed density functions for solution (including ligand) molecules are obtained as a set of discrete spatial grids that uniquely describe the continuous solvent-site distribution around the protein solute. Potentials of mean force derived from these distributions define the scoring function interfaced with the AutoDock program for an automated ranking of docked conformations. As a case study in terms of solvent composition, we analyze cooperative interactions encountered in the binding of a flexible thiamine molecule to the prion protein at near-physiological conditions. The predicted location and residency times of computed binding modes are in excellent agreement with the available experimental data.

1. INTRODUCTION

Computer-assisted[1] drug design is an important supplement to empirical screening in current pharmaceutical research,[2,3] and structural diversity is a prerequisite for further optimization of lead compounds.[4−6] *In silico* free energy-based ranking of specificities in binding small molecules is often obscured by avoided steric clashes with proteins due to the reduced size of fragments. This low specificity stems from the fact that a single fragment can potentially fit to biologically unrelated proteins.[7] Besides this, the ranking of a ligand depends on how reliable the scoring function is in terms of addressing all relevant contributions to the binding energy,[8] which are also assumed to be mutually separable and additive.[9]

Our recent *in silico* docking tests[10] showed that the standard docking programs using phenomenological descriptions of solvation are incapable of capturing the molecular structure of solvent sites inside the host cavity for the thiamine−prion protein (PrP) complex. While MD simulations can elicit time-resolved features in the collective interactions among solvent molecules, they operate on time scales too short to span the slow progress of molecular recognition in physiological solutions. Compared to explicit solvent MD simulations, the 3D-RISM-KH theory[11] yields an efficient solvent sampling at a reduced computational cost[12] by using statistical-mechanics description of solvent molecules. Moreover, 3D-RISM-KH theory provides thermodynamically reliable 3D density distributions for all solvent sites near solvated macromolecules. The presence of high-density peaks in the solvent distributions around the solutes is an indicator of a strong mutual affinity, which can be used to discriminate among small rigid molecules by matching the constituent atomic sites to the location of corresponding peaks.

The usefulness of such an approach within the standard 3D-RISM-KH framework was demonstrated earlier by Imai et al.[13] and most recently by Kiyota et al.[14] In particular, Kiyota et al.[14] argued that the 3D-RISM-KH theory presents itself as a powerful approach to analyzing binding affinities under different physical conditions imposed by chemical specificities of hydrogen bonding, hydrophobicity, and solvation entropic effects. Thus, the present work can be seen as a further development of the use of the 3D-RISM-KH theory as a rigorous computational approach for molecular recognition of proteins in solutions. Namely, by treating the ligand fragments as a part of the solvent, we compute and map the density distributions of all atomic sites onto the spatial grid around the protein and then use conventional docking algorithms to assemble, sample, score, and rank numerous ligand poses.

The 3D-RISM-KH binding affinities express the statistical preferences in protein−ligand interaction supported by the physical force field used for simulations and are interpreted in analogy to potentials of mean force (PMF). PMF-based scoring removes the need for empirical balancing of several opposing contributions to binding affinity, including desolvation, enthalpy, and entropy.[15] In contrast to the empirical scoring functions, the 3D-RISM-KH approach treats the solvation effects with atomic resolution using rigorous statistical mechanics of liquids. In particular, it accounts for possible structural solvation effects (competition between solvent particles and a ligand for binding sites), solvation entropic (including hydrophobic) effects, and solvent mediated hydrogen bonding.

Received: March 29, 2012
Published: August 7, 2012

Compared to the conventional grid-based docking algorithms, implemented for example in the AutoDock program,[16] a novel 3D-RISM-KH docking approach introduced in this paper is characterized by transferability of its solvation part, which accounts for (i) competition between different ligand species, (ii) ligand concentration and solvent composition effects, and (iii) the thermodynamic state of a system under consideration. Another conceptual advantage of our approach is that it can also be used for accurate ranking of docked conformations based on the free energy calculated within the same statistical mechanical framework and the same force field as used for docking. By employing 3D-RISM-KH for calculations of the solvation part of the binding free energy, we avoid using nontransferable implicit solvation models or computationally demanding thermodynamic integration methods for free energy calculations.

According to recent surveys,[17,18] the most popular open-source application for molecular docking is AutoDock,[16,19−22] which also constitutes a computational platform for development of pharmacological software. These numerous AutoDock spin-offs can be divided into three representative categories: (i) interfaces[23,24] to the well-tested[18,25] empirical AutoDock scoring function, (ii) refinements of sampling algorithms[26−31] and scoring parameters,[32] and (iii) parallelized implementations[33−39] for improved performance. In the current implementation of the 3D-RISM-Dock protocol, we completely replaced the empirical AutoDock scoring function by its PMF counterpart. We also state the required modifications to a 3D-RISM aware AutoDock source code that now can be used in complex mixtures of ingredients with diverse biological activities.

Previous characterization of human cerebrospinal fluid (CSF) by Wishart et al.[40] identified more than 300 compounds that are currently integrated in the Human Metabolome Database.[41] Using the heteronuclear single-quantum coherence (HSQC) experiments combined with isothermal titration calorimetry, these compounds have been systematically assessed for their level of binding specificity to a truncated form of recombinant Syrian hamster prion protein shPrP(90−232). In a recent NMR study, Perez-Pineiro et al.[42] indicated that thiamine (which can be found in human CSF) specifically binds to the prion protein with an affinity in the 60−120 μM range. In this paper, we apply the 3D-RISM-Dock protocol to aid these experimental studies and offer a plausible theoretical description of shPrP-thiamine complexes that are in good agreement with the experimental data.

2. METHODS

2.1. Theoretical Formulation. Thermodynamic properties of multicomponent liquids can be studied in the framework of integral equation theory of molecular liquids by using the method of distribution and correlation functions.[43] Briefly, the pair correlation functions $g^{ij}(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)$ describe the interaction of any two molecules belonging to any of $i, j = 1, ..., N_{comp}$ components in the solution. In homogeneous liquids, the correlation functions depend only on relative position $\boldsymbol{r}_{12} = \boldsymbol{r}_1 - \boldsymbol{r}_2$ and orientation $\boldsymbol{\Theta}_{12} = \boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2$ of molecules, and the resulting six-dimensional molecular Ornstein−Zernike integral equation can be written[44] in terms of both the total $h^{ij}(\boldsymbol{r}_{12};\boldsymbol{\Theta}_{12}) = g^{ij}(\boldsymbol{r}_{12};\boldsymbol{\Theta}_{12}) - 1$ and direct $c^{ij}(\boldsymbol{r}_{12};\boldsymbol{\Theta}_{12})$ correlation functions

$$h^{ij}(\boldsymbol{r}_{12};\boldsymbol{\Theta}_{12}) = c^{ij}(\boldsymbol{r}_{12};\boldsymbol{\Theta}_{12}) + \sum_{k=1}^{N_{comp}} \frac{\rho_k}{8\pi^2} \int_{R^3}\int_{\Omega} d\boldsymbol{r}_3\, d\boldsymbol{\Theta}_3\, c^{ik}(\boldsymbol{r}_{13};\boldsymbol{\Theta}_{13})\, h^{kj}(\boldsymbol{r}_{32};\boldsymbol{\Theta}_{32}) \quad (1)$$

such that Ω contains all possible molecular orientations defined through a set of Euler angles and R^3 defines the total liquid volume in which the component k has the bulk density ρ_k. Equation 1 relates two sets of correlation functions and thus should be complemented by an additional set of relations called closure.[44] The approximate solution of eq 1 relies on the widely used RISM approach pioneered by Chandler and Andersen.[45] In a liquid composed of a pure solvent and an infinitely diluted solute, eq 1 becomes a system of three uncoupled equations for *solute−solute* ($i = u, j = u'$), *solute−solvent* ($i = u, j = v$), and *solvent−solvent* ($i = v, j = v'$) correlation functions, each solved independently.

2.1.1. 3D-RISM-KH Theory of Molecular Solvation. Initially devised by Chandler and co-workers,[46,47] the 3D-RISM method has been formulated in the hypernetted chain (HNC) approximation by Beglov and Roux.[48] Preserving the atomic-level structural information of a solute in eq 1, Kovalenko and Hirata partially integrated the *solute−solvent* molecular Ornstein−Zernike equations over the solvent orientations, reducing it to a set of 3D integral equations, including the analytical treatment of electrostatic asymptotics.[49,50] Complemented by the Kovalenko−Hirata closure (see below), this approach is known as the 3D-RISM-KH molecular theory of solvation[11,51] For each site $\gamma = 1, ..., N_{solv}$ in a solvent molecule, we solve a system of N_{solv} 3D equations

$$h_\gamma(\boldsymbol{r}) = \sum_{\alpha=1}^{N_{solv}} \int_{R^3} d\boldsymbol{r}'\, c_\alpha(\boldsymbol{r}-\boldsymbol{r}')\, \chi_{\alpha\gamma}(|\boldsymbol{r}'|) \quad (2)$$

with respect to a given set of intermolecular *solute−solvent site* total correlation $h_\gamma(\boldsymbol{r})$ and direct correlation $c_\gamma(\boldsymbol{r})$ functions. Prior knowledge of the bulk solvent susceptibility function $\chi_{\alpha\gamma}(r)$ is required. The latter can be computed from the radial *solvent site−site* total correlation function $h^{solv}_{\alpha\gamma}(r)$

$$\chi_{\alpha\gamma}(r) = \omega^{solv}_{\alpha\gamma}(r) + \rho_\alpha h^{solv}_{\alpha\gamma}(r) \quad (3)$$

where the geometry-aware term $\omega^{solv}_{\alpha\gamma}(r) = \delta_{\alpha\gamma}\delta(r) + (1-\delta_{\alpha\gamma})\delta(r - l_{\alpha\gamma})/(4\pi l^2_{\alpha\gamma})$ contains site−site separations $l_{\alpha\gamma} = |\boldsymbol{r}_\alpha - \boldsymbol{r}_\gamma|$ in the solvent molecule. Calculation of $h^{solv}_{\alpha\gamma}(r)$ functions, see Figure 4, is carried out by using the one-dimensional dielectrically consistent RISM theory[52,53] (1D-DRISM).

The standard model for the site−site pair potential contains contributions from electrostatic and Lennard-Jones (LJ) interactions

$$u_{p\gamma}(\boldsymbol{r}) = \frac{q_p q_\gamma}{|\boldsymbol{r}_p - \boldsymbol{r}|} + 4\varepsilon^{LJ}_{p\gamma}\left[\left(\frac{\sigma^{LJ}_{p\gamma}}{|\boldsymbol{r}_p - \boldsymbol{r}|}\right)^{12} - \left(\frac{\sigma^{LJ}_{p\gamma}}{|\boldsymbol{r}_p - \boldsymbol{r}|}\right)^{6}\right] \quad (4)$$

with mixing parameters $\sigma^{LJ}_{p\gamma} = (\sigma^{LJ}_p + \sigma^{LJ}_\gamma)/2$ and $\varepsilon^{LJ}_{p\gamma} = (\varepsilon^{LJ}_p \varepsilon^{LJ}_\gamma)^{1/2}$. For a given 3D interaction potential $u_\gamma(\boldsymbol{r}) = \sum_p u_{p\gamma}(\boldsymbol{r})$ between the solute molecule and the γ site of solvent, eq 2 must be complemented by N_{solv} auxiliary closure relations derived from a cluster diagram analysis[44,54−56]

$$h_\gamma(\boldsymbol{r}) = \exp(d_\gamma(\boldsymbol{r}) + b_\gamma(\boldsymbol{r})) - 1$$
$$d_\gamma(\boldsymbol{r}) = -u_\gamma(\boldsymbol{r})/(k_B T) + h_\gamma(\boldsymbol{r}) - c_\gamma(\boldsymbol{r}) \quad (5)$$

where k_B is the Boltzmann constant, T is the temperature, and direct correlation functions decay asymptotically as $c_\gamma(\boldsymbol{r}) \rightarrow -u_\gamma(\boldsymbol{r})/(k_B T)$. Bridge functions $b_\gamma(\boldsymbol{r})$ are given as an infinite series of integrals over higher-order total correlation functions. Being computationally intractable, bridge functions have continued to receive increasing interest over the past two decades[11,57,58] through the development of new and improved closure relations. Kovalenko and Hirata[11] proposed the 3D-KH closure

$$h_\gamma(\boldsymbol{r}) = \begin{cases} \exp(d_\gamma(\boldsymbol{r})) - 1 & : \; d_\gamma(\boldsymbol{r}) \le 0 \\ d_\gamma(\boldsymbol{r}) & : \; d_\gamma(\boldsymbol{r}) > 0 \end{cases} \quad (6)$$

which applies the HNC approximation to the density-depleted regions ($h_\gamma < 0$) and switches to the mean-spherical approximation (MSA) in density-enhanced domains ($h_\gamma > 0$) of the local solvation structure. The HNC is conventionally defined by $b_\gamma(\boldsymbol{r}) = 0$ in the whole space,[59] while the MSA usually enforces $c_\gamma(\boldsymbol{r}) \rightarrow -u_\gamma(\boldsymbol{r})/(k_B T)$ everywhere outside the repulsive core. Thus, the 3D-RISM-KH theory yields the solutions to eqs 2 and 6 and provides direct structural information on the solvation shell around the solute molecule in terms of the probability density $\rho_\gamma g_\gamma(\boldsymbol{r})$ for each solvent interaction site γ. Here, ρ_γ denotes the average number density of site γ in the solvent bulk where the site-specific normalized density distribution $g_\gamma(\boldsymbol{r}) = 1 + h_\gamma(\boldsymbol{r})$ approaches unity at large distances.

