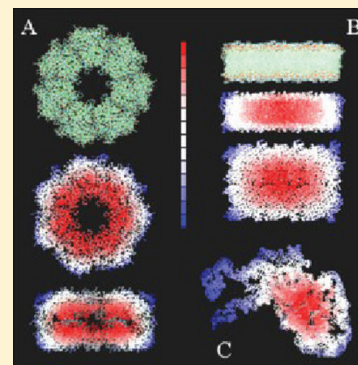


# Water-Exclusion and Liquid-Structure Forces in Implicit Solvation

Sergio A. Hassan\* and Peter J. Steinbach

Center for Molecular Modeling, DCB/CIT, National Institutes of Health, U.S. DHHS, Bethesda, Maryland 20892, United States

**ABSTRACT:** A continuum model of solvation is proposed to describe (i) long-range electrostatic effects of water exclusion resulting from incomplete and anisotropic hydration in crowded environments and (ii) short-range effects of liquid-structure forces on the hydrogen-bond interactions at solute/water interfaces. The model is an extension of the phenomenological screened coulomb potential-based implicit model of solvation. The developments reported here allow a more realistic representation of highly crowded and spatially heterogeneous environments, such as those in the interior of a living cell. Only the solvent is treated as a continuum medium. It is shown that the electrostatic effects of long-range water-exclusion can strongly affect protein–protein binding energies and are then related to the thermodynamics of complex formation. Hydrogen-bond interactions modulated by the liquid structure at interfaces are calibrated based on systematic calculations of potentials of mean force in explicit water. The electrostatic component of the model is parametrized for monovalent, divalent and trivalent ions. The conceptual and practical aspects of the model are discussed based on simulations of protein complexation and peptide folding. The current implementation is  $\sim 1.5$  times slower than the gas-phase force field and exhibits good parallel performance.



## 1. INTRODUCTION

Living cells contain high concentrations of solutes,<sup>1–3</sup> including macromolecular complexes and assemblies that displace large amounts of water, thus removing part of the highly polar/polarizable medium that modulates electrostatic interactions. These water-exclusion effects depend mainly on the size and morphology of the water-excluding solutes. The internal components of the cell also form extensive solute/liquid interfaces; thus much of cell water is indeed interfacial.<sup>2</sup> Aqueous interfaces display nonbulk behavior and can extend into the bulk a few Angstroms or several nanometers, depending on the surface size, topography, and charge distribution.<sup>4–9</sup> The behavior of molecules immersed in such regions can differ from those in the bulk solution as well.<sup>10–16</sup> As is the case in many surface-mediated phenomena, the microscopic origin of these effects is not well understood. Both electrostatic and liquid-structure effects at the interface have been invoked as possible mechanisms.

The amount of water in the interior of living cells that displays nonbulk behavior is still controversial,<sup>2,17,18</sup> and many experiments have been conducted to elucidate it. Magnetic relaxation dispersion data<sup>4,19</sup> suggest that only water in the first hydration shell is substantially modified by the solutes (up to  $\sim 20\%$  of the cell water content). In contrast, X-ray and neutron scattering,<sup>20,21</sup> as well as diffusion data of endogenous probes,<sup>22–24</sup> suggest a much larger fraction of nonbulk water (up to  $\sim 75\%$  in some cell types). Given the heterogeneous nature of the subcellular environment, it is likely that the cell contains regions of water with substantial nonbulk behavior (e.g., in the intervening space between nearby proteins, or in confined spaces) and regions with bulk-like behavior (less crowded regions in the cytoplasm). Atomistic dynamics simulations show that small solutes can induce long-range water structure in the intervening space between

polar or charged groups,<sup>25</sup> a property likely to be amplified in the crowded and heterogeneous cellular milieu.

Regardless of the overall proportion of bulk/nonbulk water suggested by experiments, at the molecular level most biology ultimately happens at interfaces. Moreover, evidence suggests<sup>2</sup> that about half of biologically active molecules in the cell do not diffuse freely in the bulk medium despite their solubility, but are transiently bound, either specifically or nonspecifically, to one another, to membranes, or to the cytoskeleton. Therefore, a model of solvation for application in crowded environments requires special consideration of at least two main effects: the long-range electrostatic modulation resulting from incomplete and anisotropic hydration owing to the presence of water-excluding solutes, and the short-range forces induced by interfacial water, in particular the forces induced by the structure of hydration shells in close proximity. Several simplified models of solvation, known as continuum or implicit solvent models (ISMs) have been proposed for simulations of macromolecules<sup>26–33</sup> that can speed up computation by several orders of magnitude in some cases. Millisecond-long dynamics simulations of small proteins using a fully atomistic model of water are now possible with special-purpose computer architecture.<sup>34,35</sup> This implies that increasingly realistic cellular environments could soon be simulated over biological time scales, provided that a computationally efficient, complete ISM is available. ISMs commonly used in macromolecular simulation focus mainly on macroscopic electrostatic effects. They do not explicitly address the problems posed by interfaces, neither dielectric nor structural effects,

Received: August 24, 2011

Revised: October 12, 2011

Published: October 18, 2011



and water-exclusion effects are not given particular consideration or are treated in rather simplified ways. Although numerical solutions of Poisson equation do account for water-exclusion effects quite naturally, they are computationally demanding, especially for large systems.

In this paper the screened Coulomb potential-based implicit solvent model (SCPISM)<sup>36–39</sup> is extended to account for the forces induced by exclusion of water from the environment of a solute and the forces induced by the structure of water in the hydration shells of closely interacting solutes. The conceptual aspects of the model and its computational efficiency are discussed based on results from simulations of a peptide and a protein complex. Emphasis is on the physical meaning of the quantities that define the model.

## II. ELECTROSTATIC EFFECTS IN THE SCPISM

Electrostatic modulation by an aqueous solution originates in the structural and dynamic response of water molecules to the electric field created by the solute's charge distribution. Bulk water is relatively simple to characterize and can be viewed as the reorientation and polarization of the water molecules subject to thermal fluctuations.<sup>40</sup> Closer to the solute the response is more complicated due to the competition between water–water and water–solute interactions. When compared to the bulk phase, these competing forces lead to changes in the structure and dynamics of the interfacial water, and bulk behavior is recovered only at a certain distance from the surface. These microscopic changes in the water response lead to two macroscopic (thermodynamic) effects:<sup>39,41</sup> local variations in the water density and local changes in the dielectric response. Therefore, at least two basic quantities are necessary in a continuum electrostatic model of solvation: a position-dependent water density,  $\rho(\mathbf{r})$ , and a position-dependent scalar dielectric permittivity,  $\epsilon(\mathbf{r})$ . A model based on these local properties has been proposed,<sup>39</sup> which is valid for any polar/polarizable solvent and solutes of arbitrary shape and charge distribution. A simplified expression for  $\epsilon(r)$  based on the Lorentz–Debye–Sack (LDS) approximation was obtained previously for a point charge  $q$  in a homogeneous fluid, which is given by<sup>42,43</sup>

$$\epsilon(r) = 1 + \frac{4\pi}{3}[\epsilon(r) + 2]\frac{\alpha}{v} + \frac{4\pi\mu}{qv}r^2\epsilon(r)L\left\{\frac{q\mu[\epsilon(r) + 2]}{[3kT + \mu R(r)]r^2\epsilon(r)}\right\} \quad (1)$$

where  $v$ ,  $\alpha$ , and  $\mu$  are the molecular volume (inverse of the number density), polarizability, and dipole moment of the solvent respectively,  $T$  is the temperature,  $r$  is the distance from the charge, and  $R$  is the Onsager reaction field. The function  $L$  is the Langevin function obtained from the Boltzmann average over all molecular orientations.<sup>42,43</sup> The relationship between the reaction field  $R$  and other electric fields commonly used in classical dielectric theory, such as the directing field, the internal (microscopic) field, the Maxwell (macroscopic) field, and the vacuum, cavity, and Debye and Lorentz fields, and their roles in determining  $\epsilon(r)$  have been discussed in detail.<sup>39</sup> Numerical solutions of eq 1 show that  $\epsilon$  is sigmoidal in  $r$  due to saturation effects close to the source, with an asymptotic value  $\epsilon_0$  at large  $r$ ;  $\epsilon_0$  is the static dielectric permittivity of the bulk solvent at the given temperature. In the limit of weak electric fields, or large  $r$  in this case, eq 1 converges to the Onsager equation,<sup>39</sup> which relates  $\epsilon_0$

and  $\mu^2$ . In the limit of high frequencies, eq 1 converges to the Lorentz–Lorenz relationship<sup>39</sup> between  $\alpha$  and the optical permittivity  $\epsilon_\infty$ . Several expressions similar to eq 1 have been proposed<sup>44–46</sup> that differ in the rate of increase of  $\epsilon$  with  $r$ , depending on the assumptions of the model. The dielectric screenings used in the SCPISM are based on solutions of eq 1. Since  $\epsilon$  depends on  $q$ , or more generally on the macroscopic field  $E$ ,<sup>39</sup> using eq 1 as the basis of an ISM has potential advantages.<sup>36,38</sup> The screening can be adjusted if  $q$  changes, e.g., when a polarizable force field is used, or when protonation states change upon  $\text{pK}_a$  shifts. The dependence of  $\epsilon_0$  with  $T$ , as well as the values of  $\alpha$  and  $\mu$  are known for most solvents, facilitating the application of the model to different solvents and temperatures. Finally, the basis of the model is essentially the same as those used in ISMs for quantum chemical calculations,<sup>47</sup> and both QM and MM methods of solvation could eventually be unified in a common framework.

In a solvent, the dielectric function  $\epsilon$  damps the vacuum electric field  $E_0$ , while the screening function  $D$  modulates the vacuum potential  $\phi_0$ . The relation between  $\epsilon$  and  $D$  is given by the definition of electric potential,<sup>48</sup> i.e.,  $E_0(r)/\epsilon(r) = -\nabla[\phi_0(r)/D(r)]$  for spherical symmetry, or explicitly by  $r \, dD(r)/dr = D(r)[D(r) - \epsilon(r)]/\epsilon(r)$ . The screening  $D$  is also sigmoidal in  $r$  with the same asymptotic value  $\epsilon_0$  at large distances.<sup>38</sup> For isolated charges, calculations show<sup>38</sup> that one particular function (Boltzmann sigmoidal), given by<sup>43</sup>

$$D(r) = \frac{(1 + \epsilon_0)}{1 + k \exp(-\alpha r)} - 1 \quad (2)$$

with  $k = (\epsilon_0 - 1)/2$ , allows a good fit of  $D(r)$  for several polar liquids using the single screening parameter  $\alpha$  (not to be confused with the polarizability  $\alpha$  in eq 1). This parameter reflects all the physics of the system embedded in eq 1, including the charge  $q$ , the properties of the solvent, and the temperature. Equation 2 is a solution of<sup>49</sup>  $dD(r)/dr = \alpha[1 + D(r)][\epsilon_0 - D(r)]/(1 + \epsilon_0)$ , which provides algebraic solutions for  $D(r)$  when  $\epsilon(r)$  is known from theory (e.g., eq 1) or experiments.