2.1.2. AutoDock. In the original AutoDock program, the change in free energy upon ligand binding is modeled by an empirical scoring function[16]

$$\Delta G^{\mathrm{aq}}_{\mathrm{bind}} = \sum_{p,l} \left\langle \left[\frac{A_{pl}}{r_{pl}^{12}} - \frac{B_{pl}}{r_{pl}^{6}}\right] W_{\mathrm{vdW}} + E(\theta)\left[\frac{C_{pl}}{r_{pl}^{12}} - \frac{D_{pl}}{r_{pl}^{10}}\right] W_{\mathrm{hb}} + \frac{q_l q_p}{r_{pl}\varepsilon(r_{pl})} W_{\mathrm{el}} + [S_l V_p + S_p V_l] e^{-r_{pl}^2/2\sigma^2} W_{\mathrm{des}} \right\rangle + N_{\mathrm{tor}} W_{\mathrm{tor}} \quad (7)$$

where the pairwise summation runs over all atomic sites l and p in both the ligand and the protein, respectively, and $\varepsilon(r_{pl})$ is the distance-dependent dielectric constant. The scoring function in eq 7 is evaluated on a spatial map of points and further calibrated for the current set of atom types: H, C, N, O, F, Mg, P, S, Cl, Ca, Mn, Fe, Zn, Br, and I. For each atom type found in the ligand, there is a probe atom that is systematically placed at every grid point around the receptor in order to compute its total interaction energy (affinity) with all sites in the receptor. This grid-based presampling produces several affinity maps that will be interpolated during the evaluation of a candidate conformation. Calibration factors W_{vdW}, W_{hb}, and W_{el} account for van der Waals, hydrogen bonding, and screened electrostatic interaction.

A departure from an ideal hydrogen bonding geometry is measured by an angle θ and penalized by a directionality weight function $E(\theta)$ that also controls the actual number of hydrogen bonds assigned to each point in the grid. The linear regression coefficient W_{des} optimizes the charge-dependent desolvation penalty within the Gaussian distance 2.5σ where V_i and $S_i = ASP_i + QASP|q_i|$ stand for atomic fragmental volumes and solvation parameters, respectively. Any restriction in the number of rotatable bonds (N_{tor}) reduces the conformational degrees of freedom in the ligand and gives rise to an unfavorable entropy of ligand binding controlled by the weight factor W_{tor}. For the current numerical values of the AutoDock linear free energy model coefficients and energetic parameters in eq 7, see Huey et al.[16]

Several map modification techniques to extend the applicability of AutoDock exist.[22] These address a variety of problems involving the prediction of bridging water positions,[60] covalently bound coordination complexes,[21] protein flexibility,[61] and exploration of chemical space via mutable atoms to manage the intermediate states in the ligand growth.[62] In the following, we deploy the automated workflow of AutoDock as a test platform for validation of the 3D-RISM-Dock method and provide the list of necessary source code changes.

2.1.3. 3D-RISM-Dock Protocol. In a recent development of a 3D-RISM-based ligand-mapping method, Imai et al.[13] deployed the all-atom model for a receptor that is immersed in a ligand−water mixture. The most probable binding modes of rigid fragments are identified via 3D-spatial distributions $g_\gamma(\boldsymbol{r})$ of their atomic sites. Kiyota et al.[14] also demonstrated the sensitivity with which the 3D-RISM method captures the multiscale effects of water on the binding affinities, opening up new avenues for computational fragment-based drug design.[63]

The present study takes this notion one step further into the realm of flexible ligands by embedding the PMF[49,50,64]

$$W_\gamma(\boldsymbol{r}) = -k_B T \ln(g_\gamma(\boldsymbol{r})) \quad (8)$$

into the modified AutoDock program.[21] Namely, each ligand site γ binds to the receptor with an affinity proportional to its PMF such that the optimal docking mode for the whole ligand corresponds to the global minimum in the complex energy landscape $\Lambda(\boldsymbol{r}) + \Delta(\boldsymbol{r})$. The total binding affinity is subject to constraints embedded in $\Lambda(\boldsymbol{r}) \propto \sum_\gamma W_\gamma(\boldsymbol{r})$ imposed by the geometry of the ligand molecule and optionally to a barrier $\Delta(\boldsymbol{r}) \propto -\sum_\sigma W_\sigma(\boldsymbol{r})$, activated at the $W_\sigma(\boldsymbol{r}) \le -(3/2)k_B T$ threshold, which accounts for the highly−conserved solvent sites σ. For example, inside the binding cavity, the energies of the interaction between the protein and docked ligand or the protein and water molecules are often comparable.[65] However, the competition between the desolvation of the ligand and the expulsion of water depends on differences in both the entropic and the energetic properties of water when it is placed in the bulk or in the highly structured binding cavity.[66] Thus, $\Delta(\boldsymbol{r})$ accounts for free energy release upon displacing the structured waters from the principal hydration sites. Within the 3D-RISM-KH approximation, the need for a nonzero $\Delta(\boldsymbol{r})$ penalty stems from the possible underestimation of ionic association peaks due to partial linearization of the closure relation found in eq 6.

As pointed out by Kiyota et al.,[14] it is straightforward to identify prime targets in drug design as those ligand sites γ that contribute the most to the overall binding affinity $\Lambda(\boldsymbol{r})$. The outlined solvent mapping algorithm can be used to evaluate the free energy minima upon protein−ligand complexation if the empirical model in eq 7 is replaced by a new scoring function

$$\Delta G^{\mathrm{3D}}_{\mathrm{RISM}} = \Lambda(\boldsymbol{r}) + \Delta(\boldsymbol{r}) + N_{\mathrm{tor}} W_{\mathrm{tor}} \quad (9)$$

which is currently being calibrated on a large set of diverse protein−ligand complexes of variable flexibility and for several charge assignment methods.

Table S-1 of the Supporting Information provides a complete list of the changes to the AutoDock source code required for incorporation of the 3D-RISM-KH scoring function, see eq 9. Thanks to its good initial design,[67] the keyword-based input for AutoGrid and AutoDock allows for inclusion of new switches to control the docking simulations. We introduced the "rism3d" keyword in the ligand input file, such that AutoDock proceeds with the adaptive global−local search of optimal ligand conformations using the scoring function in eq 9.

2.1.4. Simulation Details. Docking experiments based on the 3D-RISM-Dock algorithm involve the following steps: (i) 1D-DRISM calculations[52,53] of the radial pair distributions entering the bulk solvent (ligand) susceptibility functions $\chi_{\alpha\gamma}(r)$ in eq 3, (ii) computation of the 3D-RISM-KH density distributions found in eqs 2 and 6 required for the construction of PMF maps (eq 8), and (iii) automated ligand placement and ranking of binding modes according to the PMF-based scoring function (eq 9). In the 1D- and 3D-RISM-KH calculations, the Pm6 partial charges (q^{mop}) and the van der Waals parameters from the Amber *gaff* force field were used for the thiamine molecule. The force field parameters of the protein were assigned according to the Amber *ff11* force field. For details on the force field selection and its effect of docking results, see section SI-2 of the Supporting Information.

The maximal computational resources required to complete step i depend greatly on the convergence criterion between two successive iterations, which is currently set to a highly conservative 1×10^{-8} for the change in the cumulative residual of all site distribution functions. For this particular case, the large number of solvent sites (40 for thiamine in saline water solutions) has a typical memory consumption of about 125 MB per solvent site, whereas the number of fragments and charge distribution used to describe the thiamine molecule may also prolong the serial execution times for step i. These may vary from 18 min per solvent site in the case of rigid and chargless thiamine down to 90 s per solvent site for positively charged thiamine with nine fragments. To complete step ii with the moderately conservative convergence criterion of 1×10^{-5} for the change in the cumulative residual, the typical Intel based CPU time consumption in a serial execution mode is 13 s per solute site. The typical number of sites in the shPrP solute is about 1680 (the number varies depending on the pH value of saline water solution), and if the thiamine is represented with three fragments simultaneously present as a part of solvent, the total serial execution time for step ii is about 6 h with 4 GB of memory consumption. In this particular setting, all thiamine binding affinity maps (PMFs) were confined to a 64 Å box with 0.5 Å resolution and rounded to four decimal places as a compromise to the computational speed of step iii.

For the purpose of analyzing our docking experiments, step iii deploys AutoDock's evolutionary Lamarckian genetic algorithm in which only the flexible ligand has the role of an evolving individual while the receptor is kept rigid. Each ligand conformation is encoded as a number sequence consisting of three coordinates for the ligand translation, four components of the quaternion (the axis and the angle of rotation) describing the ligand orientation, and angles for all active torsions in the ligand molecule. Receptor and ligand were embedded in a 64 Å cubic grid with 128 points in each direction, whereas quaternion components and the torsion angle were randomly varied in the $[-1,+1]^3 \times [-\pi, +\pi]$ and $[-\pi, +\pi]$ ranges, respectively.

The population size was varied between 100 and 1000 individuals, and the recommended[68] size of 300 was used for all production runs as a trade-off between the computational cost of 5 million fitness evaluations and the overall docking performance. Generational promotion among the individual ligand conformations was conducted by both the two-point and the arithmetic crossover operators[69] at a rate of 0.8. Operating at a mutation rate of 0.02, the Cauchy distribution had the zero mean and the unit spread. The local search settings for the modified Solis and Wets algorithm[70] were such that the initial value of the standard deviation parameter $\rho_t = 1$ was expanded/contracted 2-fold after the four subsequent successes/failures and search was terminated if $\rho_t < 0.01$ or a maximum of 300 iterations had been reached.

2.2. Initial Structures for Docking Studies. We adopt the shPrP structures of James et al.[71] (PDB entry 1B10) that were previously used, for example, by Lima et al.[72] in support of small-angle X-ray scattering measurements of bound double-stranded DNA. In particular, we will illustrate our results for the 1B10 conformer #17, see Figure 1, which has been slightly

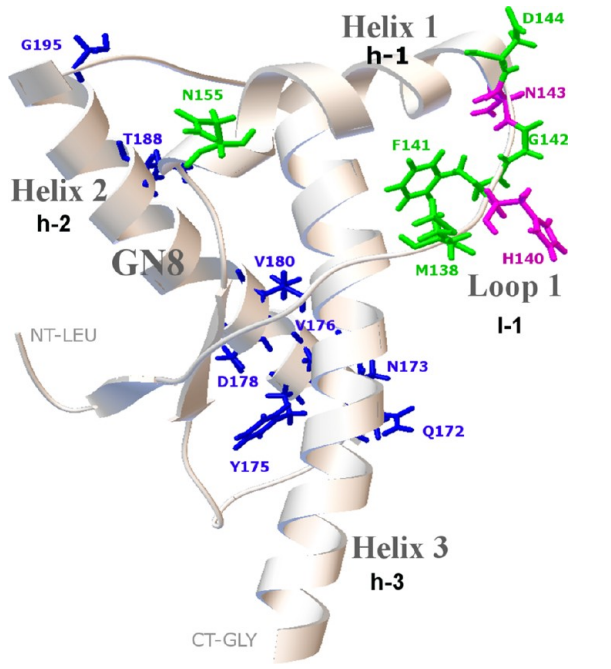

Figure 1. Ribbon representation of shPrP with locations of residues strongly perturbed due to thiamine binding:[42] blue, distal conformational changes; magenta, significant chemical shift perturbations; and green, direct contacts with the thiamine.

modified[10] to support the recent heteronuclear NMR screening of thiamine affinities.[42] The residues L125−G228 form a core domain that contains two antiparallel β-strands (M129−G131, V161−Y163) and three α-helices (h-1: D144−M154, h-2: Q172−T193, h-3: E200−D227). According to Perez-Pineiro et al.,[42] the thiamine binding site resides between Helix 1 and the preceding Loop 1 (M138−Y150). Figure 1 also indicates the location of the interaction site proposed by Kuwata et al.[73] for the 2-pyrrolidin-1-yl-*N*-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamid (termed "GN8"), which is also proven to bind to PrP.

In their comprehensive review of crystal structures and enzymatic actions of thiamine and its derivatives, Louloudi and Hadjiliadis[74] concluded that detailed roles of this molecule are

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

dependent on both the cellular environment and its structural changes. The initial 3D structure of ThOH, taken from the Human Metabolome Database,[41] is further geometry optimized by using the Pm6 semiempirical method MOPAC[75] and is shown in Figure 2. The positive charge of thiamine is

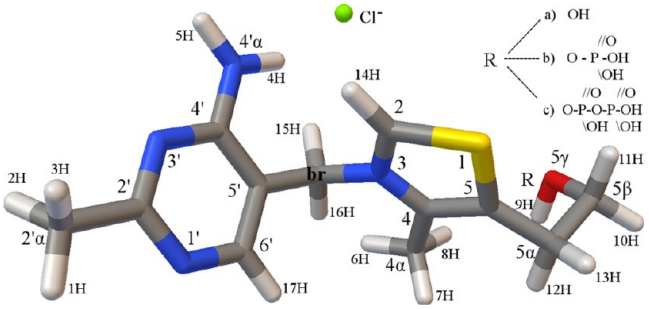

Figure 2. General structure of the *syn* configuration of thiamine and its derivatives. The pyrimidine ring and the thiazole ring with the hydroxyethyl side-chain are bridged by the methylenic C(br)H_2 group. The functional group R in the side chain relates to (a) thiamine, ThOH; (b) thiamine monophosphate, ThMP; and (c) thiamine diphosphate, ThDP. Color code: gray, carbon; white, hydrogen; red, oxygen; blue, nitrogen; and yellow, sulfur.

delocalized over the thiazole ring[76−78] where the S(1) site strongly attracts lone electron pairs and negative ions.[77,79] Consequently, the electron-deficient thiazole ring can engage in charge-transfer and stacking interactions with aromatic residues.[80] The side-chain at the C(5) location in the thiazole ring may contain the hydroxyethyl group in ThOH or the pyrophosphate group in ThDP. However, Kozik[81] reports on the irrelevance of phosphate groups in the C(5) side-chain due to the strong interaction between the thiazole ring and negatively charged carboxyl groups in buckwheat thiamine binding protein and concludes that none of the three fragments appear as absolutely essential for the thiamine binding. In the case of biologically active ThDP in solutions, the heterocyclic rings are not perturbed by β-phosphate functional groups.[82] The loss of important phosphate groups in the ThOH moiety is most likely to be compensated within the binding site by an increased π-stacking pyrimidine interaction,[83] resulting in similar dissociation constants K_d for all three thiamine analogs.[84]

In a recent NMR study,[85] several ionization states of thiamine analogues were investigated—the unprotonated 4′-aminopyrimidine form of ThOH and the protonated 4′-aminopyrimidinium (APH^+) form of thiamine hydrochloride (Th·HCl). The perturbation of the ^{13}C and ^{15}N chemical shift anisotropy tensors for C(2) and N(4′) atoms due to the absence of the chloride anion suggests its location close to the S(1) site at the interatomic N(4′)−Cl^- and C(2)−Cl^- distances of 3.2 Å and 3.7 Å, respectively.[85] Using the 3D-RISM affinity maps obtained in the water solution of sodium chloride at an infinite dilution, we investigated the most probable position of the chloride anion in the thiamine monochloride (ThOH·Cl) complex as a function of a charge assignment method. Only the Pm6 partial charges (see q^{mop} column in Table S-2 of the Supporting Information) were capable of reproducing the experimental coordination[85] of the chloride anion relative to the ThOH cation.

2.3. Fragmental Decomposition of Ligands. One of the objectives of the present study is to characterize the predictive behavior of the 3D-RISM-Dock method as a function of a fragmentation scheme and to assess a loss of specificity due to a reduced size of thiamine fragments. Namely, small fragments are less likely to have steric clashes that prevent them from fitting into a designated binding crevasse. Within the context of blind docking, however, the concomitant degradation of site specificity for small molecules results in large number of lead anchors. These leading seeds may eventually lose their potency in the linking process due to the imposed stereochemical constraints present in a viable conjugated ligand.[86]

In order to preserve the desirable orientations of constituent fragments, the fragment-linking method requires the design of suitable linkers. For this purpose, we divided the rigid thiamine molecule into a set of three or nine fragments, see Figure 3, and

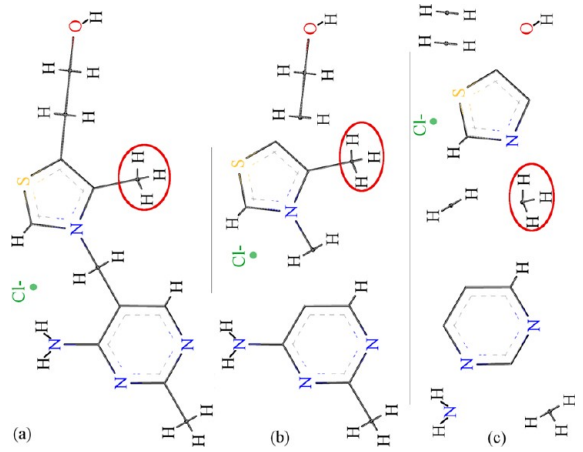

Figure 3. Partitioning scheme of thiamine: (a) as a rigid molecule, ThOH; (b) a set of three fragments, ThOH/3; (c) set of nine fragments, ThOH/9. The chloride anion is treated as a separate component in the solvent mixture.

introduced them as distinct moieties in the solvent composition. In this partitioning scheme, each of the high-affinity moieties in the three-fragment case can be obtained by assembling low-affinity units belonging to the nine-fragment set and further conjugated into a specific thiamine conformer using methylene linkers. Throughout this study, the inherent flexibility of the thiamine molecule in eqs 7 and 9 was kept constant by setting the number of rotatable bonds to $N_{tor} = 8$. These rotatable bonds are C(2′α)−C(2′), C(4′)−N(4′α), C(5′)−C(br), C(br)−N(3), C(4)−C(4α), C(5)−C(5α), C-(5α)−C(5β), and C(5β)−O(5γ).

2.4. MD Simulations and Energy Minimization. The binding conformations obtained in the 3D-RISM-Dock docking simulations are further assessed for their dynamic stability. Prior to explicit solvent simulations, the structures are minimized using the generalized Born (GB)/Surface Area (SA) implicit solvation model as implemented in the Amber11 suite of programs[87] with the *ff03* and *gaff* all-atom force fields[88,89] for the protein and thiamine, respectively. In particular, the charges of the thiamine atomic sites were assigned by using the antechamber module based on the Gasteiger charging method, but the AM1-BCC scheme was used later on in the explicit solvent simulations. The purpose of this initial implicit solvent energy minimization is to relax the protein degrees of freedom (constrained during 3D-RISM-Dock simulations) and to improve shape complementarity between the binding pocket and the ligand. This preliminary step also helps avoid artificial

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

injection of explicit water molecules between the ligand and/or the protein in an unoptimized binding pocket. Implicit solvent minimization was performed in two steps with the "igb=5" option of the AMBER11 package. As a first step, the tyrosine residues of shPrP were kept restrained to optimize possible π-stacking interactions between the protein and the ligand, whereas the second step involves an unrestrained minimization aimed at converging and reducing the energy gradient RMS values below 0.02 kcal/mol. In all minimization runs, the monovalent salt concentration was maintained at 150 mM. Both in the implicit and explicit solvation simulations described later, the disulfide bond between C179 and C214 was kept intact.

To sample the conformational space of the thiamine−shPrP complexes in bound and unbound states, explicit-solvent all-atom MD calculations were performed. The GBSA-minimized structures were placed in rectangular boxes and solvated with the SPC/E[90] water molecules such that a distance of at least 12 Å was allowed between ThOH−shPrP complexes and periodic boundaries. Charge−balancing Cl^- counterions were placed in a shell around shPrP in positions of extrema of a Coulombic potential. The above procedure was followed by two minimization runs to relax the degrees of freedom of both the solvent and the shPrP/thiamine complexes. Potentially equilibrated systems were sampled at the end of two subsequent thermalization/equilibration runs and subjected to MD productive simulations. Namely, in thermalization runs, gradual increases in temperature from 0 K to the target value of 298.15 K were followed by the constant pressure MD in order to relax solvent degrees of freedom and adjust its density to approximately 1 g/cm^3. In the equilibration MD runs with the 1 fs time step, the coordinates of the shPrP and thiamine were restrained to their minimum energy values.

In production runs, all hydrogen-containing bonds were subjected to holonomic constraints using the SHAKE algorithm[91] with a relative geometric tolerance for coordinate resetting of 10^{-5} Å, allowing a time step of 2 fs. Throughout the simulation, the system was coupled to a heat bath at T = 298.15 K via Langevin dynamics carried out with a collision frequency of 5 ps^{-1}. Isotropic position scaling with a relaxation time of 2 ps was used to maintain a constant pressure of 1 bar. The residue-based cutoff for nonbonded interactions was set to 16 Å, whereas other parameters for nonbonded interactions, including the parameters for the particle mesh Ewald method for computing the long-range Coulomb interactions, were set to their default values.[87]

3. RESULTS AND DISCUSSION

As already discussed in section 2.1.4, docking experiments based on the 3D-RISM-DOCK algorithm involve the following steps: (i) 1D-DRISM calculations of the radial pair distributions entering the bulk solvent susceptibility functions $\chi_{\alpha\gamma}(r)$ in eq 3; (ii) computation of the 3D-RISM-KH density distributions found in eqs 2 and 6 and required for the construction of PMF maps (eq 8); (iii) automated ligand placement and ranking of binding modes according to the PMF-based scoring function (eq 9). The previous two steps bare a certain resemblance to the corresponding procedures of conventional docking algorithms, such as the generation of affinity maps and search for optimal binding poses, but the first step is unique to the 3D-RISM-DOCK protocol. Namely, by accounting for solvent composition and thermodynamic states as well as ligand geometry and concentration through susceptibility functions used to calculate affinity maps, we are able to analyze binding processes under different conditions without reparametrization of the scoring function.

For illustration, in Figure 4 we show pair radial distribution (correlation) functions for selected solvent and ligand atomic

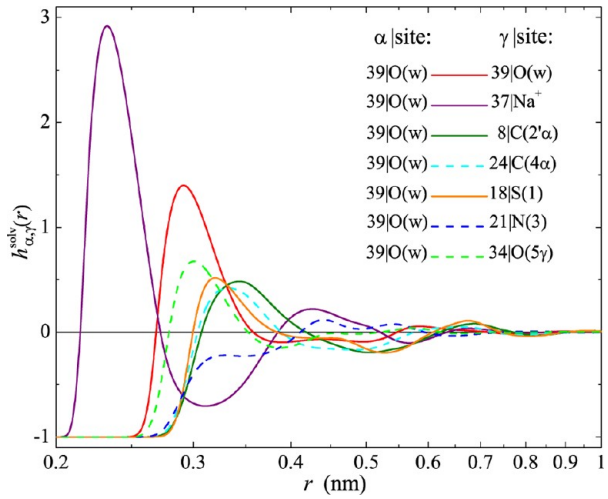

Figure 4. Radial solvent site−site total correlation functions between water oxygen and selected solvent/thiamine atomic sites calculated with the 1D-DRISM-KH theory.[52,53] For a description of ligand sites, see Table S-2.

sites around water oxygen obtained by using the 1D-DRISM theory[52,53] with the 1D-KH closure for aqueous sodium chloride solution with thiamine dissolved at 7 mM concentration. These functions describe correlations between different atomic sites in bulk solution. The convolution of the 3D direct correlation functions with the solvent susceptibility according to eq 2 propagate these bulk correlations into 3D densities, accounting for a competition between atomic sites in both the ligand and solvent molecules for most favorable placement. The intensity and location of density peaks in the first solvation shell, for example, depend both on the type (defined by the force field) of site and the geometry of the molecular fragment that the site belongs to. This is clearly seen in Figure 4 for the carbon sites from two methyl groups attached to the thiazole and pyrimidine rings of thiamine. The differences in the site−site correlation functions, especially due to geometrical restraints entering the RISM equations (both in 1D and 3D cases), contribute to a higher degree of specificity of docked conformations predicted with 3D-RISM-DOCK compared to conventional docking algorithms that ignore the actual geometrical constraints in the affinity maps. Additionally, the 3D-RISM-DOCK algorithm accounts for the effects of solvent composition and the thermodynamic state of the system in a similar fashion as the ligand mapping approach proposed by Imai et al.[13]

3.1. 3D-RISM-KH Solvation Structure and PMFs. The 3D distributions of ligand and solvent atomic sites around a receptor, $g_\gamma(\boldsymbol{r})$, define the PMF used in the 3D-RISM-DOCK scoring function for ranking binding modes. Sets of points around the receptor with $g_\gamma(\boldsymbol{r}) > 1$ and $g_\gamma(\boldsymbol{r}) < 1$ tentatively define domains of favorable and unfavorable locations of ligand (solvent) sites, respectively. Thus, visual inspection of 3D densities can provide useful information on possible binding modes prior to docking experiments. For example, electrostatic effects may play an important role in the binding of thiamine to

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

the prion protein mainly due to the net positive charge of thiamine, see Table S-2. The substantial positive fractional charges at N(3) and S(1) atomic sites in the thiazole ring are expected to associate with free anions in solution. Since the experimental binding pocket of an isoelectric shPrP contains a few ionizable residues (including negatively charged D144 and D147), desolvation effects and possible competition between thiamine and solvent particles for binding sites may play an important role in thiamine binding to shPrP.

The electrostatic potential (ESP) isosurfaces have been obtained by solving the Poisson−Boltzmann equation[92] at 0.15 mol/L ionic strength and are illustrated in Figure 5a and b at the $\pm 1.5\ k_B T/e$ level. The analysis of the electrostatic field obtained for different protonation states of the protein, reveals that favorable environments for binding positively charged species exist in the C-terminus (the bottom part) and in the experimental binding pocket of thiamine (in wire-frame representation). The monovalent salt site density isosurfaces obtained by solving 3D-RISM-KH equations are shown in blue for sodium (Figure 5c and d) and in red for chloride (Figure 5e and f) ions. Both the protein and ligand (in the experimental bound conformation) are represented in Figure 5 by their solvent exposed areas, colored in gray for the protein and in green wire-frame for the ligand.