### II.1. Total Electrostatic Energy and Hydration Free Energy.

The total electrostatic energy  $E_e$  of a molecule in a solvent can be partitioned into two terms,  $E_e = E_{\text{int}} + E_{\text{self}}$  where  $E_{\text{int}}$  is the interaction energy and  $E_{\text{self}}$  is the self-energy. Because this is a phenomenological proposition, any model based on this partitioning is necessarily phenomenological. Except in numerical solutions of the Poisson equation, this approach is the basis of most solvent models currently used in simulations of macromolecules, and is also used in the SCPISM. The electrostatic contribution to the Helmholtz free energy of solvation is given by<sup>50</sup>  $\Delta F_e = \int \mathbf{E} \cdot \mathbf{D} - \int \mathbf{E}_0 \cdot \mathbf{E}_0$ , where  $\mathbf{E}$  and  $\mathbf{D}$  are the electric and the displacement fields, and the integration is over all space. For a single charge  $q$ , the simplest form of  $\Delta F_e$  is given by the original Born model<sup>51</sup> as  $\Delta F_e = -q^2(1 - 1/\epsilon_0)/2R$ . Subsequent studies of ion hydration showed that the Born radius can be expressed as  $R = R_I + \delta_\pm$  where  $R_I$  is the ionic radius and  $\delta$  is an extension, different for anions and cations.<sup>52,53</sup> With this interpretation,  $\Delta F_e$  could be reproduced for ions of different charges ( $-1e$  to  $+3e$ ) and sizes without the need to empirically adjust  $R$  in each case. It was also shown<sup>53–55</sup> that saturation effects of the electric field had to be incorporated to reproduce both the enthalpy and the entropy of ion hydration. Thus, a more general form of  $\Delta F_e$  can be obtained for a fully solvated charge  $q$ , in

the form<sup>36,48</sup>

$$\Delta F_e = \frac{q^2}{2R(q)} \left\{ \frac{1}{D[q, R(q)]} - 1 \right\} \quad (3)$$

which depends on the screening  $D$  instead of the permittivity  $\epsilon$  and where both  $R$  and  $D$  depend on the magnitude of the charge  $q$ , and  $D$  also depends explicitly on  $R$ . The dependence of  $R$  on  $q$ , instead of its sign only, allows for partial charges, which are common in classical force fields. The SCPISM takes eq 3 as the self-energy of a partially solvated atom  $i$  in a protein, but with an effective radius  $R_i$ <sup>36,37</sup> and an effective screening  $D_i$  dependent on the protein structure, i.e.

$$\Delta F_{e,i} = \frac{q_i^2}{2R_i(q_i, \mathbf{r})} \left\{ \frac{1}{D_i[q_i, R_i(q_i, \mathbf{r}), \mathbf{r}]} - 1 \right\} \quad (4)$$

where  $\mathbf{r} \equiv \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$  indicates the explicit dependence of  $R$  and  $D$  on the coordinates of all the atoms in the protein. The total electrostatic energy of a solvated protein is then calculated as

$$E_e = \frac{1}{2} \sum_{i \neq j}^N \frac{q_i q_j}{r_{ij} D_{ij}(q_i, q_j, r_{ij}, \mathbf{r})} + \frac{1}{2} \sum_{i=1}^N \frac{q_i^2}{R_i(q_i, \mathbf{r})} \left\{ \frac{1}{D_i[q_i, R_i(q_i, \mathbf{r}), \mathbf{r}]} - 1 \right\} \quad (5)$$

where the first term is a sum of pairwise Coulomb interactions screened by effective screenings  $D_{ij}$  and the second term is a sum of the self-energies given by eq 4,  $r_{ij}$  is the distance between atoms  $i$  and  $j$ , and  $N$  is the total number of atoms in the system. The first summation in eq 5 contains  $N(N-1)/2$  screening functions  $D_{ij}$ , each depending explicitly on  $r_{ij}$  and on the positions  $\mathbf{r}$  of all other atoms in the system. Each function  $D_{ij}$  also depends on the charges  $q_i$  and  $q_j$ , although this dependence is implicit through the corresponding screening parameters  $\alpha_i$  and  $\alpha_j$  defined in eq 2. The functions  $D_{ij}$  are expected to depend not only on  $q_i$  and  $q_j$ , but also on all other charges in the system, i.e.,  $\mathbf{q} \equiv \{q_1, q_2, \dots, q_N\}$ . This is so because the dielectric response of the liquid, which is reflected in  $\epsilon$  and determines the total energy, depends on the macroscopic field  $\mathbf{E}$  generated by the solute, not only on the field generated by the two charges.<sup>39</sup> This generalization, however, is not introduced in the current model. Equation 5 also contains  $N$  screening functions  $D_i$  in the second term. As discussed above, these functions depend explicitly on  $R_i$  and  $\mathbf{r}$ , and implicitly on  $q_i$  through  $\alpha_i$ ; here also a general dependence on  $\mathbf{q}$  is expected. Finally, the effective radii  $R_i$  in eq 5 depend explicitly on both  $\mathbf{r}$  and  $q$ . These radii appear only in the self-energy term and play no role in the interaction term. The dielectric permittivity is the only physical quantity that affects both interaction and self-energy. Although phenomenological, the form  $E_e$  given by eq 5 contains all the basic ingredients of a complete electrostatic model of solvation. Nonlocal effects<sup>56</sup> and other electrostatic effects<sup>39</sup> known to affect the behavior of water at interfaces are not introduced explicitly in eq 5. In the current development, however, such effects are expected to be accounted for by a suitable parametrization, mainly by the behavior of  $D$ 's at the solute/solvent interface.<sup>39</sup>

The transfer free energy of a solute from the bulk of a solvent  $a$  to the bulk of a solvent  $b$  is obtained from eq 5 in each solvent, as  $\Delta F_e^{a \rightarrow b} = E_{e,b} - E_{e,a}$ . If  $a$  is vacuum and  $b$  is water,  $\Delta F_e^{a \rightarrow b}$  is the electrostatic hydration energy  $\Delta F_e$ , and  $E_e$ 's are "effective" energies containing the potential energy of the solute and the free energy of the solvent.<sup>30</sup> A simplified form of  $\Delta F_e$  is used in the SCPISM, namely

$$\Delta F_e = \frac{1}{2} \sum_{i \neq j}^N \frac{q_i q_j}{r_{ij}} \left( \frac{1}{D_{ij}} - \frac{1}{D_v} \right) + \frac{1}{2} \sum_{i=1}^N q_i^2 \left[ \frac{1}{R_i} \left( \frac{1}{D_i} - 1 \right) - \frac{1}{R_{i,0}} \left( \frac{1}{D_v} - 1 \right) \right] \quad (6)$$

where,  $D_{ij} \equiv D_{ij}(q_i, q_j, r_{ij}, \mathbf{r})$ ,  $D_i \equiv D_i[q_i, R_i(q_i, \mathbf{r}), \mathbf{r}]$ , and  $R_i \equiv R_i(q_i, \mathbf{r})$ , and  $R_{i,0} \equiv R_{i,0}(q_i, \mathbf{r})$  is the effective radius of atom  $i$  when the solute is in vacuum. Equation 6 is valid for a solute that can be assumed to be spatially homogeneous in its dielectric properties and characterized by a constant permittivity  $\epsilon_v = D_v > 1$ . Experiments and simulations have shown that proteins are highly heterogeneous, with local dielectric variations.<sup>57–60</sup> Therefore, eq 6 is not adequate to estimate  $\Delta F_e$  for such systems, and  $D_v$  must be replaced by the complete position-dependent  $D_v(\mathbf{r})$ . As for any thermodynamic quantity, estimation of  $D_v(\mathbf{r})$  requires first a simulation of the protein dynamics, and then the calculation of the corresponding time average once adequate sampling and convergence are achieved.<sup>59,60</sup> These calculations are computationally demanding, and the results depend on the quality of the protein force field. On the other hand, estimating  $D_v(\mathbf{r})$  simply from a rigid protein structure is conceptually more difficult and also beyond the scope of this paper. Because the parametrization of the SCPISM is based on hydration energies of small molecules,<sup>36</sup> the assumption of dielectric homogeneity is reasonable, and eq 6 can then be used.

**II.2. Effective radii  $R$  and effective screenings  $D$ .** In the SCPISM the dependence of  $R$  with  $\mathbf{r}$  (protein conformation) in eq 5 is given by<sup>36</sup>

$$R_i = R_{0,i} + \delta_{w,i} \xi_i + \delta_{p,i} (1 - \xi_i) \quad (7)$$

where  $p$  and  $w$  indicate "protein" and "water", respectively. In this equation  $R_{0,i} + \delta_{w,i}$  is the effective radius of the atom when it is fully hydrated, while  $R_{0,i} + \delta_{p,i}$  is the radius when the atom is fully dehydrated, i.e., in the interior of an infinitely large generic protein. The dependence on the charge  $q$  is through  $\delta_{w,i}$  and  $\delta_{p,i}$ , and  $R_{0,i}$  is taken as the covalent radius of the atom.<sup>36</sup> The function  $\xi_i$  is a measure of exposure of the atom  $i$  to water, which is approximated by the function<sup>37</sup>

$$\xi_i = 1 - \sum_{j \neq i}^N A_{ij} \exp(-B_{ij} r_{ij}) \quad (8)$$

where  $A$  and  $B$  are parameters that determine the degree of exclusion of water by the protein. Complete hydration implies  $r_{ij} \rightarrow \infty$  for all atoms  $j$ , so  $\xi_i \rightarrow 1$ . Complete dehydration can be represented by  $i$  at the center of an infinitely large solvent-excluding cavity (e.g., a protein with  $N \rightarrow \infty$ ), where  $A$  and  $B$  are such that  $\xi_i \rightarrow 0$ .

Depending on its position in a protein, an atom polarizes the solvent only partially and anisotropically. Since this is manifested in the dielectric response, the screening  $D$  must also depend on the degree of water exclusion. To account for the dependence of



$D$  on the protein structure, the screening parameter  $\alpha$  depends on the proximity of the atom to the solvent, as  $\xi$  does in eq 8. If the particle is gradually transferred from water to vacuum, then  $D(r)$  must converge gradually to unity as the particle moves further away from the water interface. This condition requires  $\alpha$  to vanish smoothly, i.e.,  $\alpha \rightarrow 0$ , leading to  $D(r) \rightarrow 1$  for all  $r$ . A similar situation arises for an atom at the center of an increasingly large solvent-excluding cavity, as water moves away from the atom. The screening parameter of a fully hydrated atom  $i$  is here denoted by  $\alpha_{w,i}$ . The corresponding parameter  $\alpha_i$  for the atom in the protein can then be described by the perturbation

$$\alpha_i = \alpha_{w,i} + \delta\alpha_i \quad (9)$$

where  $\delta\alpha_i < 0$ . In analogy with eq 8 the degree of solvent exclusion is approximated by

$$\delta\alpha_i = \sum_{j \neq i}^N A_{ij}' \exp(-B_{ij}' r_{ij}) \quad (10)$$

where  $\delta\alpha_i \rightarrow 0$  for  $r_{ij} \rightarrow \infty$  (full hydration); and  $A'$  and  $B'$  are such that  $\delta\alpha_i \rightarrow -\alpha_{w,i}$  when the particle is at the center of an infinitely large solvent-excluding cavity (complete dehydration). In the latter case,  $\Delta F_e$  in eq 4 vanishes, although both  $R$  and  $D$  retain finite values:  $R \rightarrow R_0$ , as  $\delta = 0$  in vacuum (cf. eq 7), and  $D \rightarrow 1$ , as  $\alpha \rightarrow 0$  [cf. eq 9]. Thus, in the current extension of the SCPISM, both  $R$  and  $\alpha$  contain information of the protein structure. Straightforward generalizations of the original Born model, e.g., so-called GB models,<sup>27,33</sup> account for solvent-exclusion effects by adjusting the size of the Born radius  $R$  to the appropriate value. In this case,  $R$  diverges as the protein becomes infinitely large. The approach used in the SCPISM, where both  $R$  and  $\alpha$  can adjust to the environment is expected to better account for solvent exclusion effects in nonglobular environments, such as peptides, membranes, and molecular interfaces, or in heterogeneous environments in general.

**II.3. SCPISM Implementation.** *II.3.1. Molecular Solutes.* The hydration energy of a charge  $q$  is given by<sup>39</sup>

$$\Delta F_e = -\frac{1}{2}q^2 \lim_{L \rightarrow \infty} \int_R^L \frac{\varepsilon(r) - 1}{r^2 \varepsilon(r)} dr \quad (11)$$

where  $\varepsilon(r)$  is given by solutions of eq 1. Equation 3 is an analytical solution of the integral of eq 11 obtained upon a suitable change of variable and after using the relation between  $\varepsilon$  and  $D$  given in section II.<sup>36,48</sup> Calculations show<sup>39</sup> that, as a function of the upper limit  $L$ , this integral increases rapidly within the first two hydration shells (accounting for 50–70% of the total hydration energy of single ions), and much slower at longer distances. Therefore, water-exclusion effects on electrostatics can be partitioned into a short-range and a long-range component. This is exploited in the SCPISM to speed up computation: the short-range dependence is introduced through the effective radii  $R$ , and the long-range dependence through the screening parameters  $\alpha$ . The structure dependence of  $R$  is fine grained, and depends on the atomistic details of the protein within a certain distance  $d_c$  from the solute atom. This distance is chosen as  $d_c = 7.5$  Å, enough to accommodate two hydration shells. In contrast, the structure dependence of  $\alpha$  is assumed to be coarse grained, since water-exclusion effects due to a distant solute depend less critically on its internal atomic structure. This partitioning of the system into short-range/fine-grain and long-range/coarse-grain

contributions allows for fast computation and efficiency that increases with system size.