While the ESP may serve as an indicator of energetically favorable binding pockets of charged species, the actual binding modes will depend on a subtle balance of electrostatic, dispersion, hydrophobic, and entropic effects. In fact, hydrophobic and dispersion interactions are enhanced by shape complementarity between the ligand and receptor, which may make the binding of larger (neutral) ligands or water molecules more preferable compared to, for example, atomic ions. Even for small charged particles, identification of possible binding modes and statistical distributions around the protein is a rather complex problem which cannot be solved by simple analysis of the electrostatic potential alone. Thus, under neutral conditions, the distribution of chloride anions found in Figure 5e is as much defined by the ESP as it is by the distribution of sodium cations in Figure 5c. The implication is that compared to chloride anions, sodium cations may have a more profound effect on ligand binding modes. On an empirical level, these effects can be partially accounted for by adjusted desolvation potentials, hindering the transferability of the scoring function to different solvent compositions and thermodynamic states.

In general, it is expected that high density domains will be located near possible binding sites such as in the grooves of a receptor surface, whereas lower density regions due to steric restraints are expected to be distributed uniformly around a receptor. For illustration, we also show in Figure 6 the computed 3D-RISM-KH density distributions for selected atomic sites of the thiamine molecule. The sulfur S(1) site density in Figure 6a is represented by an isosurface corresponding to 0.1% of its bulk density, which translates to a PMF of ∼+4 kcal/mol. Thus, due to possible steric clashes with the protein, this isosurface illustrates energetically unfavorable locations for the sulfur S(1) site in thiazole. The favorable locations of the methyl carbon C(4α) site, shown in Figure 6a at 150% of its bulk density, correspond to a PMF of ∼−0.24 kcal/mol. It is expected (see Imai et al.[13]) that the most probable binding modes could be determined by a maximum overlap of high density domains among all ligand sites, as shown in Figure 6b, resulting in competition between different binding modes.

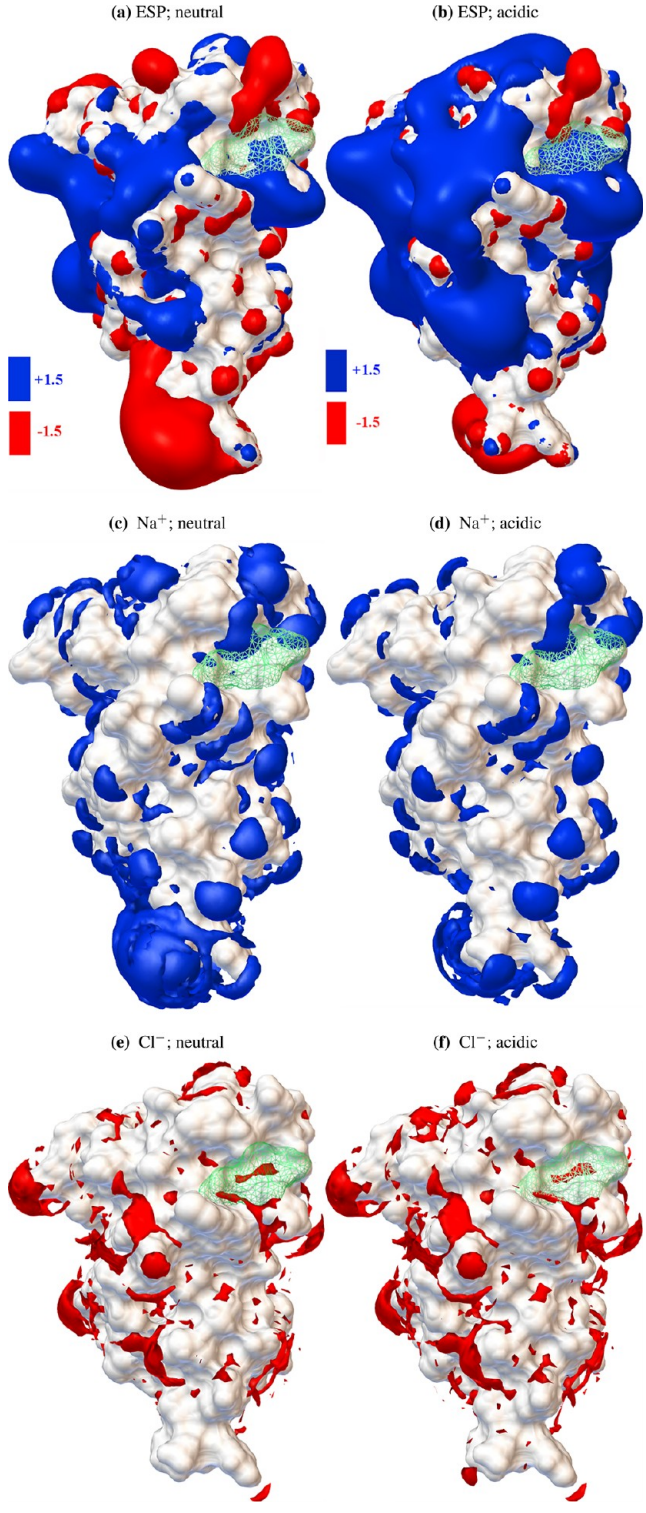

Figure 5. Isosurface representation of the electrostatic potential (a,b) and the 3D-RISM-KH density distributions of sodium (c,d) and chloride (e,f) ions around the prion protein under neutral and acidic conditions; see the text for details. The protein is represented by the solvent-accessible surface.

Using eq 8, the 3D-RISM-KH density distributions of all solvent/ligand sites can be converted into their PMF analogs, as shown in Figure 7 for the isoelectric shPrP. Overall, Figure 7 indicates that monovalent salt cations are expected to compete with the thiazole ring for a common binding region, most

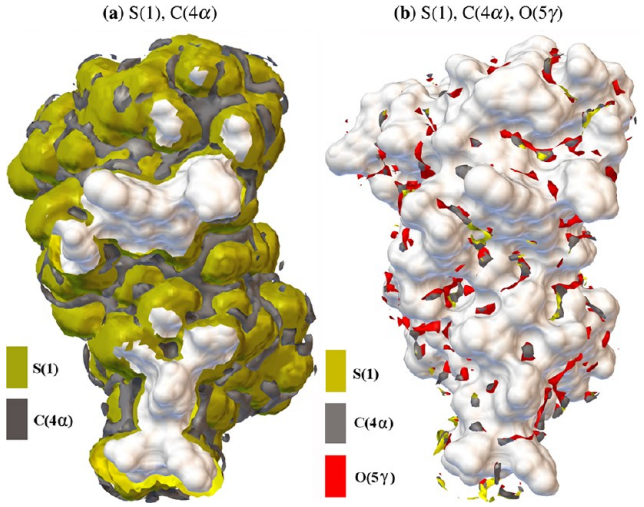

Figure 6. Isosurface representation of the 3D-RISM-KH density distributions for selected thiazole sites: (a) sulfur and methyl carbon at isolevels of 0.1% and 150% of their bulk values, respectively; (b) isolevels of thiazole sulfur and methyl carbon at 400% of their bulk density and side-chain hydroxyl oxygen at 500% of its bulk density. The front part of isosurfaces in panel a are clipped to show the solvent exposed surface of the protein.

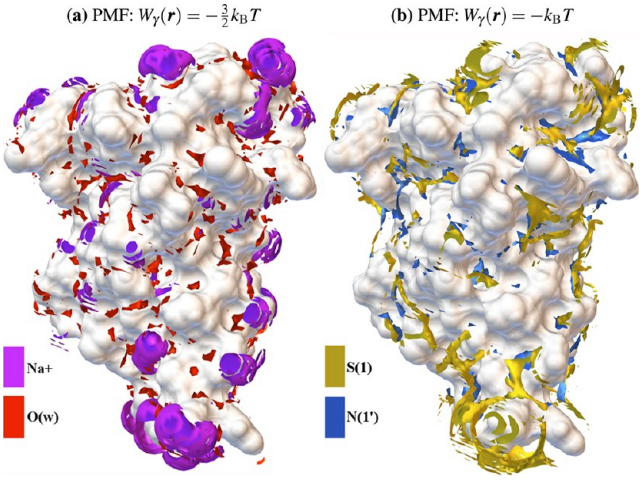

Figure 7. The PMF representation of the 3D-RISM-KH solvation structure of the isoelectric shPrP in 0.15 mol/L saline water solution. See Table S-2 for the additional details.

notably the carboxyl groups of the residues D144, D147, and E152 in Helix 1 (h-1 in Figure 1). In this region, the established patterns of hydration contain highly conserved waters around the amide group of the G142 side-chain and between the hydroxyl groups of partially hydrophobic Y150 and Y157 residues.

3.2. Linking of Ligand Fragments. The final step of a docking experiment with the 3D-RISM-Dock protocol involves ligand placement and ranking of binding modes according to the PMF-based scoring function in eq 9. PMFs are obtained by solving 3D-RISM-KH equations with ligand degrees of freedom treated as a part of a complex solvent. The constraints imposed by the ligand geometry are explicitly embedded in calculations through the term $\omega^{solv}_{\alpha\gamma}(r) = \delta_{\alpha\gamma}\delta(r) + (1 - \delta_{\alpha\gamma})\delta(r - l_{\alpha\gamma})/(4\pi l^2_{\alpha\gamma})$ entering the bulk susceptibility (see section 2.1.1 for details). Choosing different ligand fragmentation schemes defines a particular set of geometrical constraints imposed on the system which in turn affects docking results. In this context, the conventional grid-based docking software can be regarded as an extreme case with no geometrical constraints used to calculate affinity maps for ligand sites. Within the 3D-RISM-Dock protocol, such an approach corresponds to calculating PMFs for each ligand site independently (similar to, for example, how affinity maps are generated by AutoGrid module of the AutoDock suite), or by treating sites as individual particles in a complex solvent, partially retaining the ligand site−site correlations. On the other hand, a ligand molecule can be treated as a rigid particle by the 3D-RISM-Dock protocol, thus fully accounting for the fixed ligand geometry. Such an approach does not involve a linking stage in docking simulations (only ligand placement and ranking of docked poses), thus removing the need for using the assumption about PMF's additivity. An obvious drawback of such treatment is a lack of ligand flexibility, which may impede fitting a ligand into the binding pocket.

The results of our docking simulations (see next sections and section SI-4 of the Supporting Information) show that any reduction in the number of fragments in the thiamine molecule leads to an increased binding specificity. At the same time, based on the scoring function derived from the 3D-RISM-KH densities, a continued fragment growth may impede the optimal packing of the linked fragments. Thus, in order to improve the efficiency of the docking protocols, any practical application should possess the right balance between specificities and optimal packing of drug fragments. In most cases, the decision on the fragmentation scheme to be used can be made by examining the internal flexibility of the ligand set by the number of rotatable bonds.

We first proceed by discussing the fragmental binding preferences supported by shPrP using the fully charged thiamine molecule. In that respect, Figure 8 illustrates how the recognition patterns for three isolated thiamine fragments may change upon their linking into the thiamine. Namely, the pyrimidine fragment in Figure 8a and b has the highest affinity toward the experimental binding pocket, followed by the thiazole fragment in Figure 8c and d, whereas the C(5) side-chain in Figure 8e and f ranks the lowest. Upon linking, see Figure 8g and h, the assembled thiamine molecule only partially inherits the affinities of disparate fragments, and new viable docking sites may emerge.

The apparent difficulty in building a precompiled library of drug-like fragments using the 3D-RISM-KH approach is associated with eq 9 and the assumption that fragmental PMFs are additive. In other words, how different will the scoring function ΔG^{3D}_{RISM} be if all fragments of interest are jointly kept as a part of solvent compared to the computationally more efficient alternative of introducing each fragment separately? To quantify these differences, we have calculated the binding affinities (PMFs) between a single water molecule and three chargeless thiamine fragments in saline water solutions at infinite dilution. Results are illustrated in Figure 9 as the radial distribution of fragmental PMFs obtained by binning individual site PMFs into 0.1-Å-thick spherical shells centered around the single water molecule.

In this conceptual example, by treating the single water molecule as a receptor and keeping all three chargeless thiamine fragments as part of the solvent, we were able to compute the many-body PMFs for each site belonging to one of mutually interacting fragments. Summing all individual site PMFs that constitute a single fragment resulted in the solid curves given in

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

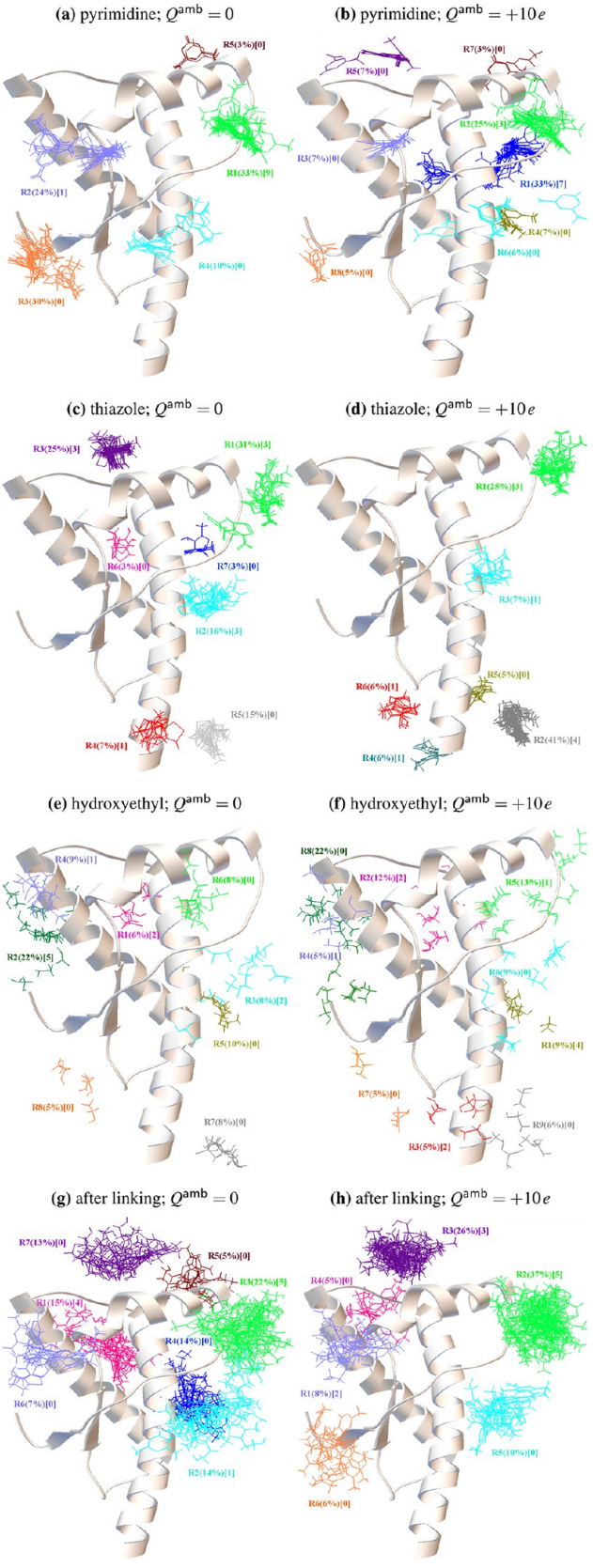

Figure 8. Docking and linking of disparate thiamine fragments carrying Pm6 charges, q^{mop}, as supported by the 3D-RISM-DOCK protocol. See text and Figure 3 for details.

Figure 9. The same procedure is repeated for each fragment separately, resulting in fragmental PMFs that are deficient in

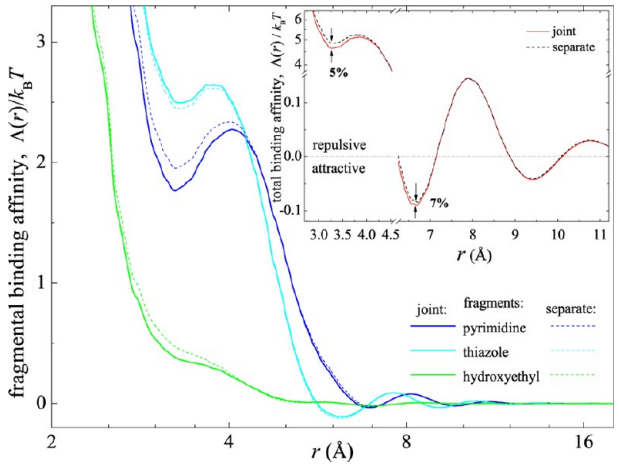

Figure 9. Radially resolved differences in fragmental short-range PMFs: (solid lines) fragments are simultaneously present as a part of the solvent; (dashed lines) each fragment is separately treated as a part of the solvent and thus independent of the other two. Partition of the thiamine follows Figure 3b, where three chargeless fragments are pyrimidine (blue), thiazole (cyan), and hydroxyethyl (green).

site−site interactions with other fragments and shown as dashed curves in Figure 9. The inset found in Figure 9 contains the final PMFs for the whole thiamine molecule: a red solid curve corresponds to mutually competing fragments, whereas a black dashed curve ignores any interfragment interaction. It is evident that the neglect of interfragmental competition only causes the local minima in the total binding affinity to be slightly less pronounced. However, the radial positions of both the repulsive and attractive total affinities remain insensitive to short-range interfragment interaction. This result not only justifies the use of 3D-RISM-KH affinity maps produced independently for small fragments but more importantly suggests that the optimal binding pose of linked fragments may also inherit the individual preferences obtained separately for each fragment.

Even though docking of isolated fragments offers a tremendous computational advantage with an acceptable loss of accuracy for binding affinities, in the remainder of this article, all thiamine fragments were simultaneously introduced as part of the solvent. By keeping all thiamine fragments as part of the solvent, we can gain a more complete insight on how the fragment size influences the displacement of localized water, as illustrated in Figure S-2 of the Supporting Information. This figure also shows how the choice of ligand fragment size may affect the density distributions of ligand atomic sites (fragmental additivity of PMFs entering 3D-RISM-DOCK scoring function). The most dramatic fragmentation effect has been found in the radial distribution of water oxygen around thiazole's methyl group (see circled fragment in Figure 3). When compared to the rigid thiamine (ThOH/1), the water coordination in the first solvation shell of the C(4α) exhibits a sharp increase if thiazole's methyl group is treated as a free fragment, as in ThOH/9, and becomes less pronounced upon linking to the thiazole aromatic ring (ThOH/3). Therefore, the capability of the thiazole fragment to release the strongly bound water molecules from binding pockets of proteins directly depends on the degree of flexibility of its methyl group.

3.3. Steric Fitness and Hydrophobic Interactions. Each cavity structure in the receptor has different characteristics for the inclusion of guest molecules, and it is widely accepted that

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

thiamine-binding sites must be hydrophobic in character.[93−98] For example, the hydrophobic isoleucine in the thiamine binding pocket found in pyruvate decarboxylase from *Saccharomyces cerevisiae* appears to be crucial[98] for an optimal alignment of the ionizable groups and for the enhanced hydrogen bonding to 4′-aminopyrimidine of ThDP. The central catalytic role of isoleucine is demonstrated by Guo et al.[98] via substitution by methionine (in pyruvate oxidase) and by leucine (in transketolase), confirming that each thiamine-dependent enzyme requires an optimal hydrophobic side chain at this position. Thomas et al.[99] investigated the enhanced potency of noncharged thiamine analogs in terms of increased hydrophobicity[100] and found an unprecedented *in vitro* activity of deazathiamine derivatives attributed to the thiopene ring due to a strong reduction in the desolvation penalty compared to the thiazole ring.

Electrostatic contributions, such as dipole−dipole, induced polarization, and hydrogen bonding interactions, also play an important role in the recognition of ligands. As follows from Table S-2, the conventional charge assignment techniques may lead to significant differences in the atomic partial charges in both the shPrP and the thiamine, which in turn may cause subtle variations in the concentration gradients of solvent species and affect the ranking of docked poses. In order to eliminate the effect of charge distribution in the thiamine molecule, the initial assessment of functional regions in the shPrP was performed with the reference model in which all of the atomic charges in the thiamine were set to zero. We then proceeded with shape complementarity studies by keeping the protein and the rest of solvent components electrically charged. Because solvent particles retain their charges, docking experiments with chargeless thiamine account for solvent-mediated electrostatic effects, including hydrophobicity.

In this recognition regime at hydrophobic interfaces, the 3D-RISM-Dock binding efficiency of thiamine fragments depends on both the solvent composition and the short-range interaction in eq 4. Figure 10 summarizes the trends in shape complementarity between the chargeless thiamine and the isoelectric/positive shPrP in an aqueous solution of monovalent salt (150 mM NaCl). For salt-free mixtures, Figures S-3 and S-4 of the Supporting Information illustrate the complexation sensitivity of the chargeless thiamine in the presence of bound active-site waters. The nomenclature used to describe different populations in the identified functional regions is as follows. Binding affinities for clusters of docked conformations are ranked in order of the increasing average energy from the most negative value R1. For clarity, only the clusters with a relative size above the 5% level have been presented and are reported in parentheses. Numbers enclosed in square brackets denote how many conformations from the given cluster globally rank among the 10 best.

Figure 10 also addresses the thiamine flexibility results obtained for the ThOH/3 (panels a and c) and the ThOH/9 (panels b and d), where the thiamine has been assembled by linking three and nine fragments, respectively. Ranking of chargeless thiamine conformations depends on an ability of the constituent fragments to maximize the number of short-range contacts with protein residues, a restriction that favors small-size fragments. Being the smallest of the three fragments (see Figure 3b), the hydroxyethyl C(5) side-chain penetrates the most into the observed binding region[42] near helix 1, followed by the thiazole and pyrimidine fragments.

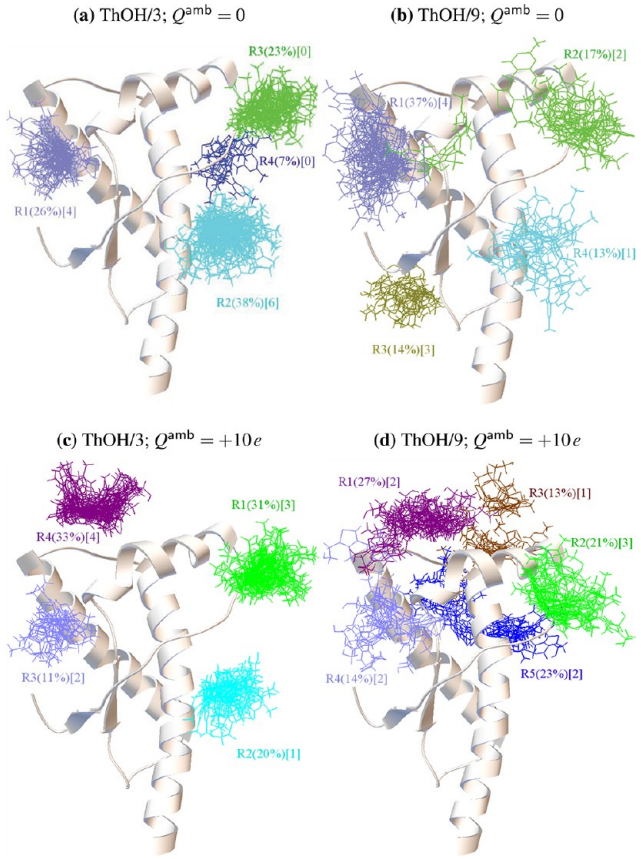

Figure 10. Analysis of the solvent-induced effects on 3D-RISM-Dock binding patterns for the chargeless thiamine. In the Amber charging scheme, panels a and b and panels c and d correspond to the isoelectric (Q^{amb} = 0) and the net positive (Q^{amb} = +10e) prion protein, respectively. Strongly persistent water and sodium sites are included in the scoring function, e.g., $\Delta(\boldsymbol{r}) = -\sum_{\sigma=37}^{39} W_\sigma(\boldsymbol{r})$. See text and Table S-2 for details.

The modulation of the scoring function ΔG^{3D}_{RISM} in eq 9 by strongly bound waters and monovalent cations was first studied by neglecting the high-density regions of solvent sites (see Figures S-3a, S-3b, S-4a, and S-4b), then by retaining only the water sites (see Figures S-3c, S-3d, S-4c, and S-4d), and finally by including both the water and sodium cation sites (see Figure 10). By comparing Figure 10 and Figures S-3c and S-4, we find two recurring regions in which the chargeless thiamine persistently dwells, regardless of the net charge on the prion protein and regardless of whether we used a deployed fragmentation scheme or an explicit inclusion of persistent solvent sites in the 3D-RISM scoring function.

The first region is located between helix 1 and the preceding loop (l-1) at the experimental binding site of thiamine, as reported by Perez-Pineiro et al.[42] and represented by the green thiamine clusters in Figure 10. The second region coincides with the GN8 interaction site proposed by Kuwata et al.[73] and is located between helix 2 and helix 1, as shown in lavender in Figure 10. The binding region between helix 1 and the l-2 loop subsequent to helix 2, is found to be sensitive to the inclusion of the preserved sodium cation sites only in the case of the isoelectric shPrP (see Figure 10a and b) and otherwise remains stable under highly acidic conditions in Figure 10c and d. The accessibility of the C-terminus interface appears to be hindered the most by the inclusion of strongly bound solvent-rich

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

regions (especially water), as illustrated in Figures S-3 and S-4 of the Supporting Information.

3.4. Selection of Top-Ranked Binding Modes. Even though the results of the previous section allow us to identify possible binding sites of thiamine based on the analysis of steric fitness and hydrophobicity, we first proceed with blind docking experiments with fully charged thiamine and then select most probable binding modes for further study. Section SI-6 of the Supporting Information contains a detailed description of the charge-induced changes in computationally viable thiamine–ShPrP complexes. When compared to the standard all-atom AMBER11 protein force field, the use of PM6/Mozyme protein charges contributes little in the context of the 3D-RISM-DOCK method. Therefore, we used the AMBER11 protein force field for the protein in subsequent docking experiments. All following MD studies of the dynamical stability of top-ranked shPrP thiamine complexes were also carried out using the AMBER11 force field. A summary of our docking studies with the AMBER11 force field for shPrP can be found in Table S-3 of the Supporting Information. Table S-3 also lists two different protein charges corresponding to neutral ($Q^{amb} = 0$) and acidic ($Q^{amb} = +10e$) conditions, which could be used to address pH-dependent ligand interactions.