The short-range dependence of  $R$  on the protein structure has been discussed,<sup>37</sup> and is given by a simplification of eq 8, with  $N$  replaced by the number of atoms  $N_c$  within a sphere of radius  $d_c$  centered at  $i$ . The focus here is on the long-range effects of water-exclusion, which are represented by  $\alpha$  in eq 9, and  $\delta\alpha$  given by a simplification of eq 10, in the form

$$\alpha_i = \alpha_{0,i} + \sum_{j \neq i}^M A_{IJ}' \exp(-B_{IJ}' r_{IJ}) \quad (12)$$

where  $M$  is the total number of residues in the protein (or nucleotides, or any basic units in which the system can be conveniently divided), and  $r_{IJ}$  is the distance between the  $C_\alpha$  atoms of residues  $I$  and  $J$ . In eq 12 the unperturbed parameter  $\alpha_{0,i}$  determines the screening assigned to atom  $i$  in the fully hydrated residue  $I$ , whereas  $\alpha_{w,i}$  in eq 9 determines the screening assigned to the same atom when it is fully hydrated. Thus,  $\alpha_{0,i}$  is itself a perturbation of  $\alpha_{w,i}$  due to the exclusion of water by all other atoms ( $j \neq i$ ) of residue  $I$ . The perturbation of eq 12 then contains information on the additional water-exclusion effects due to all other residues ( $J \neq I$ ) surrounding  $I$ . For monatomic ions, clearly  $\alpha_{0,i} = \alpha_{w,i}$ . In the SCPISM the parameters  $\alpha_{0,i}$  are the actual quantities used in the parametrization of the model through eq 6.<sup>36</sup>

The parameters  $A'$  and  $B'$  in eq 12 quantify the strength of the long-range effects of solvent exclusion, like  $A$  and  $B$  in eq 8 quantify the corresponding short-range effects. In the current implementation  $A'$  and  $B'$  are independent of  $I$  and  $J$ . For a particle in the interior of an infinitely large solvent-excluding cavity,  $D \rightarrow 1$ , thus the parameter  $A'$  can be determined by setting  $0 = \alpha_{0,i} + \delta\alpha_i$ , and eq 12 can be rewritten as

$$\alpha_i = \alpha_{0,i} - \frac{\alpha_{0,i}}{S(\sigma)} \sum_{j \neq i}^M \exp(-r_{IJ}/\sigma) \quad (13)$$

where  $\sigma \equiv 1/B'$ . In eq 13,  $S(\sigma)$  is a summation over the volume of an idealized infinitely large protein ( $M \rightarrow \infty$ ), that is

$$\begin{aligned} S(\sigma) &= \sum_{j \neq i}^M \exp(-r_{IJ}/\sigma) \approx \int_p^\infty \frac{4\pi}{v_r} r^2 \exp(-r/\sigma) dr \\ &= 3 \left( \frac{\sigma}{p} + 2 \frac{\sigma^2}{p^2} + 2 \frac{\sigma^3}{p^3} \right) \exp(-p/\sigma) \end{aligned} \quad (14)$$

where the sum has been approximated by an integral using  $v_r = \langle V_P/n_P \rangle = 183 \pm 10$  Å<sup>3</sup> as the average volume per residue, and  $p = (3v_r/4\pi)^{1/3} \approx 3.5$  Å is a radius;  $V_P$  and  $n_P$  are the volume and the number of residues of a protein  $P$ , and the average  $\langle \rangle$  is taken over all of the proteins in the PDB with more than 500 amino acids. Volumes were calculated numerically using a three-dimensional mesh of cubic unit cells with sides of 1 Å.

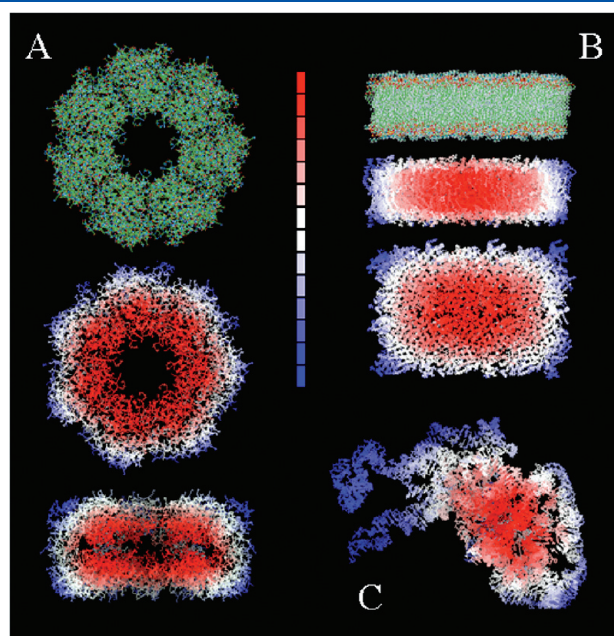
A similar procedure can be followed to estimate  $\alpha_{ij}$  and the correction  $\delta\alpha_{ij}$  of the interaction terms  $E_{e,ij} = q_i q_j / r_{ij} D_{ij}$  in eq 5. In the original SCPISM implementation<sup>36</sup> the parameter  $\alpha_{ij}$ , which defines the two-particle screening function  $D_{ij}$  is given by  $\alpha_{ij}^2 = \alpha_i \alpha_j$ . This form is admittedly arbitrary and chosen only to preserve the symmetry of  $E_{e,ij}$  upon index permutation. This restriction is relaxed here, although the symmetry is still preserved. The correction  $\delta\alpha_{ij}$  can be obtained in terms of

single-particle corrections  $\delta\alpha'_i$  and  $\delta\alpha'_j$  to first order, that is

$$\alpha_{ij} \approx \alpha_{0,ij} + \delta\alpha'_i + \delta\alpha'_j \quad (15)$$

where  $\alpha_{0,ij}^2 = \alpha_{0,i}\alpha_{0,j}$  is the uncorrected parameter.<sup>36</sup> The interaction energy of two particles  $i$  and  $j$  in the middle of an infinitely large solvent-excluding cavity requires  $D_{ij} \rightarrow 1$ ; that is,  $E_{e,ij} \rightarrow q_i q_j / r_{ij}$ . Thus, imposing  $0 = \alpha_{0,ij} + \delta\alpha_{ij}$ , the perturbation  $\delta\alpha'_i$  is given by

$$\delta\alpha'_i = -\frac{\alpha_{0,ij}}{2S(\sigma')} \sum_{K \neq I}^M \exp(-r_{IK}/\sigma') \quad (16)$$



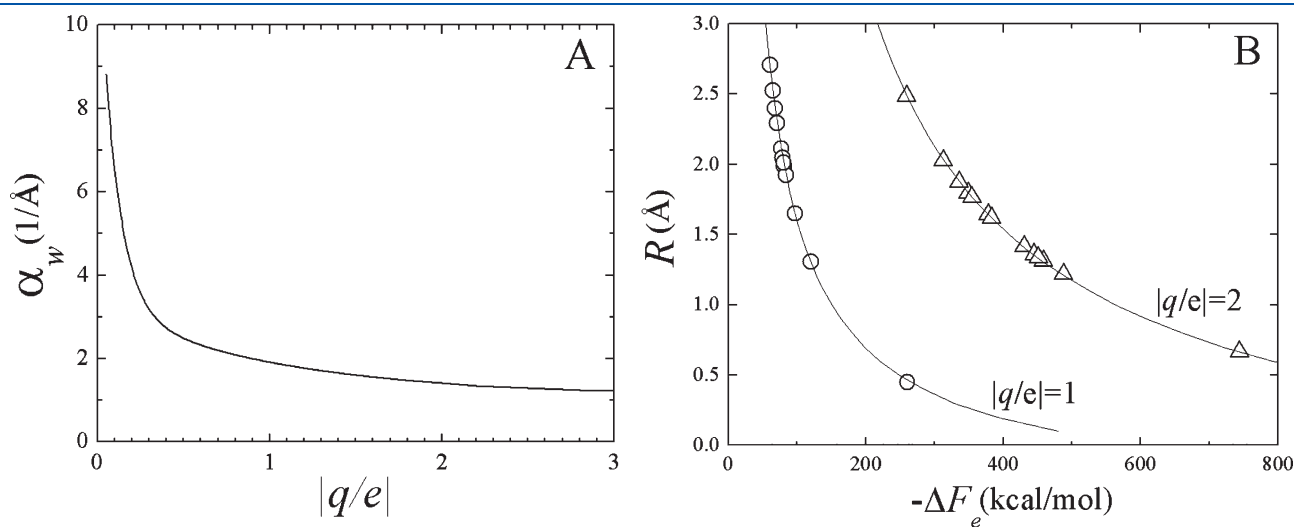
**Figure 1.** Representation of water-exclusion effects for (A) a multi-protein assembly, (B) a rectangular slice of lipid bilayer, and (C) a bundle of RNA. Regions more excluded from water are red [larger  $\delta\alpha$  in eq 9], and regions more exposed to water are blue (smaller  $\delta\alpha$ ). Similar behavior could be illustrated for  $\delta\alpha'_i$  and  $\delta\alpha'_j$  in eq 15. Upper panel of A and B colored by atom (C: green, O: red, N: blue, S: yellow). Lower panel of A and B rotated by 90°. All figures are cross sections.

where  $S$  is also given by eq 14. The correction  $\delta\alpha'_j$  for atom  $j$  is obtained from eq 16 by changing index  $I$  to  $J$ . In eqs 13 and 16,  $\sigma$  and  $\sigma'$  are the characteristic decay lengths of the long-range effects of water exclusion on the self-energy and interaction terms, respectively. Since  $\sigma$  and  $\sigma'$  can now be chosen independently, the exclusion of solvent can have different effects on the self-energy and the interaction terms. Therefore, the two terms in eq 5 are completely decoupled. The correct balance between these terms is discussed in section IV.1.

Both  $\sigma$  and  $\sigma'$  are expected to be larger than  $d_c$ . The values of these parameters and their relation to the system thermodynamics is better discussed in the context of protein complexation (cf. section IV.1). The effects of solvent exclusion on  $\delta\alpha$ 's are illustrated in Figure 1 for three different cases: a protein aggregate, a bundle of RNA, and a rectangular slice of lipid bilayer.

**II.3.2. Single Charges.** To obtain solvation parameters for a single charge  $q$ ,  $\epsilon(r)$  is first calculated numerically from eq 1, which requires the value of  $q$ , the temperature  $T$ , and the properties of the solvent molecules. From  $\epsilon(r)$  the screening function  $D(r)$  is then calculated<sup>38</sup> (cf. section II), and the screening parameter  $\alpha$  is thus obtained from eq 2. This parameter is denoted by  $\alpha_w$  and is also subject to the solvent-exclusion effects indicated by eq 9 when the charge is partially solvated. Finally, the effective radius  $R$  is calculated from eq 3 using experimental values of  $\Delta F_e$ . The radius can be further partitioned as<sup>39</sup>  $R = R_I + \delta(q)$  and the extension  $\delta$  obtained when the ionic radii  $R_I$  are known.<sup>53</sup> The procedure just outlined combines the LDS theory (cf. eq 1), the SCPISM self-energy (cf. eq 3), and experimental  $\Delta F_e$  data, and can be used to obtain parameters for other temperatures or solvents.

Figure 2A shows the dependence of  $\alpha_w$  on  $q$  obtained from numerical solutions of eq 1. The values of  $\alpha_w$  in Figure 2A indicate that the dielectric behavior of bulk water is typically reached at  $r \approx 5$  Å from the central ion, although other models<sup>44,46</sup> suggest much larger distances, in some cases beyond  $\sim 10$  Å. Figure 2B shows the corresponding radii  $R$  versus  $\Delta F_e$  obtained from a numerical inversion of eq 3. Experimental data for common ions are also shown (open symbols). The effective radii tend to fall in the range  $\sim 1$ – $2.5$  Å with the exception  $\text{Be}^{2+}$



**Figure 2.** (A) Effective screening parameters  $\alpha_w$  as a function of the charge  $q$  ( $e$  is the electron charge) and (B) effective radii  $R$  as a function of the electrostatic free energy of hydration  $\Delta F_e$ . Theoretical curves in (B) from inversion of eq 3; experimental data (symbols) taken from ref 53.