Regardless of the net charge of the shPrP and as long as salt-free water solutions contain the neutral thiamine component, the experimental helix 1–loop 1 region of Perez-Pineiro et al.[42] persistently scores the highest. The representative summary of docking results obtained with 3D-RISM-DOCK protocol for the isoelectric shPrP is given in Figure 11. The best-scored 3D-RISM-DOCK complexes are due to the $\pi-\pi$ stacking of the pyrimidine ring with the Y150 residue in the helix 1–loop 1 region (h-1/l-1 region in Figure 1). These complexes are additionally stabilized by anchoring the hydroxyethyl group of the thiamine molecule to the nearby M138, H140, and G142 side chains. An analysis of docking results obtained for different fragmentation/linking schemes, with and without solvent-induced correction $\Delta(r) = -\sum_\sigma W_\sigma(r)$ (see section SI-6 of the Supporting Information), shows a systematic improvement of scoring of the docked conformations in the experimental binding pocket upon accounting for the ligands' geometrical constraints (higher degree of specificity for larger fragments) and desolvation penalties (gradual removal of irrelevant binding modes by enhancing the desolvation penalty). The former is consistent with the docking results for chargeless thiamine from the previous section, thus providing justification for the use of a neutral ligand for the shape complementarity study, while the latter highlights the importance of accounting for solvation effects to explain the mechanisms of thiamine binding to the prion protein, as first proposed by Pagadala et al.[10]

Although many strategies exist for the enrichment of hit rates, they are all based on the process of elimination involving a certain degree of subjectivity in defining the undesirable compounds. The initial removal of docked thiamine poses, which are believed to have a low probability of providing useful information, is performed with respect to their ranking status. Given the overall frequency with which the experimental helix 1–loop 1 region[42] appears in the present 3D-RISM-DOCK simulations, we continued our analysis of blind docking by selecting two candidates that were *globally* ranked among the top five complexes and which also belonged to the R1 region in Figure 11. These are shown in detail in Figure 12a and b as Q00G1 and Q00G5 models, respectively, with residues M138, H140, G142, and D144 given as the molecular surface.

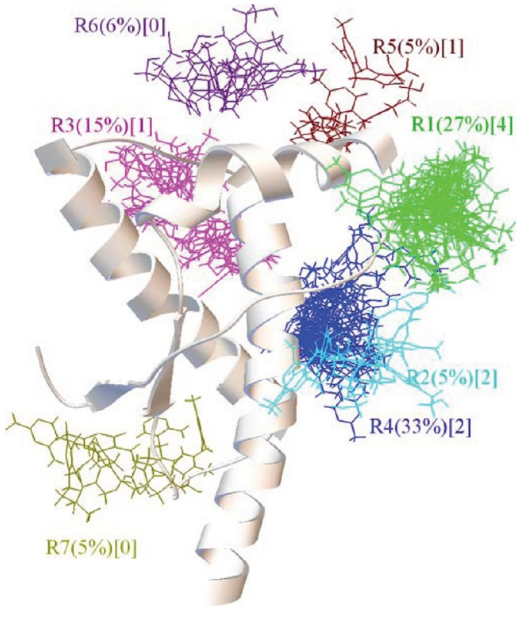

Figure 11. Clustering analysis of the results of 3D-RISM-DOCK docking experiments in a NaCl–water solution containing the isoelectric ($Q^{amb} = 0$) shPrP and 7 mM of electropositive thiamine (with q^{mop} partial atomic charges and assembled by linking nine fragments). Highly conserved water and sodium sites are included in the scoring function via $\Delta(r) = -\sum_{\sigma=37}^{39} W_\sigma(r)$. See text and Table S-2 for details.

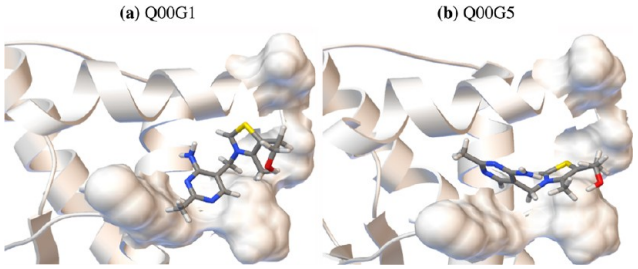

Figure 12. Two representative 3D-RISM-DOCK candidates from the unrestricted R1 region of Figure 11 that *globally* ranked first (a) and fifth (b) in the isoelectric shPrP. The thiamine molecule has been assembled by linking the nine fragments carrying Pm6 charges, q^{mop}.

To determine the efficiency with which 3D-RISM-DOCK samples the confined helix 1–loop 1 region of Perez-Pineiro et al.[42] (an experimentally observed binding pocket of thiamine), we introduced a penalty in eq 9 for the thiamine to go outside this active domain. The dimensions of the confinement region were sufficiently large to allow for an unrestricted thiamine motion across its volume. Orientational preference of the thiamine inside this confined grid was evaluated using Figure 5 of Perez-Pineiro et al.[42] Namely, a docked conformation is considered putative only if visually similar to the experiment. Thus, if the docking pose contains the pyrimidine fragment in the vicinity of M138 *and* the thiazole fragment in the proximity of H140, we consider it viable and assign it to a descriptive cluster. Two representative binding poses with lowest energy are shown in Figure 13.

The descriptor Dn<p> of every docked pose has an index "n" that ranks its binding energy and an indicator "p" for the

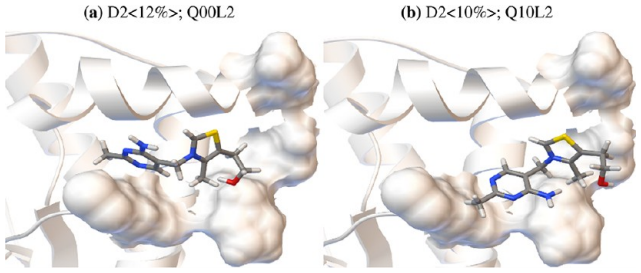

Figure 13. Two representative 3D-RISM-Dock candidates from the confined h-1/l-1 region of Perez-Pineiro et al.,[42] both *locally* ranked second in the isoelectric (a) and net positive (b) shPrP. The thiamine molecule has been assembled by linking the nine fragments carrying Pm6 charges, q^{mop}.

relative size of the host cluster, as illustrated in Figure 13 for two charge states of shPrP. The only difference between Figure 13a and b comes from the E146 residue located in the experimental helix 1−loop 1 region.[42] Namely, the E146 residue is electronegative in Figure 13a and electroneutral in Figure 13b due to protonation. Therefore, as seen earlier in Figure 7a, it is of no surprise that sodium anions are highly preserved around the E146 residue in the case of isoelectric ShPrP and eventually dissipate to bulk density for the net positive ShPrP. Compared to the Q00L2 model in Figure 13a, the reduced presence of sodium anions and subsequent rearrangement of water around the protonated E146 residue in Figure 13b are the main driving forces behind repositioning of the thiazole fragment in the Q10L2 model.

In the adopted description scheme and in terms of clustering analysis, the use of Mozyme[101] charges for the prion protein resulted in a somewhat inferior ranking compared to Amber charges, as demonstrated in Figures S-10 to S-12 of the Supporting Information. These findings stand in contrast to the results of Bikadi and Hazai,[102] which favor Pm6/Mozyme charges over default AutoDock charges. In the framework of phenomenological treatment of solvent effects, as found in the default AutoDock settings, the apparent improvements caused by Pm6/Mozyme charges may come from a pronounced electrostatic potential.[102] On the other hand, due to the rigorous treatment of solvation effects in the 3D-RISM-Dock framework, our method accurately accounts for local structural rearrangement of solvent sites in response to charge redistribution in the protein.[103,104] The real merits of using Mozyme charges for proteins within the 3D-RISM-Dock protocol remain to be benchmarked on the complete AutoDock training set and will be addressed in a subsequent publication. We therefore proceed with the evaluation of the dynamic stability of candidate binding modes found in Figures 12 and 13 against thermal fluctuations in the protein structure.

3.5. Dynamical Stability of Binding Modes. A few representative binding modes identified in the previous subsection (see also section SI-6 of the Supporting Information) were selected for further analysis. These included the global models Q00G1 and Q00G5 from Figure 12 and the local models Q00L2 and Q10L2 from Figure 13. The structures of these thiamine−prion complexes were first minimized with the implicit solvation GBSA model (see section 2.4 for details). Then, all-atom explicit solvent MD simulations were carried out with the Amber11 package using the *ff03* force field.[88,89] The dynamic stability of different binding modes was assessed based on visual inspection of MD trajectories and monitored by using the time evolution of distances between the centers of mass of the thiamine molecule and that of the binding pocket (consisting of M138−G142, N143, E146, D147, Y150, R151, and M154 residues, see Figure 1). Note that most of the residues in this binding pocket exhibit the attenuation of the NMR signal upon thiamine binding.[42] We show in Figure 14 the initial (energy minimized) structures for the two dynamically most stable models, each having a residency time in the binding pocket of about 7 ns, see Figure 15.

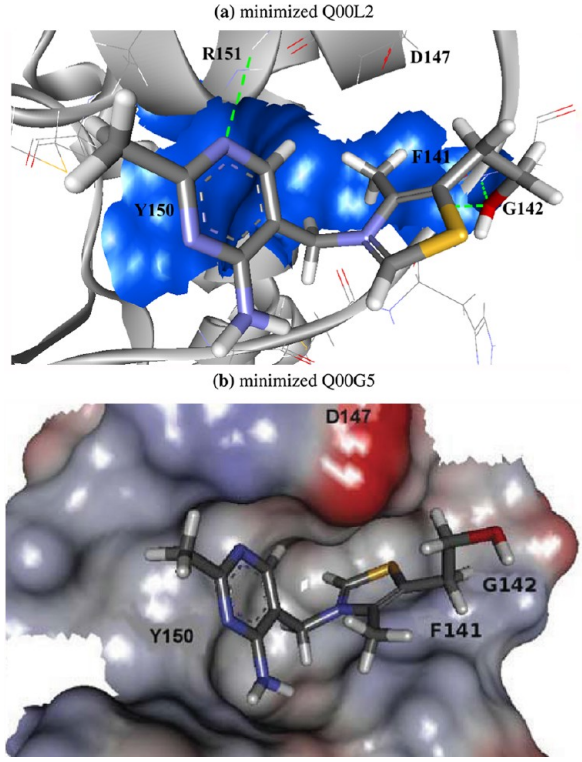

Figure 14. Thiamine in the binding pocket of prion protein. (a) The solvent exposed surfaces of Y150 and F141 are shown in blue. Hydrogen bonds between thiamine and F141, G142, and R151 are represented by dashed green lines. (b) Solvent exposed surface of residues in the binding pocket are colored according to partial charges of their atoms (red and blue colors correspond to negative and positive charges, respectively).

Both binding modes in Figure 14 are characterized by a favorable orientation of the thiamine pyrimidine ring, allowing formation of a π-stacking interface with the aromatic ring of Y150 in the course of MD simulations. In addition, the Q00L2 mode of thiamine supports the formation of hydrogen bonds with the G142 and R151 residues, see Figure 14a, such that the thiazole's methyl group forms a hydrophobic contact with the F141 residue. Note that the same methyl group is solvent exposed in the Q00G5 mode, which may contribute to its lower stability compared to the Q00L2 mode. However, the solvent exposure of the thiazole's methyl group is partially compensated by the favorable electrostatic contact between the thiazole's positively charged sulfur and nitrogen sites and the prion's negatively charged D147 residue (see Figure 14b). For both the Q00L2 and Q00G5 models, the distance between the Y150 and the pyrimidine ring of thiamine remains within the range of π-stacking interaction involving aromatic rings of these residues (see Figure 15). For the Q00L2 model, the contact between the thiazole's methyl group and F141 remains

dx.doi.org/10.1021/ct300257v | J. Chem. Theory Comput. 2012, 8, 3356−3372

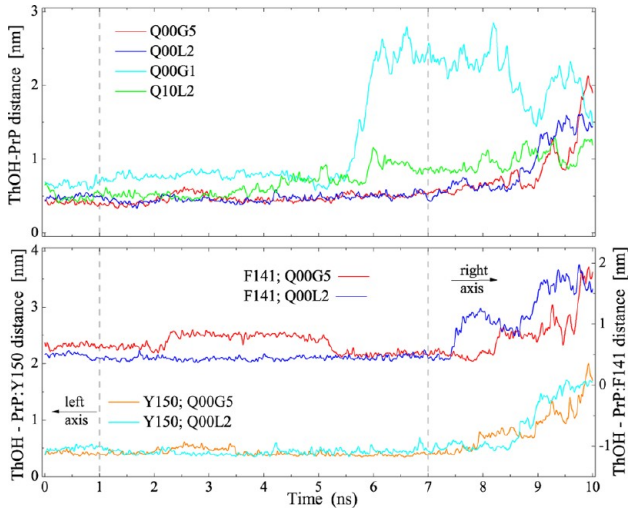

Figure 15. Time evolution for thiamine migration within the h-1/l-1 region: (top) changes in the distance between the centers of mass of thiamine and the experimental binding pocket;[42] (bottom) changes in aromatic/hydrophobic contacts between the pyrimidine ring and Y150 (left axis) and between the methyl group in thiazole and F141 (right axis). The segment of the trajectory enclosed by vertical lines is used in free energy calculations for the two dynamically most stable docked poses.

preserved throughout the course of the MD simulations of bound conformations. In the case of the Q00G5 model, the conformational changes in thiamine that occur between 5 and 6 ns lead to an optimization of hydrophobic contacts between the thiazole's methyl group and the residues in the binding pocket.

As illustrated in Figure 16, the most prominent feature of the binding pocket is the orientation of Y150 and F141 residues

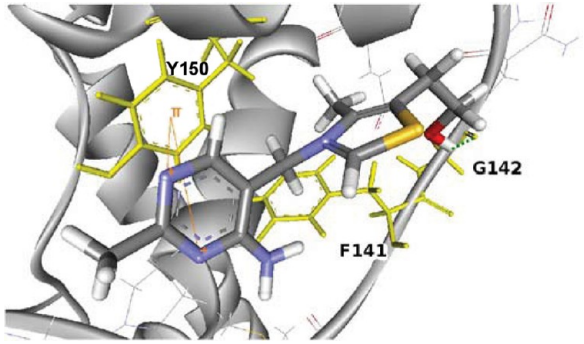

Figure 16. Snapshot of thiamine residency in the experimental binding pocket assisted by π-stacking and hydrogen bonding.

such that their aromatic rings are held at right angles to each other. This creates a favorable environment for π-stacking interactions between these residues and the pyrimidine and thiazole rings of thiamine. While thiazole's methyl group remains buried in the hydrophobic pocket throughout the course of the MD simulations, the positively charged sulfur and nitrogen sites in the same fragment engage in transient salt bridges with D147. Furthermore, the hydrogen bond between thiamine's hydroxyl group and the G142 residue remains stable for the duration of thiamine residency in the binding pocket. For both dynamically stable binding modes shown in Figure 14, the residence time of thiamine in the binding pocket is approximately 7 ns. The experimental average residence time can be estimated on the basis of the transition state theory as an inverse of the barrier crossing rate.[105] For the experimental binding affinity of 60 μM, the typical time of the barrier crossing is around 3 ns, which agrees with the results of our MD simulations.

4. SUMMARY

In summary, this work describes 3D-RISM-Dock—a novel all-atom protocol to evaluate the role of chemical specificities in molecular recognition in solutions of arbitrary complexity. This is demonstrated by the blind docking studies of thiamine against the prion protein. By introducing ligands as a part of a solvent, the conformational sampling of the solvent degrees of freedom can be carried out in terms of statistical-mechanical distribution functions that define the site-resolved PMF around the receptor. Subject to the solvent composition and ligand flexibility, these thermodynamically correct PMFs provide a quantitative estimate of binding affinities based on detailed physical contributions of hydrogen bonding, hydrophobicity, and solvation entropic effects.

In order to assess the roles of steric fitness and short-range interactions in the recognition process of thiamine by the prion protein, the most probable binding sites are initially sampled with the chargeless thiamine using the 3D-RISM-Dock approach. We then investigate the stability of the identified active sites with respect to several charge assignment methods in both the ligand and the receptor, and only the most persistent binding regions are selected for further docking analysis. Each position and orientation of thiamine with respect to the rigid prion protein is ranked according to the binding energy. A few top-ranked models are selected as input structures for explicit solvent MD simulations of the thiamine-induced conformational changes in the prion protein. The propagation of unrestrained thiamine motion within each of the candidate binding sites is monitored with respect to the total residence time during which the effective interaction energy between the thiamine and the nearby residues is sampled. This information is then used for final ranking and discrimination of thiamine−prion protein complexes and benchmarking against the experimentally obtained binding mode. The final theoretically viable binding conformation of thiamine to the prion protein is found to be in excellent agreement with the experimental data.

■ ASSOCIATED CONTENT

Supporting Information

Table S-1: The list of AutoDock4.2 files that are modified by the 3D-RISM-Dock protocol. Table S-2: Lennard-Jones potential parameters and partial atomic charges of solvent sites. Table S-3: Summary of blind docking results with respect to the thiamine flexibility and persistence of solvent sites. Figure S-1: *pH*-modulated charging data for prion protein structures. Figure S-2: Water coordination around thiazole's methyl group as a function of thiamine flexibility. Figures S-3 and S-4: Water-induced effects on 3D-RISM-Dock binding topography for the chargeless thiamine. Figure S-5: Comparison of AutoDock and 3D-RISM-Dock protocols. Figure S-6 to S-9: Solvent-induced effects on 3D-RISM-Dock binding topography for the electroneutral and positively charged thiamine. Figure S-10: The 3D-RISM-Dock orientational analysis of thiamine assembled from the nine fragments and confined to the h-1/l-1 region. Figures S-11 and S-12: The 3D-RISM-Dock orientational analysis of thiamine assembled from

the three fragments and confined to the h-1/l-1 region. This material is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*E-mail: andriy.kovalenko@nrc.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was funded by the PrioNet Canada, the Alberta Prion Research Institute, and the National Research Council of Canada. The images in Figures 1−3, 5−8, and 10−13 were generated using AutoDockTools,[106] release 1.5.6. Figures 4, 9, and 15 were prepared with Origin, release 8.1. Illustrations of the docked structures in Figures 14 and 16 were obtained with Discovery Studio Visualizer, release 3.1 (San Diego: Accelrys Software Inc., 2011).

REFERENCES

(1) McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Drug Discovery Dev.* **2007**, *11*, 494−502.

(2) Schneider, G. Virtual screening: an endless staircase? *Nat. Rev. Drug Discovery* **2010**, *9*, 273−276.

(3) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862−865.

(4) Venkatraman, V.; Pérez-Nueno, V. I.; Mavridis, L.; Ritchie, D. W. Comprehensive Comparison of Ligand-Based Virtual Screening Tools Against the DUD Data set Reveals Limitations of Current 3D Methods. *J. Chem. Inf. Model.* **2010**, *50*, 2079−2093.

(5) von Korff, M.; Freyss, J.; Sander, T. Comparison of Ligand- and Structure- Based Virtual Screening on the DUD Data Set. *J. Chem. Inf. Model.* **2009**, *49*, 209−231.

(6) Köppen, H. Virtual screening - what does it give us. *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 397−407.

(7) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358−364.

(8) Recanatini, M.; Bottegoni, G.; Cavalli, A. In silico antitarget screening. *Drug Discovery Today* **2004**, *1*, 209−215.

(9) Dill, K. A. Additivity Principles in Biochemistry. *J. Biol. Chem.* **1997**, *272*, 701−704.

(10) Pagadala, N. S.; Blinov, N.; Bjorndahl, T.;Wishart, D. S.; Kovalenko, A. Water and Ions Play Important Role in Binding of Thiamine and its Derivatives to Prion Protein. **2012**, in preparation.

(11) Kovalenko, A. In *Molecular Theory of Solvation*; Hirata, F., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003; Vol. *36*, Chapter 4. Three−Dimensional RISM Theory for Molecular Liquids and Solid−Liquid Interfaces, pp 169−275.

(12) Stumpe, M. C.; Blinov, N.; Wishart, D.; Kovalenko, A.; Pande, V. S. Calculation of Local Water Densities in Biological Systems: A Comparison of Molecular Dynamics Simulations and the 3D-RISM-KH Molecular Theory of Solvation. *J. Phys. Chem. B* **2011**, *115*, 319−328.

(13) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.* **2009**, *131*, 12430−12440.

(14) Kiyota, Y.; Yoshida, N.; Hirata, F. A New Approach for Investigating the Molecular Recognition of Protein: Toward Structure-Based Drug Design Based on the 3D-RISM Theory. *J. Chem. Theory Comput.* **2011**, *7*, 3803−3815.

(15) Muegge, I. PMF Scoring Revisited. *J. Med. Chem.* **2006**, *49*, 5895−5902.

(16) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145−1152.

(17) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: Current status and future challenges. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 15−26.

(18) Park, H.; Lee, J.; Lee, S. Critical assessment of the automated AutoDock as a new docking tool for virtual screening. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 549−554.

(19) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.

(20) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

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

(22) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual screening with AutoDock: theory and practice. *Expert Opin. Drug Discovery* **2010**, *5*, 597−607.

(23) Mezei, M.; Zhou, M. M. Dockres: a computer program that analyzes the output of virtual screening of small molecules. *Source Code Biol. Med.* **2010**, *5*, 1−6.

(24) Durrant, J. D.; Amaro, R. E.; McCammon, J. A. AutoGrow: A Novel Algorithm for Protein Inhibitor Design. *Chem. Biol. Drug Des.* **2009**, *73*, 168−178.

(25) Ördög, R.; Grolmusz, V. In *Proceedings of the 4th International Conference on Bioinformatics Research and Applications*; Măndoiu, I., Sunderraman, R., Zelikovsky, A., Eds.; Springer-Verlag: Berlin, 2008; ISBRA 2008, pp 402−413.

(26) Atilgan, E.; Hu, J. Improving Protein Docking Using Sustainable Genetic Algorithms. *Int. J. Comput., Inf., Syst., Ind. Man.* **2011**, *3*, 248−255.

(27) Atilgan, E.; Hu, J. *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*; GECCO; ACM: New York, 2010; pp 211−212.

(28) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(29) Antes, I. DynaDock: A new molecular dynamics-based algorithm for protein−peptide docking including receptor flexibility. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1084−1104.

(30) de Magalhães, C. S.; Barbosa, H. J. C.; Dardenne, L. E. A genetic algorithm for the ligand−protein docking problem. *Genet. Mol. Biol.* **2004**, *27*, 605−610.

(31) Thomsen, R. Flexible ligand docking using evolutionary algorithms: investigating the effects of variation operators and local search hybrids. *BioSystems* **2003**, *72*, 57−73.

(32) Buzko, O. V.; Bishop, A. C.; Shokat, K. M. Modified AutoDock for accurate docking of protein kinase inhibitors. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 113−127.

(33) Norgan, A.; Coffman, P.; Kocher, J.-P.; Katzmann, D.; Sosa, C. Multilevel Parallelization of AutoDock 4.2. *J. Cheminf.* **2011**, *3*, 12−20.

(34) Collignon, B.; Schulz, R.; Smith, J. C.; Baudry, J. Task-parallel message passing interface implementation of Autodock4 for docking of very large databases of compounds using high-performance super-computers. *J. Comput. Chem.* **2011**, *32*, 1202−1209.

(35) Prakhov, N. D.; Chernorudskiy, A. L.; Gainullin, M. R. VSDocker: a tool for parallel high-throughput virtual screening using AutoDock on Windows-based computer clusters. *Bioinformatics* **2010**, *26*, 1374−1375.

(36) Zhang, S.; Kumar, K.; Jiang, X.; Wallqvist, A.; Reifman, J. DOVIS: an implementation for high-throughput virtual screening using AutoDock. *BMC Bioinf.* **2008**, *9*, 126−129.