( $R \approx 0.67$  Å), which reduces its size upon double ionization, and  $H^+$  ( $R \approx 0.45$  Å). Although  $Li^+$  has the same electronic configuration as  $Be^{2+}$  ( $1s^2$ ), its effective radius is twice as large ( $R \approx 1.31$  Å) because of the reduced charge, i.e., less saturation. Quantum mechanical effects, although always present, do not in general lead to covalent bonds between the central ion and water, so the classical LDS assumptions are quite appropriate. For  $H^+$ , however, there are transient formations of valence bonds with nearest oxygen atoms, leading to (unstable)  $H_3O^+$  species. This is expected to decrease the average distance between the proton and the surrounding water, thus leading to slightly smaller effective radii than would be obtained if a purely classical treatment applies. This inconsistency is not expected to be a major concern here since the present development is aimed at developing parameters for biologically relevant cations, such as  $Na^+$  and  $K^+$  for use in classical simulations, for which these effects are less important.

The optimization of the SCPISM for ions just outlined is only the first step prior to introducing explicit ions in an implicit solvent. A number of additional developments are needed before ions and other cosolutes can be used routinely in simulations. These include the estimation of the short-range liquid-structure effects,<sup>25,39</sup> as done for amino acids in the next section, the optimization of Langevin parameters to reproduce experimental self-diffusion coefficients as a function of the ion concentration, and a computational algorithm to mimic anisotropic collisions with water in Langevin dynamics. These effects are important for all solutes but are particularly critical for ions and small molecular species for which the granularity of the solvent becomes a significant challenge.

### III. LIQUID-STRUCTURE EFFECTS IN THE SCPISM

Although the behavior of interfacial water is not yet fully understood, the first two hydration shells are known to be essential in modulating short-range intermolecular interactions. The structure of the liquid in these shells determines the characteristic shape of the intermolecular potentials of mean force (PMF), including the contact minimum, the desolvation barrier, and the solvent-separated minimum. The forces between closely interacting solutes are influenced by the hydrogen-bond (HB) network of the surrounding water, which undergoes important changes as the hydration shells come into contact.<sup>25,61</sup> These liquid-structure forces are also known as solvent-induced forces<sup>25,62–64</sup> (SIF). At the microscopic level, hydrophobic and hydrophilic forces are different manifestations of SIF. These forces are known to play a role in the action of ions and cosolutes on protein denaturation, stabilization, precipitation, and self-assembly.<sup>65,66</sup> They also operate at the surface of proteins, e.g., between water-exposed side chains or during ligand binding. Although SIF have been studied extensively with computer simulations<sup>25,61–64</sup> and in theories of nonhomogeneous liquids,<sup>41,67</sup> ISMs commonly used in macromolecular simulations have largely neglected their effects, except for simple treatments of hydrophobic forces. Part of the challenge is that SIF cannot be described by a purely electrostatic model as they are not solely electrostatic in origin, although they are affected by the solute charge distribution. Therefore, an explicit treatment is needed to introduce the effects of SIF in a computationally efficient model of solvation, and the approach used in the SCPISM is described below.

**III.1. Hydrogen Bond Interactions at Protein Interfaces.** A hydrogen bond is characterized by its strength and its

geometry.<sup>68</sup> In the gas-phase, HB geometries are determined by the electron density at the donor and acceptor atoms, with a characteristic directionality provided by the orientation of the electron lone pairs. These gas-phase geometries can still be identified in the condensed phase, although they are perturbed by the specific environment. Thus, in the solid state the perturbation is provided by the crystal field; while in aqueous solutions the geometry is modulated by the SIF. The original SCPISM addressed the problem of SIF modulation on both HB strength and geometry.<sup>68</sup> It was shown that both HB features are important for correct structure prediction in Monte Carlo simulations of small peptides. For Langevin and molecular dynamics simulations, however, the algorithm that controls the HB geometry in the SCPISM has been removed.<sup>37</sup> This parallels the development of the molecular force field itself (in this case, CHARMM) which originally contained an explicit HB-geometry term that was subsequently removed for practical purposes. Because the HB geometry is mostly determined by the electron structure of the interacting groups, it must be contained in the molecular force field in the first place, with the SIF only providing the additional perturbation to this gas-phase HB geometry. Therefore, this section deals only with an optimization of the HB strength in the SCPISM, i.e., the SIF modulation of the gas-phase energy.

Small hydrophilic molecules provide an ideal framework to quantify the effects of SIF and the resulting PMF at short intermolecular distances. A systematic study of the PMF in explicit water has been reported using a representative set of 42 amino-acid dimers containing polar and charged groups.<sup>61,69</sup> These include the extreme cases of charge–charge interactions (i.e., ionic or salt-bridge), as well as charge–polar, polar–polar, and nonpolar–nonpolar (hydrophobic) interactions. The PMF is given by  $V(r) = \phi(r) + V_w(r)$ , where  $\phi$  is the intermolecular potential energy in the gas phase and  $V_w$  is the intermolecular energy calculated from the forces exerted by the solvent. The “reaction coordinate”  $r$  is assumed to be the distance between the acceptor (A) atom and the proton shared by the acceptor and the donor (D) atoms. Both  $\phi(r)$  and  $V_w(r)$  vanish as  $r \rightarrow \infty$ . The microscopic behavior of water as the monomers dissociate from close contact has been discussed in detail.<sup>61</sup> The potentials show in all cases the usual desolvation barrier [ $V(r) = V_c$  at  $r = r_c$ ] separating the close-contact ( $V_c$  at  $r_c$ ) and the solvent-separated ( $V_{ss}$  at  $r_{ss}$ ) minima. Additional, less pronounced maxima and minima appear beyond the water-separated distance, as a result of further restructuring of water in the intervening space between the monomers.<sup>25,61</sup> This restructuring, which is stronger for charged groups, can be visualized from the positions and relocation of the peaks of water density as the distance between the monomers varies.<sup>25,39</sup> An analysis of the contributions to the PMF made by bulk and nonbulk water shows the fundamental role of SIF in determining the form of  $V_w(r)$ ,<sup>61</sup> in particular, the value  $V_c$  which quantifies the “HB energy” between the monomers in water-accessible regions. These HB energies are determined by the solvent, and must be accurately reproduced with an ISM. By contrast, the gas-phase HB is given by  $\phi(r)$ , which is accounted for by the parametrization of the protein force field, and is not addressed in this paper.

**III.2. SCPISM Implementation.** The challenge here is to incorporate the effects of SIF between polar/charged groups into a computationally efficient algorithm that reproduces  $V_c$  at  $r_c$ .



In the SCPISM the total energy  $E_{\text{SCP}}$  of a solute in a solvent is partitioned into a purely electrostatic contribution  $E_e$ , given by eq 5, and a short-range correction  $E_{\text{SIF}}$  due to the effects of SIF,<sup>39</sup> that is

$$E_{\text{SCP}} = E_e + E_{\text{SIF}} \quad (17)$$

For a nonpolar solute  $E_e = 0$  and then  $E_{\text{SCP}} = E_{\text{SIF}}$ , which in this case is simply a “hydrophobic” energy, typically modeled as a cavity term  $E_{\text{cav}}$ . For a polar solute  $E_e \neq 0$  and  $E_{\text{SIF}}$  is rather complicated, and depends on the solute topography and charge distribution. A physically complete ISM must reproduce the PMF for all values of  $r$ , including  $V_c$  (which controls thermodynamic equilibrium),  $V_t$  (kinetics), and  $V_{ss}$ . A numerical algorithm has been developed to achieve this goal with the SCPISM in the context of Langevin dynamics simulations.<sup>39</sup> A less demanding but highly efficient approach has been reported earlier,<sup>68</sup> which is used here, and aims at reproducing  $V_c$  only. The method is based on an extension of eq 7 for polar hydrogen atoms

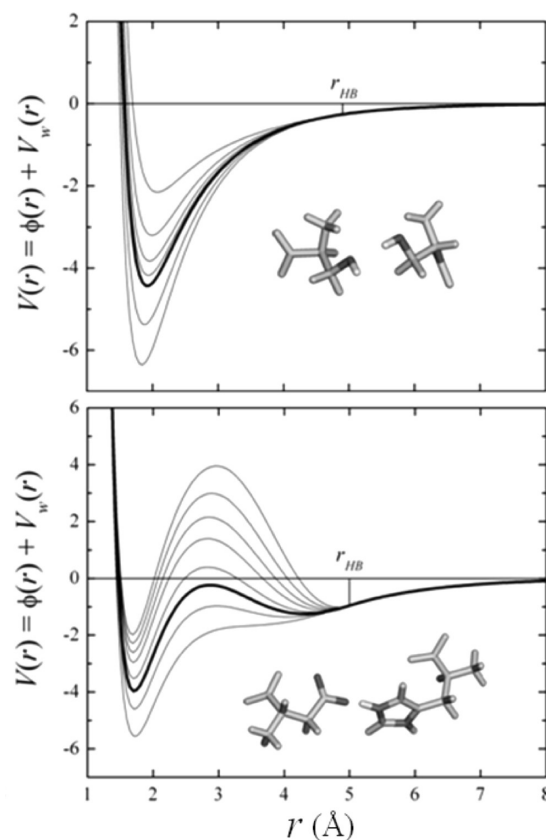
$$R_i = R_{0,i} + \delta_{w,i}\xi_i + \delta_{p,i}(1 - \xi_i) + \delta_{A,i}\zeta_i \quad (18)$$

where the fraction  $(1 - \xi_i)$  in eq 7 has been divided into a fraction  $\zeta_i$  and a fraction  $(1 - \xi_i - \zeta_i)$ . As long as the proton  $i$  remains away from an acceptor  $A$ , then  $\zeta_i = 0$  and eq 18 is identical with eq 7. If the proton approaches  $A$ , then  $\zeta_i$  measures the degree of exposure of the proton to  $A$ . The energy  $E_{\text{SCP}}$  of eq 17 is then approximated as

$$E_{\text{SCP}} \approx E_e + E_{\text{cav}} \quad (19)$$

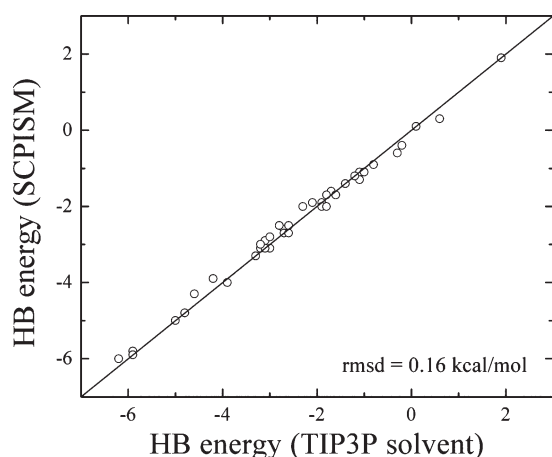
where  $E_e$  is given by eq 5, using  $R_i$  of eq 18 if  $i$  is a polar proton, or eq 7 otherwise. Equation 19 represents the complete SCPISM as currently implemented, and yields the total energy of a solute in a solvent. The cavity term  $E_{\text{cav}}$  in eq 19 is calculated as  $E_{\text{cav}} = a + bS$ , where  $a$  and  $b$  are optimized parameters based on hydration energies of small nonpolar molecules and  $S$  is the solvent-accessible surface area of the solute.<sup>36,70</sup> In the original SCPISM implementation  $S$  was calculated numerically;<sup>36</sup> in the current implementation an approximation similar to that in eq 8 is used instead, which allows analytical derivatives for the calculation of forces.<sup>37</sup> In the current implementation, this cavity term is the only source of hydrophobic forces between nonpolar solutes, which is clearly a simplification, and a better description is needed; three potentially useful approaches are currently being considered for use in the SCPISM.<sup>71–73</sup>

The parameter  $\delta_{A,i}$  in eq 18 allows fitting to HB energies. To obtain  $\delta_{A,i}$  for each amino acid pair the corresponding HB energies must be known, preferably experimentally. Such determination is difficult, although attempts have been reported. For example, HB energies in water-accessible regions of proteins have been estimated by site directed mutagenesis in enzyme–substrate complexes.<sup>74</sup> These studies showed that deleting a side chain that forms one HB between the enzyme and the substrate tends to weaken the free energy of binding by  $\sim 0.5$  to  $\sim 4.5$  kcal/mol, depending on the charge of the group left unpaired. During enzyme–substrate binding, the donor and acceptor atoms must break their HB with water molecules at the interfaces. Thus, these experiments do measure the HB energies controlled by SIF.<sup>74</sup> The original implementation of the SCPISM<sup>37,68</sup> used these estimates to calibrate all the HB energies between amino acid pairs. In this case the  $\delta_A$  values were adjusted individually for all possible side chain–side chain, side chain–backbone, and backbone–backbone interactions in proteins. The extensive



**Figure 3.** Intermolecular energy  $V(r)$  (in kcal/mol) between Ser–Ser (top) and Asp<sup>−</sup>–His<sup>+</sup> (bottom) as a function of the shared proton-acceptor atom distance  $r$ . Energy calculated with the SCPISM (eq 19) and the all-atom CHARMM van der Waals potential. Thick lines correspond to the optimized HB energy based on PMF calculations in explicit water;<sup>69</sup> thin lines illustrate the dependence of the HB energy with  $\delta_A$  (eq 18).  $V_w(r)$  and  $\phi(r)$  are the water-induced and the gas-phase intermolecular energies, respectively.

calculations of the PMF reported previously<sup>69</sup> confirmed the overall range of HB energies determined by these experiments. However, the relationship between the charge/polarity of the interacting groups and the HB strengths was less clear. Because of the limited number of HB studied experimentally, and because of the inherent difficulties in deriving individual HB interactions from kinetics data in protein complexes,<sup>74</sup> side chain–side chain HB energies in the SCPISM are here recalibrated based on the systematic calculations of PMF using explicit water simulations.<sup>69</sup> Figure 3 shows the effects of  $\delta_A$  on the HB energies in the case of two representative interactions. The thick lines show the optimized potentials that reproduce  $V_c$  and  $r_c$  obtained from the atomistic simulations. This potential is obtained with an optimized value  $\delta_A = \delta^*$  in eq 18. The thin lines are the interaction energies for different values of  $\delta_A$ , showing overstabilization ( $\delta_A < \delta^*$ ) and destabilization ( $\delta_A > \delta^*$ ) of the dimers. Since the correction is effective only for  $r < r_{\text{HB}}$ , the long-range electrostatic forces described by eq 5 are not affected by  $\delta_A$ . If the dimers are transferred to the interior of a protein, the SIF no longer play a role. In this case the modulation of HB energies becomes a purely electrostatic problem, and is then addressed by the water-exclusion effects controlled by  $\alpha$ . In the limit of complete exclusion (infinitely large solvent-excluding cavity)  $V_w \rightarrow 0$ , and the HB energies converge to the gas-phase interaction



**Figure 4.** Optimization of 42 representative HB interactions of amino acids pairs.<sup>68,69</sup> All energies in kcal/mol. HB energies are calibrated within the statistical errors in explicit water simulations ( $\sim 0.2$ – $0.4$  kcal/mol); HB distances are reproduced within the estimated errors ( $\sim 0.2$  Å).

energy as given by the vacuum protein force field. Figure 4 shows the new HB calibration for the dimers reported in ref 69 with  $V_c$  and  $r_c$  accurate to within the statistical errors of the simulations.

The computational efficiency of this approximation stems from the relatively small number of shared protons for any given conformation of the system and from the fact that  $\zeta_i = 0$  for proton-acceptor distance  $r > r_{HB}$ , where  $r_{HB} \approx 5$  Å, i.e., the distance at which the first hydration shells of the acceptor and donor groups begin to overlap.

#### IV. APPLICATIONS

The SCPISM has been implemented in the CHARMM program<sup>75</sup> and was shown to be among the most computationally efficient ISMs currently available.<sup>76</sup> The developments reported in this paper add little computational cost. In the current implementation for dynamics, and only for practical purposes, the corrections  $\delta\alpha$ 's in eq 9 and 15 are updated heuristically when any inter-residue distance  $r_{IJ}$  changes more than  $0.5$  Å from the last update. This screening-update protocol was motivated by the nonbonded list update method commonly used in macromolecular simulations, but may lead to energy conservation problems. The threshold  $0.5$  Å, however, was determined after substantial experimentation to keep energy-conservation errors small during each screening update. The current SCPISM implementation in CHARMM (version c36) is only  $\sim 1.5$  times slower than the vacuum calculation. Its parallel performance is good, with a speedup of  $\sim 30$  for 32 CPU cores using 48-core AMD Opteron 2.2 GHz machines, for 1000 dynamics steps of a  $\sim 6500$ -atom protein complex. To test and discuss the conceptual aspects of the model, two systems are considered: the complex of Barnase with its inhibitor protein Barstar and the short peptide known as Trp-cage. Both systems are well characterized experimentally, and their conformations in aqueous solutions have been reported.

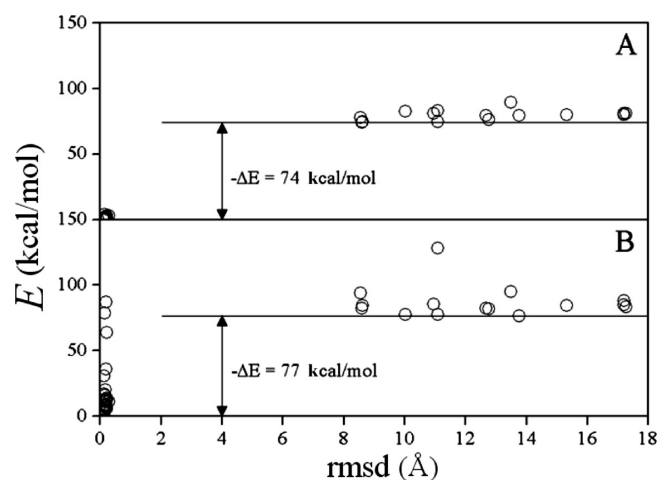
**IV.1. Protein Association.** Barstar is a small protein (89 aa) that binds to barnase (110 aa) and inhibits its activity in vivo. The crystal structure of the complex (PDB code: 1BRS) has been resolved at  $2$  Å resolution.<sup>77</sup> The proteins are commonly used as a paradigm for protein complexation.<sup>77–82</sup> Barnase and barstar associate mainly by electrostatic forces, and bind tightly to one

another due to electrostatics and HB interactions. Recent studies have also suggested a role of interfacial water in the stability of the complex.<sup>83</sup> This protein system is thus well suited to discuss all aspects of the SCPISM developed in this paper. Chains A (barnase) and D (barstar) of the crystal structure were used. The calculations assume that the proteins are immersed in pure water, a common but imperfect assumption.<sup>61</sup> Standard ionization states at neutral pH are used. Asp and Glu residues are assumed to be unprotonated, whereas Arg and Lys are protonated. There are three His residues in the complex, which are assumed to be protonated, as well. The complex is thus net neutral. H102 is located at the barnase/barstar interface and its  $pK_a$  is known to shift upon complexation.<sup>77</sup> Calculations were repeated with a neutral H102, but the qualitative observations reported here were not affected.

The simulations consist of Monte Carlo (MC) annealing. The polar-hydrogen (param19) representation of the protein force field is used to speed up computation. Because the original SCPISM implementation in CHARMM was parametrized for the all-atom (param22) force field,<sup>36</sup> these simulations required a parametrization of the model for param19. The HB energies were also calibrated for param19 as discussed in section III.2. Two types of standard metropolis MC simulations were carried out. In “RBMC” runs, the proteins were treated as rigid bodies, whereas in “RBDMC” runs, selected dihedral angles at the protein surface were allowed to rotate in addition to rigid-body movements. Many proteins that form heterodimers are reasonably well represented as rigid bodies, since their backbone conformations change little from the uncomplexed to the complex forms. This is the case for barnase and barstar. RBMC moves consisted of rotation, translation, and rotation-translation of one of the proteins, chosen randomly with equal probability. Rigid body rotations were represented by quaternions. Both translations and rotations were generated randomly, with distances and angles taken from Gaussian probability distributions using the Box-Muller method. The acceptance rate  $\lambda$  was monitored on the fly at each temperature  $T$ , and the widths of the distributions were adjusted as needed to keep  $\lambda$  above a certain threshold (here,  $\lambda > 0.4$ ) as  $T$  was decreased. In the RBDMC simulations, rotations, translations, roto-translations, and single-dihedral rotations were considered, each with equal probability. Dihedral angles were selected randomly and moved based on a Gaussian probability distribution with a variable width used to keep the acceptance rate above the threshold. The updates of the widths of the Gaussian distributions in both the RBMC and RBDMC violate detailed balance, which affects convergence to Boltzmann distributions unless proper corrections to the acceptance probability are introduced. Simulations were also carried out without  $\lambda$ -stabilization and the results reported here were similar. The goal was to find low-energy complexes, and no thermodynamic quantities were calculated from the ensembles generated in this study.

Initially, the proteins were distributed randomly in a spherical simulation cell of radius  $R_c$ . This radius can be selected heuristically or fixed based on the molar concentration of the proteins. The latter setup is more convenient when several types of proteins and ions are used to mimic realistic experimental setups. For the analysis intended here, the radius was set to  $R_c = 85$  Å, which corresponds to a very low concentration and allows enough mobility of the proteins within the container. If the center of mass of a protein moved outside of the container during a translation, the protein is relocated in a random position inside the container. This is equivalent to imposing an infinitely



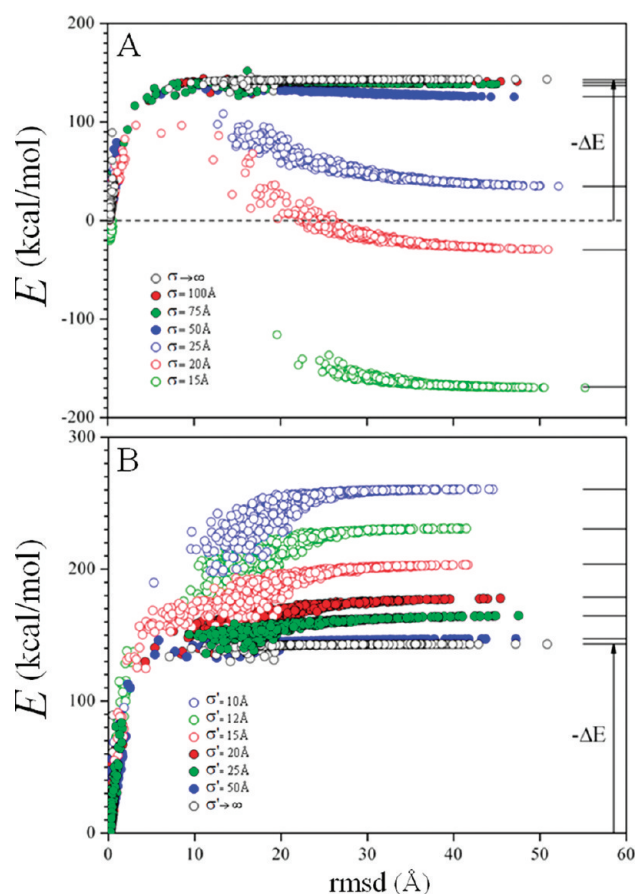


**Figure 5.** Results of the rigid-body Monte Carlo (RBMC) simulated annealing of barnase and barstar: Fifty complexes before (A) and after (B) energy minimization. Energy  $E$  is the sum of the CHARMM protein force field (param19) and the SCPISM energy  $E_{\text{SCP}}$  (cf. eq 19; parametrized for param19).  $C_{\alpha}$ -rmsd measured with respect to the crystal structure.  $\Delta E$ 's are energy differences between the non-native and native complexes.

repulsive potential outside the container. All of the simulations were started at  $T = 3000$  K and ended at  $T = 296$  K (crystal growth temperature), and annealing was achieved through a logarithmic schedule in 20 temperature steps. A total of  $2 \times 10^5$  MC steps was performed per temperature for both RBMC and RBDMC simulations. This MC protocol was based on the authors' previous experience but was not optimized. No cutoffs were used for the treatment of nonbonded interactions. For comparison, an additional RBMC simulation was carried out with a cutoff of 14 Å, which sped up computation by a factor of 4, but the main results did not change and are not discussed. The widths of the distributions were initially set to 10 Å for the translations, and 180° for rigid-body and dihedral-angle rotations.