(37) Jiang, X.; Kumar, K.; Hu, X.; Wallqvist, A.; Reifman, J. DOVIS 2.0: an efficient and easy to use parallel virtual screening tool based on AutoDock 4.0. *Chem. Cent. J.* **2008**, *2*, 1−7.
(38) Khodade, P.; Prabhu, R.; Chandra, N.; Raha, S.; Govindarajan, R. Parallel implementation of AutoDock. *J. Appl. Crystallogr.* **2007**, *40*, 598−599.
(39) Thormann, M.; Pons, M. Massive docking of flexible ligands using environmental niches in parallelized genetic algorithms. *J. Comput. Chem.* **2001**, *22*, 1971−1982.
(40) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. The human cerebrospinal fluid metabolome. *J. Chromatogr. B* **2008**, *871*, 164−173.
(41) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, *37*, D603−D610.
(42) Perez-Pineiro, R.; Bjorndahl, T. C.; Berjanskii, M. V.; Hau, D.; Li, L.; Huang, A.; Lee, R.; Gibbs, E.; Ladner, C.; Dong, Y. W.; Abera, A.; Cashman, N. R.; Wishart, D. S. The prion protein binds thiamine. *FEBS J.* **2011**, 1−13.
(43) Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.
(44) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*, 3rd ed.; Academic Press; Elsevier Academic Press: Waltham, MA, 2006; Chapter IV: Distribution-Function Theories, pp 97−103.
(45) Chandler, D.; Andersen, H. C. Optimized Cluster Expansions for Classical Fluids. II. Theory of Molecular Liquids. *J. Chem. Phys.* **1972**, *57*, 1930−1937.
(46) Chandler, D.; McCoy, J. D.; Singer, S. J. Density functional theory of nonuniform polyatomic systems. I. General formulation. *J. Chem. Phys.* **1986**, *85*, 5971−5976.
(47) Chandler, D.; McCoy, J. D.; Singer, S. J. Density functional theory of nonuniform polyatomic systems. II. Rational closures for integral equations. *J. Chem. Phys.* **1986**, *85*, 5977−5982.
(48) Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.
(49) Kovalenko, A.; Hirata, F. Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *J. Chem. Phys.* **2000**, *112*, 10391−10402.
(50) Kovalenko, A.; Hirata, F. Potentials of mean force of simple ions in ambient aqueous solution. II. Solvation structure from the three-dimensional reference interaction site model approach, and comparison with simulations. *J. Chem. Phys.* **2000**, *112*, 10403−10417.
(51) Kovalenko, A.; Hirata, F. Self-consistent description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999**, *110*, 10095−10112.
(52) Perkyns, J. S.; Pettitt, B. M. A dielectrically consistent interaction site theory for solvent-electrolyte mixtures. *Chem. Phys. Lett.* **1992**, *190*, 626−630.
(53) Perkyns, J. S.; Pettitt, B. M. A site-site theory for finite concentration saline solutions. *J. Chem. Phys.* **1992**, *97*, 7656−7666.
(54) Rosenfeld, Y.; Ashcroft, N. W. Theory of simple classical fluids: Universality in the short-range structure. *Phys. Rev. A* **1979**, *20*, 1208−1235.
(55) Lee, L. L. In *Molecular Thermodynamics of Nonideal Fluids*; Brenner, H., Ed.; Butterworths series in chemical engineering; Butterworth-Heinemann: Boston, MA, 1988; Chapter VI: Integral Equation Theories, pp 133−184.
(56) Martynov, G. A. *Fundamental Theory of Liquids: Method of Distribution Functions*; Adam Hilger: New York, 1992; Chapter V: Fundamental Equations of Statistical Mechanics, pp 122−156.
(57) Duh, D. M.; Haymet, A. D. J. Integral equation theory for uncharged liquids: The Lennard-Jones fluid and the bridge function. *J. Chem. Phys.* **1995**, *103*, 2625−2633.
(58) Monson, P. A.; Morriss, G. P. In *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; JohnWiley & Sons: Hoboken, NJ, 2007; Vol. *77*; Chapter VIII: Recent Progress in the Statistical Mechanical Mechanics of Interaction Site Fluids, pp 451−550.
(59) Singer, S. J.; Chandler, D. Free energy functions in the extended RISM approximation. *Mol. Phys.* **1985**, *55*, 621−625.
(60) Lu, S. Y.; Jiang, Y. J.; Lv, J.; Zou, J. W.; Wu, T. X. Role of bridging water molecules in GSK3β-inhibitor complexes: Insights from QM/MM, MD, and molecular docking studies. *J. Comput. Chem.* **2011**, *32*, 1907−1918.
(61) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 34−40.
(62) Olson, A. J.; Goodsell, D. S. Automated Docking and the Search for HIV Protease Inhibitors. *SAR QSAR Environ. Res.* **1998**, *8*, 273−285.
(63) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3DRISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114*, 8505−8516.
(64) McQuarrie, D. A. *Statistical Mechanics*, 2nd ed.; University Science Books: Sausalito, CA, 2000; Chapter XIII: Distribution Functions In Classical Monoatomic Liquids, p 266.
(65) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein−ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 808−813.
(66) Snyder, P. W.; Mecinović, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Héroux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17889−17894.
(67) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293−304.
(68) Hetenyi, C.; van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2002**, *11*, 1729−1737.
(69) Michalewicz, Z.; Fogel, D. B. *How to Solve It: Modern Heuristics*, 2nd ed.; Springer: Berlin, Germany, 2004; Chapter VII: Designing Evolutionary Algorithms, pp 161−184.
(70) Hart, W. E.; Rosin, C.; Belew, R. K.; Morris, G. M. In *Improved Evolutionary Hybrids for Flexible Ligand Docking in AutoDock*; Floudas, C. A., Pardalos, P. M., Eds.; Optimization in Computational Chemistry and Molecular Biology; Kluwer Academic Publishers B. V.: Dordrecht, The Netherlands, 2000; pp 209−230.
(71) James, T. L.; Liu, H.; Ulyanov, N. B.; Farr-Jones, S.; Zhang, H.; Donne, D. G.; Kaneko, K.; Groth, D.; Mehlhorn, I.; Prusiner, S. B.; Cohen, F. E. Solution structure of a 142-residue recombinant prion protein corresponding to the in fectious fragment of the scrapie isoform. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10086−10091.
(72) Lima, L. M. T. R.; Cordeiro, Y.; Tinoco, L. W.; Marques, A. F.; Oliveira, C. L. P.; Sampath, S.; Kodali, R.; Choi, G.; Foguel, D.; Torriani, I.; Caughey, B.; Silva, J. L. Structural Insights into the Interaction between Prion Protein and Nucleic Acid. *Biochemistry* **2006**, *45*, 9180−9187.
(73) Kuwata, K.; Nishida, N.; Matsumoto, T.; Kamatari, Y. O.; Hosokawa-Muto, J.; Kodama, K.; Nakamura, H. K.; Kimura, K.; Kawasaki, M.; Takakura, Y.; Shirabe, S.; Takata, J.; Kataoka, Y.; Katamine, S. Hot spots in prion protein for pathogenic conversion. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11921−11926.

(74) Louloudi, M.; Hadjiliadis, N. Structural aspects of thiamine, its Derivatives and their metal complexes in relation to the enzymatic action of thiamine enzymes. *Coord. Chem. Rev.* **1994**, *135*−*136*, 429−468.
(75) Stewart, J. Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009**, *15*, 765−805.
(76) Jordan, F. Semiempirical calculations on the electronic structure and preferred conformation of thiamine (vitamin B1) and thiamine pyrophosphate (cocarboxylase). *J. Am. Chem. Soc.* **1974**, *96*, 3623−3630.
(77) Turano, A.; Pletcher, J.; Furey, W.; Sax, M. Charge Density Analysis Of Thiamine. *Ann. N.Y. Acad. Sci.* **1982**, *378*, 91−106.
(78) Shin, W.; Oh, D. G.; Chae, C. H.; Yoon, T. S. Conformational analyses of thiamin-related compounds. A stereochemical model for thiamin catalysis. *J. Am. Chem. Soc.* **1993**, *115*, 12238−12250.
(79) DuPre, D. B.; Wong, J. L. Topological Analysis of the Electron Density in Model Azolium Systems for Thiamin Structure-Function: Sulfur Is the Electron Sink and Positively Polarized Carbanions Act as Nucleophiles. *J. Phys. Chem. A* **2005**, *109*, 7606−7612.
(80) Friedmann, R.; Uslar, W. Biochemistry and Physiology of Thiamin Diphosphate Enzymes. In *Proceedings of the International Meeting on the Function of Thiamin Diphosphate Enzymes*; Bisswanger, H., Ullrich, J., Eds.; Verlag Chemie: Weinheim, Germany, 1990; pp 27−33.
(81) Kozik, A. Current Advances in Buckwheat Research. *Proceedings of the 6th International Symposium on Buckwheat*; Matano, T., Ujihara, A., Eds.; Shinshu University Press: Shinshu, Japan, 1995; Chapter Thiamin-Binding Protein from Buckwheat Seeds: Some Molecular Properties, Ligand-Protein Interaction, Bioanalytical Applications, pp 823−831.
(82) Soriano, E. V.; Rajashankar, K. R.; Hanes, J. W.; Bale, S.; Begley, T. P.; Ealick, S. E. Structural Similarities between Thiamin−Binding Protein and Thiaminase-I Suggest a Common Ancestor. *Biochemistry* **2008**, *47*, 1346−1357.
(83) Sippel, K. H.; Venkatakrishnan, B.; Boehlein, S. K.; Sankaran, B.; Quirit, J. G.; Govindasamy, L.; Agbandje-McKenna, M.; Goodison, S.; Rosser, C. J.; McKenna, R. Insights into Mycoplasma genitalium metabolism revealed by the structure of MG289, an extracytoplasmic thiamine binding lipoprotein. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 528−536.
(84) Hanes, J. W.; Chatterjee, D.; Soriano, E. V.; Ealick, S. E.; Begley, T. P. Construction of a thiamin sensor from the periplasmic thiamin binding protein. *Chem. Commun.* **2011**, *47*, 2273−2275.
(85) Paramasivam, S.; Balakrishnan, A.; Dmitrenko, O.; Godert, A.; Begley, T. P.; Jordan, F.; Polenova, T. Solid−State NMR and Density Functional Theory Studies of Ionization States of Thiamin. *J. Phys. Chem. B* **2011**, *115*, 730−736.
(86) Ciulli, A.; Blundell, T. L.; Abell, C. In *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*; Stroud, R. M., Finer−Moore, J., Eds.; RSC Biomolecular Sciences; Royal Society of Chemistry Publishing: London, 2008; Chapter XVI: Discovery and Extrapolation of Fragment Structures towards Drug Design, pp 293−319.
(87) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber11*; University of California: San Francisco, CA, 2010.
(88) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(89) Lee, M. C.; Duan, Y. Distinguish protein decoys by Using a scoring function based on a new Amber force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 620−634.
(90) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing Term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(91) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science Publications; Clarendon Press: New York, 1989.
(92) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(93) Wittorf, J. H.; Gubler, C. J. Coenzyme Binding in Yeast Pyruvate Decarboxylase. *Eur. J. Biochem.* **1970**, *14*, 53−60.
(94) Gutowski, J. A.; Lienhard, G. E. Transition state analogs for thiamin pyrophosphate-dependent enzymes. *J. Biol. Chem.* **1976**, *251*, 2863−2866.
(95) O'Brien, T. A.; Gennis, R. B. Studies of the thiamin pyrophosphate binding site of Escherichia coli pyruvate oxidase. Evidence for an essential tryptophan residue. *J. Biol. Chem.* **1980**, *255*, 3302−3307.
(96) Ullrich, J. Structure-Function Relationships In Pyruvate Decarboxylase Of Yeast And Wheat Germ. *Ann. N.Y. Acad. Sci.* **1982**, *378*, 287−305.
(97) Hawkins, C. F.; Borges, A.; Perham, R. N. A common structural motif in thiamin pyrophosphate-binding enzymes. *FEBS Lett.* **1989**, *255*, 77−82.
(98) Guo, F.; Zhang, D.; Kahyaoglu, A.; Farid, R. S.; Jordan, F. Is a Hydrophobic Amino Acid Required To Maintain the Reactive V Conformation of Thiamin at the Active Center of Thiamin Diphosphate-Requiring Enzymes? Experimental and Computational Studies of Isoleucine 415 of Yeast Pyruvate Decarboxylase. *Biochemistry* **1998**, *37*, 13379−13391.
(99) Thomas, A. A.; De Meese, J.; Le Huerou, Y.; Boyd, S. A.; Romoff, T. T.; Gonzales, S. S.; Gunawardana, I.; Kaplan, T.; Sullivan, F.; Condroski, K.; Lyssikatos, J. P.; Aicher, T. D.; Ballard, J.; Bernat, B.; DeWolf, W.; Han, M.; Lemieux, C.; Smith, D.; Weiler, S.; Wright, S. K.; Vigers, G.; Brandhuber, B. Non-charged thiamine analogs as inhibitors of enzyme transketolase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 509−512.
(100) Kluger, R. Thiamin diphosphate: a mechanistic update on enzymic and nonenzymic catalysis of decarboxylation. *Chem. Rev.* **1987**, *87*, 863−876.
(101) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(102) Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminf.* **2009**, *1*, 15.
(103) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 804−813.
(104) Hirano, K.; Yokogawa, D.; Sato, H.; Sakaki, S. An Analysis of 3D Solvation Structure in Biomolecules: Application to Coiled Coil Serine and Bacteriorhodopsin. *J. Phys. Chem. B* **2010**, *114*, 7935−7941.
(105) Hills, R. D., Jr.; Brooks, C. L., III. Hydrophobic Cooperativity as a Mechanism for Amyloid Nucleation. *J. Mol. Biol.* **2007**, *368*, 894−901.
(106) Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graphics Modell.* **1999**, *17*, 57−61.
(107) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242−2254.
(108) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J. Mol. Biol.* **2007**, *371*, 1118−1134.

(109) Parkin, G. Valence, Oxidation Number, and Formal Charge: Three Related but Fundamentally Different Concepts. *J. Chem. Educ.* **2006**, *83*, 791−799.
(110) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(111) Stewart, J. J. P. MOPAC: A General Molecular Orbital Package (Version 6.0). *Quant. Chem. Program Exchange* **1990**, *455*, 1−1.
(112) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(113) Tsai, K.-C.; Wang, S.-H.; Hsiao, N.-W.; Li, M.; Wang, B. The effect of different electrostatic potentials on docking accuracy: A case study using DOCK5.4. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3509−3512.
(114) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(115) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(116) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity−a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(117) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(118) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(119) Stewart, J. J. P. Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *J. Comput. Chem.* **1991**, *12*, 320−341.
(120) Stewart, J. J. P. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *J. Mol. Model.* **2004**, *10*, 155−164.
(121) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101−1111.
(122) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein−Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.
(123) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368−W371.
(124) Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 928−938.
(125) Bashford, D.; Karplus, M. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **1990**, *29*, 10219−10225.
(126) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735−1747.
(127) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical electrostatics for biomolecules: Beyond the generalized-Born approximation. *J. Chem. Phys.* **2006**, *124*, 124902.
(128) Olsson, M. Improving the desolvation penalty in empirical protein pK_a modeling. *J. Mol. Model.* **2011**, 1−10.
(129) Gilson, M. K. Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins: Struct., Funct., Bioinf.* **1993**, *15*, 266−282.
(130) Onufriev, A. In *Modeling Solvent Environments: Applications to Simulations of Biomolecules*; Feig, M., Ed.; Wiley-VCH: Weinheim, Germany, 2010; Chapter VI: Continuum Electrostatics Solvent Modeling with the Generalized Born Model, pp 127−166.
(131) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25*, 265−284.
(132) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(133) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.
(134) Sproviero, E.; Newcomer, M.; Gascón, J. A.; Batista, E.; Brudvig, G.; Batista, V. The MoD-QM/MM methodology for structural refinement of photosystem II and other biological macromolecules. *Photosynth. Res.* **2009**, *102*, 455−470.
(135) Zuegg, J.; Bliznyuk, A. A.; Gready, J. E. Comparison of electrostatic potential around proteins calculated from Amber and AM1 charges: application to mutants of prion protein. *Mol. Phys.* **2003**, *101*, 2437−2450.
(136) Ponder, J. W.; Case, D. A. In *Protein Simulations*; Daggett, V., Ed.; Advances in Protein Chemistry; Academic Press: San Diego, 2003; Vol. *66*; Chapter II: Force Fields for Protein Simulations, pp 27−85.
(137) Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.
(138) Fuhrmann, J.; Rurainski, A.; Lenhof, H. P.; Neumann, D. A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J. Comput. Chem.* **2010**, *31*, 1911−1918.