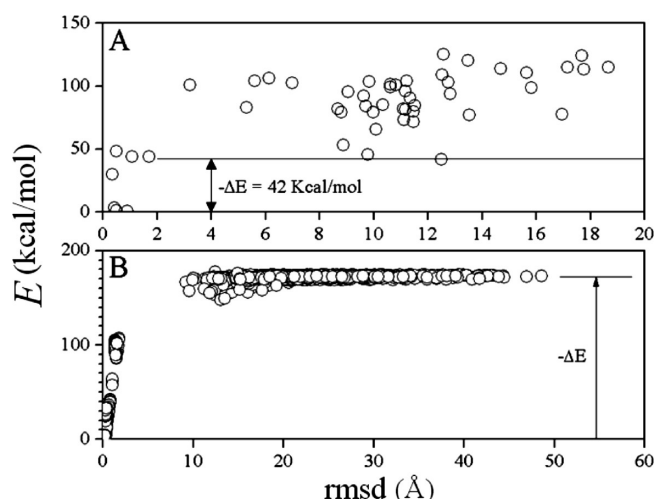
Each simulation consisted of fifty independent MC runs. Figure 5 summarizes the results for the RBMC simulation. Only the final structures in each run are shown (circles). About 75% of the runs identified the native population as the one with lowest energy ( $C_{\alpha}$ -rmsd < 0.5 Å with respect to the crystal structure). The non-native conformations show no common structural features. All of the complexes were subsequently energy-minimized without restraints for 50 steps using an adopted basis Newton–Raphson algorithm, and the energies are plotted in Figure 5B. The gap  $|\Delta E|$  changes little, although the native conformations show a spread in energy due to internal structural relaxations.

The simulations discussed above were carried out with  $\sigma \rightarrow \infty$  and  $\sigma' \rightarrow \infty$  in eq 13 and 16, which means that the model neglected the long-range effects of water exclusion. The short-range effects are still accounted for through eq 8. However, long-range water-exclusion effects are large when two proteins come into contact, as there is a sudden exclusion of water at the protein/protein interface during complexation. Thus, protein association is an ideal process to gain insight into the long-range electrostatic effects described by the SCPISM. To carry out a systematic study of these effects, the dependences of the binding energy on the parameters  $\sigma$  and  $\sigma'$  were analyzed. Here, the crystal structure was heated from  $T = 296$  K up to  $10^3$  K, generating a set of decoys which includes near-native structures and fully dissociated proteins.



**Figure 6.** Long-range water-exclusion effects on the self-energy (upper panel) and interaction-energy (lower panel) terms of the total electrostatic energy  $E_e$  in the SCPISM (cf. eq 13). Decoys are generated from the crystal structure by heating. Characteristic lengths  $\sigma$  and  $\sigma'$  are defined in eq 13 and eq 16. The limits  $\sigma \rightarrow \infty$  and  $\sigma' \rightarrow \infty$  means that only short-range water-exclusion effects are accounted through eq 8 (original electrostatic description in the SCPISM<sup>36,37</sup>). HB energies are based on systematic PMF calculations in explicit water.<sup>69</sup>  $\Delta E$ 's are energy differences between the fully dissociated and the native complexes, i.e., binding energies). A reasonable balance between  $\sigma$  and  $\sigma'$  can be obtained from experimental binding energies or dissociation constants.

Simulations were performed as a function of  $\sigma$  and  $\sigma'$ , with each parameter varying in the interval  $[10 \text{ Å}, \infty)$ . Figure 6A shows the energies of the decoys as a function of the  $C_{\alpha}$ -rmsd for selected values of  $\sigma$  and  $\sigma' \rightarrow \infty$ . By setting  $\sigma' \rightarrow \infty$ , only the effect of water exclusion on the self-energy is assessed. In this case  $|\Delta E|$  is the binding energy, i.e., the energy difference between the native complex and the fully dissociated proteins, and thus larger than  $|\Delta E|$  in Figure 5. Figure 6A shows that water-exclusion effects on the self-energy tend to decrease the binding energy. This result is physically correct for any polar solute, as association implies partial dehydration, which is energetically unfavorable for polar and charged groups. Binding can actually become unfavorable for small  $\sigma$ . In this case the proteins remain dissociated for  $\sigma < 22 \text{ Å}$ , as confirmed with a set of RBMC simulations (not shown). Figure 6B shows the effect of water exclusion on the interaction energy, i.e., for  $\sigma \rightarrow \infty$ . Here the trend reverses, and the smaller the value of  $\sigma'$  the more favorable the binding energy. This reflects the fact that partial dehydration during complexation leads



**Figure 7.** (A) As in Figure 5B, for the RBDMC simulation (interaction energy only); (B) as in Figure 6, for the RBDMC simulation (interaction energy only) using short-range water-exclusion effects only ( $\sigma \rightarrow \infty$  and  $\sigma' \rightarrow \infty$ ). The qualitative behavior of  $\Delta E$  with  $\sigma$  and  $\sigma'$  is the same as in Figure 6 (not shown).

to stronger pair interactions at the protein/protein interface, which usually favors attraction.

The qualitative behavior illustrated in Figure 6 using the native complex holds for any of the non-native structures obtained in the MC simulation (not shown). The opposing trends in binding energy with  $\sigma$  and  $\sigma'$  mean that for a given set of parameters ( $\sigma < \infty$ ,  $\sigma' < \infty$ ) the resulting binding energy is a balance between the unfavorable and favorable effects of water exclusion on the different terms in eq 5. This observation provides a rationale to select values of  $\sigma$  and  $\sigma'$  based on experimental thermodynamic measurements, such as binding affinities or dissociation constants (see below).

Similar analysis was carried out for the RBDMC simulations. Figure 7A shows the energies of the final structures as a function of the all-atom rmsd from the native complex. When compared to the RBMC simulations in Figure 5, only four out of 50 structures are native, implying that better sampling is needed to accommodate the new degrees of freedom. Also, the binding energy decreased by  $\sim 30$  kcal/mol, indicating that sampling of the dihedral space allowed the MC to find new conformations that were not allowed with the dihedral angles fixed at the crystal-structure values.

Figure 7B shows the results of the decoys obtained by heating the native complex in a RBDMC simulation, for  $\sigma \rightarrow \infty$  and  $\sigma' \rightarrow \infty$ . The same dependence with  $\sigma$  and  $\sigma'$  observed in Figure 6 is obtained in these simulations (not shown), as dihedral movements do not change the major effects of water-exclusion discussed above. As for the RBMC simulations, the binding energy is larger than  $|\Delta E|$  in Figure 7A and for the same reasons. In addition, the binding energy in the RBDMC simulations is larger than the binding energy in the RBMC simulations, indicating that sampling the dihedral space helps find lower-energy conformations in the complex.

Three different approaches can be used to obtain values of  $\sigma$  and  $\sigma'$  that yield a correct balance between the interaction and the self-energy terms in eq 5. One method, as mentioned above, is based on experimental data on the thermodynamics of protein–protein binding. A number of enzyme–inhibitor and

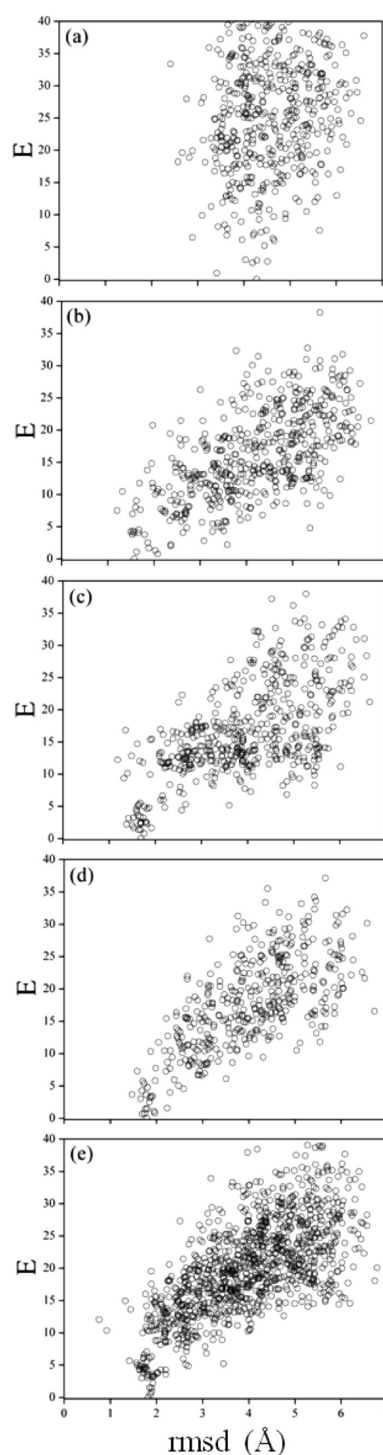
antigen–antibody complexes, for which the rigid-body approximation is known to be reasonable, are well characterized thermodynamically, as in the case of barnase and barstar. This is the method of choice in the SCPISM, as the model has been parametrized in the past using experimental data only (excepting the new calibration of the HB energies reported here, which is based on data from atomistic dynamics simulations). The optimization of  $\sigma$  and  $\sigma'$  based on thermodynamic data requires extensive calculations, and is not reported in this paper. It is possible, however, to obtain tentative values for these parameters using simpler methods, such as numerical solutions of the Poisson equation (PE). In general, solutions of the PE for peptides and proteins are known to be sensitive to the parameters used in the calculations, especially for charged atoms close to the solute/solvent interface.<sup>38</sup> For larger systems and buried atoms, i.e., away from the solute/solvent interface, solutions of the PE are less sensitive to the parameters used.<sup>38</sup> These are the conditions required to probe long-range solvent-exclusion effects. Thus, the simplest method to evaluate  $\sigma$  and  $\sigma'$  is to obtain effective screening functions  $D_{ij}$  and  $D_i$  and  $D_j$  for atoms  $i$  and  $j$  in a protein using a standard PE solver, and then to equate these PE-based effective screening functions with the corresponding ones in the SCPISM. From these calculations values of  $\alpha_{ij}$ ,  $\alpha_i$ , and  $\alpha_j$  can be obtained from eq 2, and  $\sigma$  and  $\sigma'$  can then be estimated from eq 13 and 15. Systematic calculations of PE-based screenings in proteins have been shown<sup>84</sup> to have sigmoidal behavior, similar to those used in the SCPISM (cf. section II). It has also been shown<sup>84</sup> that solvent-exclusion effects modify the PE-based effective  $D$ 's in ways that are qualitatively similar to those proposed here (cf. section II.2), i.e., the larger the exclusion, the smaller the value of  $\alpha$ . An alternative method to evaluate  $\sigma$  and  $\sigma'$  involves similar calculation of effective screenings but using atomistic dynamics simulation of water, preferably with a polarizable water model. This is more computationally demanding than solving the PE, but far more practical than using thermodynamic data of protein binding. Here, the PE based method will be used, but to avoid extensive PE calculations in proteins, an idealized spherical cavity of radius  $r_c$  centered at (0,0,0) is used instead. Two charges,  $q_1 = +1$  and  $q_2 = -1$  are located at positions  $(-r/2, 0, 0)$  and  $(r/2, 0, 0)$  respectively. The interparticle distance  $r$  is varied in the interval 2–10 Å, and for each value of  $r$  the radius  $r_c$  is changed in the 25–500 Å range. For a given  $r$  and  $r_c$  the PE is solved ( $\epsilon_i = 1$  and  $\epsilon_w = 78.4$ ) and  $\sigma$  and  $\sigma'$  are obtained as explained above. To evaluate the self-energy terms the charges are switched off, as needed. The procedure to obtain the interaction and the self-energy components from the PE solution was described.<sup>36</sup> It was shown that the SCPISM correlates well with solutions of the PE in small peptides with different conformations and charge distributions<sup>36</sup> and in decoys of small globular proteins.<sup>48</sup> To use the SCPISM in this idealized system, the fully hydrated parameter  $\alpha_{0,i}$  of eq 13 and  $\alpha_{0,ij}$  in eq 15 are taken from Figure 2 for a unit charge, and set to  $2 \text{ Å}^{-1}$ ; the fully dehydrated radii in eq 7 is set to  $R_{0,i} + \delta_{p,i} = 2 \text{ Å}$ , which is a characteristic value in proteins. Also, the summations over  $r_{IJ}$  in eq 13 and eq 16 are performed numerically in the continuum, as done in eq 14, but with an upper limit of  $r_c$ , instead of  $\infty$ , i.e.,  $M = (r_c/p)^3$ . Although the exact values of  $\sigma$  and  $\sigma'$  depends on which portion of the  $25 \text{ Å} < r_c < 500 \text{ Å}$  interval the curves are fitted, in the limit of large  $r_c$  the calculations yield  $\sigma \approx 75 \text{ Å}$  and  $\sigma' \approx 50 \text{ Å}$ . These values are assumed to be the best estimates for the current implementation.

**IV.2. Peptide Folding.** The original SCPISM has been used to predict the structure of small  $\alpha$ -helix- and  $\beta$ -hairpin-forming

peptides in solution.<sup>85</sup> To test the new implementation, and for comparison with other implicit solvent models,<sup>86–89</sup> the 20-residue peptide NLYIQWLKDGPPSSGRPPPS known as Trp-cage is used here. The structure of this peptide has been determined<sup>90</sup> by NMR (PDB code: 1L2Y) at 282 K and pH 7. All residues are assumed to be neutral, except Lys<sup>+</sup>, Arg<sup>+</sup>, and Asp<sup>−</sup>. The NMR coordinates indicate standard NH<sub>3</sub><sup>+</sup> and COOH<sup>−</sup> termini, resulting in a total charge +1. In explicit-solvent simulations, charge neutralization is enforced by introducing a set of monovalent ions in the simulation cell. This strategy is usually avoided in implicit solvation where the excess charge is generally not removed. However, some implicit models enforce electric neutralization by effectively deleting the excess charge from the system, e.g., neutralizing all the charged side chains and termini in a protein. This approach is advantageous as charged groups are known to introduce strong local perturbations that can produce artifacts, especially in structure prediction. However, local charges are important modulators of biological function, and charge modification at specific sites and times, e.g., by proton transfer, or phosphorylation, can trigger specific dynamic responses. In section II.3.2 the electrostatic component of the SCPISM was parametrized for multivalent ions. The system could then be neutralized by introducing a single Cl<sup>−</sup> ion into the simulation cell. However, it was also mentioned that the physics of solvation for small solutes is more complicated than only electrostatic interactions, and that additional developments are needed prior to using explicit ions in an ISM simulation. Pending these developments the system here retains the +1 charge, and the simulations were carried out at zero ionic strength. The all-atom representation of the CHARMM force field was used, including the CMAP correction of the backbone dihedral energy terms.<sup>76</sup> Backbone rmsd from the NMR structure was calculated relative to the first model deposited in the PDB, excluding the first and last residues.

A Monte Carlo-minimization/annealing (MCMA) conformational search<sup>86,91,92</sup> was used to find low-energy minima in the Trp-cage energy landscape. Following biased, random changes in side-chain or main-chain dihedral angles, the energy is minimized before acceptance or rejection of the trial move. Occasionally, simulated annealing is performed prior to the energy minimization to increase acceptance of structures following relatively large conformational changes. Thus, MCMA is used to search for the global energy minimum, not to achieve Boltzmann sampling of a canonical ensemble.<sup>92</sup>

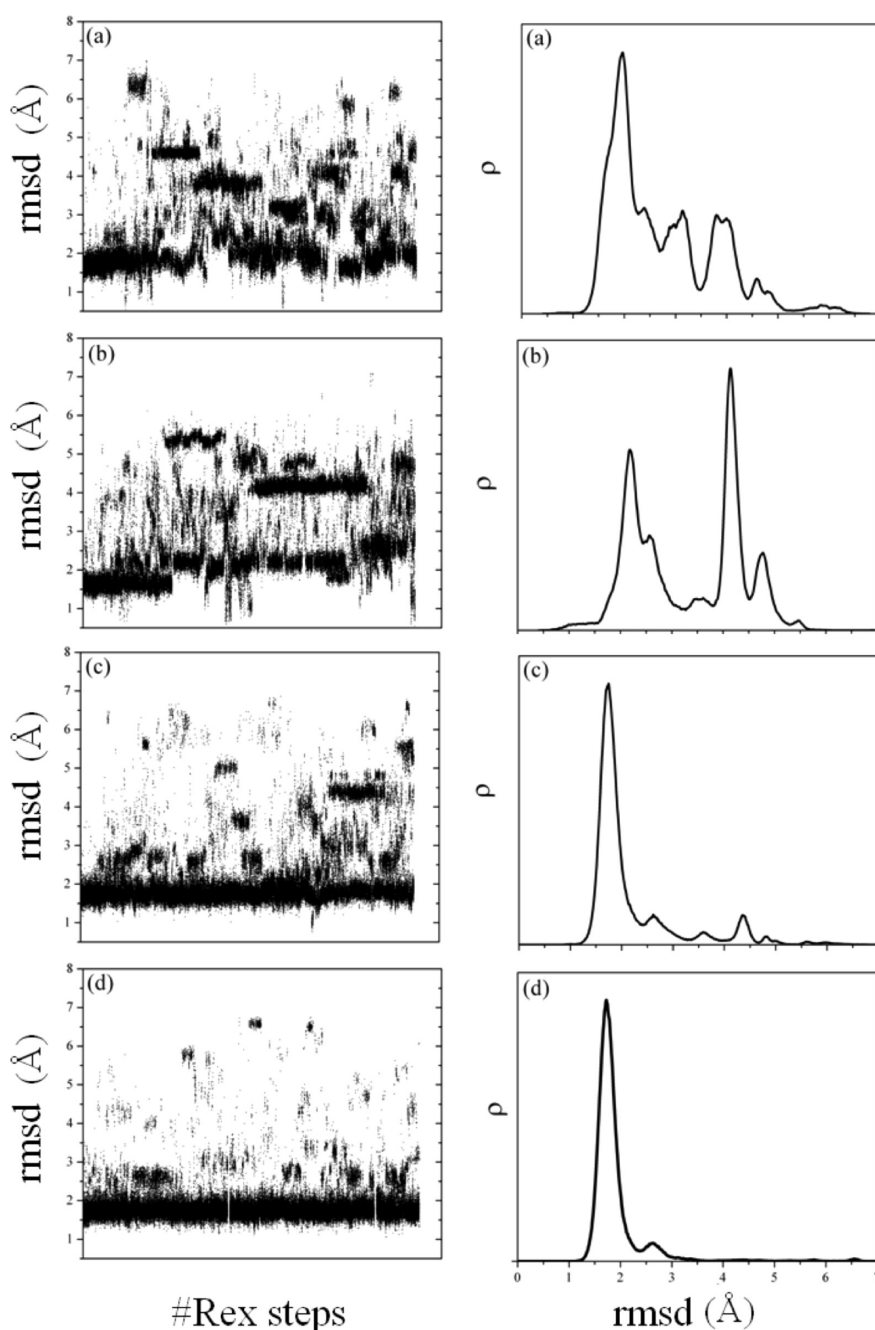
Replica-exchange simulations (REx) were performed using  $N = 24$  replicas, with each replica initiated using a structurally distinct low-energy conformation found by MCMA, as follows. The structure found using MCMA possessing the lowest energy was assigned to the lowest temperature, here 282 K (the NMR temperature). All other MCMA minima with backbone-rmsd within a cutoff distance of this first structure were assumed to belong to the same conformational family and were excluded from further consideration. Of the remaining structures, the lowest-energy structure was assigned to the second temperature, and structures within the cutoff distance of this structure were excluded. The process was repeated to produce 24 different starting structures of low energy. The cutoff distance was chosen to ensure inclusion of structures that were both low in energy and quite different from the lowest-energy structure. Temperatures of the replicas are given by  $T_n = \text{int}[282(1 + 19/282)^{n-1} + 0.5]$  with  $n = 1, \dots, 24$ , which produced a relatively uniform acceptance rate in the 0.3–0.4 range. Swaps of neighboring replicas were



**Figure 8.** Results from MCMA simulations of Trp-cage (all energies  $E$  in kcal/mol). (a) Gas-phase CHARMM force field (param22); (b) original SCPISM (HB energies based on experimental estimates; only short-range water-exclusion effects included); (c) as in (b) but with charges on the N- and the C-terminus switched off; (d) current SCPISM but with short-range water-exclusion effects only (HB interactions based on atomistic PMF calculations); to be compared with plot in (b); (e) as in (d) but with the complete SCPISM, including short- and long-range water-exclusion effects ( $\sigma = 75$  Å and  $\sigma' = 50$  Å).

considered after 100 steps of Langevin dynamics, using a collision frequency of  $2 \text{ ps}^{-1}$  for non-hydrogen atoms. The SHAKE





**Figure 9.** Results from REx simulations of Trp-cage. Left column: backbone-rmsd as a function of the number of replica-exchange steps (maximum of  $\sim 10^6$ ). Right column: histogram  $\rho$  of conformations at 282 K (NMR temperature), as a function of the backbone rmsd. (a) Original SCPISM (HB energies based on experimental estimates and only short-range water-exclusion effects included); (b) as in (a) but with terminal charges switched off; (c) current SCPISM but with short-range water-exclusion effects only (PMF-based HB energies); (d) as in (c) but with short- and long-range water-exclusion effects included ( $\sigma = 75$  Å and  $\sigma' = 50$  Å).

algorithm was applied to bonds involving hydrogen atoms. The integration time step was 1.5 fs for  $T < 700$  K and 1.0 fs for  $T \geq 700$  K. The relatively large number of replicas and the high temperatures (up to 1263 K) used represent a stringent test of the energy landscape.

Results of the MCMA simulations are shown in Figure 8. The lowest energy conformation found in vacuum (Figure 8A) has a backbone-rmsd of  $\sim 4.2$  Å with respect to the NMR structure, and all low-energy conformations have an rmsd above  $\sim 3$  Å. Secondary structure elements resembling the NMR conformation

were not observed in low-energy structures. The effect of entropy at 282 K, as evidenced by the REx simulation, further stabilizes these non-native structures (not shown), and none of the major conformational families are native-like. Thus, the MCMA/REx results confirm that the peptide does not adopt native-like structures in the gas-phase, at least with the CHARMM protein force field.

Figure 8B shows results for the original SCPISM. Only short-range effects of water exclusion are modeled, and the HB energies are calibrated based on enzyme–substrate data.<sup>68,74</sup> In contrast

to protein complexation, long-range water-exclusion effects are not expected to play a major role for small systems. The lowest-energy structure is native-like with a backbone rmsd of  $\sim 1.5$  Å. Several low-energy structures are non-native, some within  $\sim 3.5$  kcal/mol from the NMR structure. Figure 9A shows the rmsd of structures sampled at 282 K (left panel), and the conformation histogram  $\rho$  as a function of the backbone rmsd (right panel). The distribution at 282 K suggests that the misfolded states are entropically favored, and the energy gap between the native and misfolded structures (inferred from Figure 8B) is not wide enough to offset conformational entropy effects. From a predictive standpoint, however, the largest population is native-like, with a maximum located at rmsd  $\sim 1.75$  Å, and some conformations reaching rmsd  $\sim 1.1$  Å. The misfolded states detected at 282 K with REx, evidenced by the peaks in  $\rho$ , have not been discussed in the experimental literature and are apparently spurious. Simulations at lower temperatures (not shown) indicate that these misfolded states become less populated, as expected from the MCMA data in Figure 8B. Inspection of the misfolded structures shows that they are stabilized by HB interactions of the charged termini with either the backbone or the charged side chains. In the original SCPISM the N- and C-termini are treated as Lys<sup>+</sup> and Asp<sup>-</sup> side chains in the HB energy calibration. To analyze the effects of these interactions, the excess charge on each terminus was artificially removed. The corresponding MCMA data are plotted in Figure 8C, showing a rather dramatic reshaping of the energy surface. The lowest-energy structure is still native like, with rmsd of  $\sim 1.7$  Å, but the energy gap between lowest and misfolded states increased. Results of the corresponding REx simulation are shown in Figure 9B. Although the uncharged termini play no role in stabilizing the misfolded structures, one conformational family with an rmsd of  $\sim 4.2$  Å is now overstabilized due to HB interactions involving other charged side chains. Structures very close to the NMR structure (rmsd close to  $\sim 0.5$  Å) are also explored in the REx simulation. Although there is a relatively large native-like population, this is insufficiently populated for successful blind structure prediction.

The above results confirm that charged groups can affect the proper folding but are also prone to engage in spurious local interactions that might overstabilize non-native structures. This also shows why careful consideration must be given to the strength of HB interactions. Figure 8D shows results of the current implementation of the SCPISM. The original SCPISM calibration of the HB strength is here replaced by the new calibration shown in Figure 4. Results in Figure 8D are obtained with short-range water-exclusion effects only ( $\sigma \rightarrow \infty$  and  $\sigma' \rightarrow \infty$ ), and the termini charges are restored, thus panel (D) can be compared with panel (B). The improvement is apparent, and all misfolded structures have now relatively large energies compared to the native structure. As shown in Figure 9C the effects of conformational entropy are not sufficient to stabilize the misfolded state at 282 K. Some non-native families can still be detected, but a blind prediction could unambiguously identify native-like structures, with a maximum in the probability distribution located at rmsd  $\sim 1.75$  Å. An analysis of the long-range effects of water-exclusion can also be made. Figure 8E shows the current implementation with  $\sigma = 75$  Å and  $\sigma' = 50$  Å, values based on results from the protein complex discussed above. This simulation retains the current HB strength based on PMF calculations, and thus panel (E) can be compared directly to panel (D). Results are similar to those in panel (D), although a modest

improvement is observed in the histogram of Figure 9D. The native population can be identified, and misfolded structures have destabilized slightly, leading to negligible populations above rmsd  $\sim 3.0$  Å. The backbone rmsd of the first 15 residues of the predicted native conformation is  $\sim 0.8$  Å with respect to the NMR structure, and the main conformational differences appear in the last five residues, three of which are Pro residues. These last residues are bent with respect to the corresponding sequence in the NMR structure, and the Trp residue is then not properly packed in the predicted structure.

## V. CONCLUSIONS

Highly crowded aqueous environments, such as the cytosol, are characterized by the presence of many solute/water interfaces and large, nonuniform regions devoid of water. Such environments display unique properties that affect the structural and dynamic behavior of biomolecules, mainly due to electrostatic and liquid–structure forces not present in the bulk phase. Implicit models of solvation suitable for large-scale computer simulation of subcellular processes, for example, are necessarily phenomenological. However, careful design and parametrization can satisfactorily incorporate the most relevant physics of solvation. Such models are needed to study thermodynamic properties such as specific and nonspecific binding, protein translocation, structural stabilization and denaturation, and other essential biological/biochemical mechanisms. In this paper, an extension of the SCPISM model of solvation is proposed to incorporate two of the most relevant effects observed in crowded environments: the long-range water-exclusion forces and the effects of water-structure on the short-range forces between polar and charged groups. The model is parametrized on the basis of experimental data and on the strengths of HB interactions inferred from simulations in explicit water. The model provides a prescription to incorporate explicit ions and other cosolutes into the system.

## ■ ACKNOWLEDGMENT

This study was supported by the Intramural Research Program of the NIH, Center for Information Technology, and utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH, Bethesda, MD (<http://biowulf.nih.gov>). The authors thank Rick Venable, Aaron Bornstein, and Milan Hodoscek for computational help.

## ■ REFERENCES

- (1) Zimmermann, S. B.; Minton, A. P. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 27.
- (2) Luby-Phelps, K. *Int. Rev. Cytol.* **2000**, *192*, 189.
- (3) Ellis, R. J. *TIBS* **2001**, *26*, 597.
- (4) Halle, B. *Phil. Trans. R. Soc. London B* **2004**, *359*, 1207.
- (5) Mancinelli, R.; Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2959.
- (6) Parsegian, V. A. *Int. Rev. Cytol.* **2002**, *215*, 1.
- (7) Parsegian, V. A.; Rau, D. C. *J. Cell Biol.* **1984**, *99*, 196.
- (8) Jensen, T. R.; Ostergaard, M.; Reitzel, N.; Balashev, K.; Peters, G. H.; Kjaer, K.; Bjornholm, T. *Phys. Rev. Lett.* **2003**, *90*, 086101.
- (9) Zheng, J.-M.; Pollack, G. H. *Phys. Rev. E* **2003**, *68*, 031408.
- (10) Hoppert, M.; Mayer, F. *Am. Sci.* **1999**, *87*, 518.
- (11) Hoppert, M.; Braks, I. J.; Mayer, F. *FEMS Microbiol. Lett.* **1994**, *118*, 249.

- (12) Wichmann, C.; Naumann, P. T.; Spangenberg, O.; Konrad, M.; Mayer, F.; Hoppert, M. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 1104.
- (13) Larsen, A. E.; Grier, D. G. *Nature* **1997**, *385*, 230.
- (14) Kepler, G. M.; Fraden, S. *Phys. Rev. Lett.* **1994**, *73*, 356.
- (15) Crocker, J. C.; Grier, D. G. *Phys. Rev. Lett.* **1996**, *77*, 1897.
- (16) Xu, X.-H. N.; Yeung, E. S. *Science* **1998**, *281*, 1650.
- (17) Luby-Phelps, K. Cytoarchitecture and Physical Properties of Cytoplasm: Vol., Viscosity, Diffusion, Intracellular Surface Area. In *Microcompartmentalization and Phase Separation in Cytoplasm*; Walter, H., Brooks, D. E., Srere, P. A., Eds.; Academic Press: New York, 1999; Vol. 192.
- (18) Persson, E.; Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6266.
- (19) Qvist, J.; Persson, E.; Mattea, C.; Halle, B. *Faraday Discuss.* **2009**, *141*, 131.
- (20) Frolich, A.; Gabel, F.; Jasnin, M.; Lehnert, U.; Oesterhelt, D.; Stadler, M.; Tehei, M.; Weik, M.; Wood, K.; Zaccai, G. *Faraday Discuss.* **2009**, *141*, 117.
- (21) Tehei, M.; Franzetti, B.; Wood, K.; Gabel, F.; Fabiani, E.; Jasnin, M.; Zamponi, D.; Oesterhelt, D.; Zaccai, G. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 766.
- (22) Foster, K. R.; Resing, H. A.; Garroway, A. N. *Science* **1976**, *194*, 324.
- (23) Mastro, A. M.; Babich, M. A.; Taylor, W. D.; Keith, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3414.
- (24) Williams, S. P.; Haggel, P. M.; Briendle, K. M. *Biophys. J.* **1997**, *72*, 490.
- (25) Hassan, S. A. *J. Phys. Chem. B* **2005**, *109*, 21989.
- (26) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
- (27) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
- (28) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591.
- (29) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578.
- (30) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133.
- (31) Wagner, F.; Simonson, T. *J. Comput. Chem.* **1999**, *20*, 322.
- (32) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Comput. Chem.* **2004**, *25*, 1015.
- (33) Feig, M.; Brooks, C. L. *Curr. Opin. Struc. Biol.* **2004**, *14*, 217.
- (34) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47.
- (35) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struc. Biol.* **2009**, *19*, 120.
- (36) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6478.
- (37) Hassan, S. A.; Mehler, E. L.; Zhang, D.; Weinstein, H. *Proteins* **2003**, *51*, 109.
- (38) Hassan, S. A.; Mehler, E. L. *Int. J. Quantum Chem.* **2005**, *102*, 986.
- (39) Hassan, S. A. *J. Phys. Chem. B* **2007**, *111*, 227.
- (40) Bottcher, C. J. F. *Theory of Electric Polarisation*; Elsevier: Amsterdam, 1952.
- (41) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Elsevier: London, 1986.
- (42) Ehrenson, S. J. *Comput. Chem.* **1989**, *10*, 77.
- (43) Mehler, E. L. The Lorentz-Debye-Sack Theory and Dielectric Screening of Electrostatic Effects in Proteins and Nucleic Acids. In *Molecular Electrostatic Potential: Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier Science: Amsterdam, 1996; Vol. 3, p 371.
- (44) Hasted, J. B.; Ritson, D. M.; Collie, C. H. *J. Chem. Phys.* **1948**, *16*, 1.
- (45) Booth, F. J. *Chem. Phys.* **1951**, *19*, 391.errata: *ibidem* 19.
- (46) Gong, H. P.; Hocky, G.; Freed, K. F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11146.
- (47) Tomasi, J.; Perisco, M. *Chem. Rev.* **1994**, *94*, 2027.
- (48) Hassan, S. A.; Mehler, E. L. *Proteins* **2002**, *47*, 45.
- (49) Mehler, E. L.; Eichele, E. *Biochemistry* **1984**, *23*, 3887.
- (50) Jackson, J. D. *Classical Electrodynamics*, 2nd ed.; Wiley: New York, 1975.
- (51) Born, M. *Z. Phys.* **1920**, *1*, 45.
- (52) Latimer, W. M.; Rodebush, W. H. *J. Am. Chem. Soc.* **1920**, *42*, 1419.
- (53) Bucher, M.; Porter, T. L. *J. Phys. Chem.* **1986**, *90*, 3406.
- (54) Laidler, K. J.; Pegis, C. *Proc. R. Soc. London A* **1957**, *241*, 80.
- (55) Bucher, M. *J. Phys. Chem.* **1986**, *90*, 3411.
- (56) Rubinstein, A.; Sherman, S. *Biopolymers* **2007**, *87*, 149.
- (57) Cohen, B. E.; McAnaney, T. B.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. *Science* **2002**, *296*, 1700.
- (58) Rees, D. C. *J. Mol. Biol.* **1980**, *141*, 323.
- (59) Pitera, J. W.; Falt, M.; van Gunsteren, W. F. *Biophys. J.* **2001**, *80*, 2546.
- (60) Simonson, T.; Perahia, D.; Brunger, A. T. *Biophys. J.* **1991**, *59*, 670.
- (61) Hassan, S. A.; Mehler, E. L. Modeling Aqueous Solvent Effects through Local Properties of Water. In *Modeling Solvent Environments: Applications to Simulation of Biomolecules*; Feig, M., Ed.; Wiley-VCH: Weinheim, Germany, 2010.
- (62) Ben-Naim, A. *J. Phys. Chem.* **1990**, *94*, 6893.
- (63) Bruge, F.; Fornilli, S. L.; Malenkov, G. G.; Palma-Vittorelli, M. B.; Palma, M. U. *Chem. Phys. Lett.* **1996**, *254*, 283.
- (64) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198.
- (65) Timasheff, S. M. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67.
- (66) Parsegian, V. A.; Rand, R. P.; Rau, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3897.
- (67) Evans, R. *Adv. Phys.* **1979**, *28*, 143.
- (68) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6490.
- (69) Hassan, S. A. *J. Phys. Chem. B* **2004**, *108*, 19501.
- (70) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2925.
- (71) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, E. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8951.
- (72) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.
- (73) Weeks, J. D. *Annu. Rev. Phys. Chem.* **2002**, *53*, 533.
- (74) Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M.; Winter, G. *Nature* **1985**, *314*, 235.
- (75) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (76) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Borech, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodosek, M.; Im, W.; Kucera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
- (77) Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878.
- (78) Gabbouline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001**, *306*, 1139.
- (79) Hartley, R. W. *Biochemistry* **1993**, *32*, 5978.
- (80) Lee, L. P.; Tidor, B. *Nature* **2001**, *8*, 73.
- (81) Schreiber, G.; Fersht, A. R. *Biochemistry* **1993**, *32*, 5145.
- (82) Schreiber, G.; Fersht, A. R. *J. Mol. Biol.* **1995**, *248*, 478.
- (83) Urakubo, Y.; Ikura, T.; Ito, N. *Protein Sci.* **2008**, *17*, 1055.
- (84) Mallik, B.; Masunov, A.; Lazaridis, T. *J. Comput. Chem.* **2002**, *23*, 1090.
- (85) Hassan, S. A.; Mehler, E. L. *Int. J. Quantum Chem.* **2001**, *83*, 193.
- (86) Steinbach, P. J. *Proteins* **2004**, *57*, 665.
- (87) Qiu, L. L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952.
- (88) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587.



- (89) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.
- (90) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425.
- (91) Abagyan, R.; Totrov, M. *J. Mol. Bio.* **1994**, *235*, 983.
- (92) Li, Z. Q.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.