HHS Public Access

Author manuscript

J Comput Chem. Author manuscript; available in PMC 2015 November 10.

Published in final edited form as:

J Comput Chem. 2014 August 15; 35(22): 1621–1629. doi:10.1002/jcc.23655.

Implicit treatment of solvent dispersion forces in protein simulations

Sergio A. Hassan

Center for Molecular Modeling, DCB, CIT, National Institutes of Health, Bethesda, MD 20892

Abstract

A model is proposed for the evaluation of dispersive forces in a continuum solvent representation for use in large-scale computer simulations. It captures the short and long-range effects of water-exclusion in conditions of partial and anisotropic hydration. The model introduces three parameters, one of which represents the degree of hydration (water occupancy) at any point in the system, which depends on the solute conformation, and two that represent the strength of water-water and water-solute dispersive interactions. It is optimized for proteins, using hydration data of a sub-optimally hydrated binding site and results from dynamics simulations in explicit water. The model is applied to a series of aliphatic-alcohol/protein complexes and a set of binary and ternary complexes of various sizes. Implications for weak and ultra-weak protein-protein association and for simulation in crowded media are discussed.

I. Introduction

When a solute approaches another solute in an aqueous solution the water displaced is reorganized structurally and dynamically. The overall effect of water removal and reorganization can be divided into electrostatic and non-electrostatic forces. A continuum solvent representation for use in large-scale (i.e., long-time, big-size) simulations requires both effects to be properly incorporated into the force field, and to be computationally efficient as well. To this end, effective potentials have been developed that represent the solvent effects implicitly,[1–5] but these have focused mainly on the treatment of electrostatics. Non-electrostatic effects include solvent-induced forces (SIF) and dispersion forces (DF), both essential components of the hydration process. SIF are short-ranged, non-pairwise forces that originate in the rearrangements of the solute-water and water-water hydrogen-bond network in the solute hydration shells,[6–9] and thus modulate molecular interactions at close proximity. In purely nonpolar solutes, SIF are hydrophobic forces, for which several continuum theories have been proposed.[10–13] For polar/charged solutes the treatment of SIF is more complicated, especially if the surfaces are topographically irregular, as most proteins are.[14,15] SIF make important contributions to the intermolecular potentials of mean force,[16] and partially determine the height of the desolvation barriers and the strength of hydrogen bonds (HB) between hydrated groups.[5] A model has been proposed to account for HB modulation by SIF in protein simulations,[5,17] and an algorithm developed to incorporate the non-pairwise nature of SIF in Langevin dynamics.[18]

The focus here is on the implicit treatment of DF. These forces are weak but pervasive,[19] and their importance in implicit solvation has long been recognized.[20–23] The fundamental

role of DF in protein hydration energies,[21,24] protein-ligand interactions,[21,25–28] thermal stability of proteins,[29,30] and preferential hydration/binding underlying protein denaturation by cosolutes[31] have been probed experimentally and computationally. Increasing awareness of the importance of water-mediated DF has led to a number of extensions of implicit solvent models (ISMs) to incorporated DF in protein simulations.[32,33]

Because DF are attractive at all distances, failure to account for solvent DF may lead to overly compact structures of peptides and proteins, over-stabilization of non-native conformers, unphysical orientations and reduced fluctuations of side chains at protein surfaces, and stronger protein/ligand binding. In the latter case the problem is typically circumvented by neglecting the direct protein-protein dispersive energy altogether. This ad hoc solution is based on the notion that proteins and the water they displace upon association make equal contributions to the dispersive energy, and thus cancel out. This assumption is not always justified because the density of interfacial water varies substantially throughout the protein surface, and range from mild over-hydration,[34] to partial dewetting,[35] to significant dehydration in crevices and narrow pockets.[26] Indeed, it has been shown that dispersive interactions are the main contributions to the binding energy in sub-optimally hydrated binding sites as a result of uncompensated dispersive attraction upon binding.[26,27] In addition, DF between large solutes have a long-range, shape-dependent component[36] (e.g., ~1/L for spheres, and ~1/L^2 for planes, where L is the separation between the surfaces), which may result in an effective attraction or repulsion between the solutes in an aqueous solution,[28,37] depending on their densities and materials, as determined by the relative values of the Hamaker constants.[38] The strength of protein-water dispersion interactions has been estimated by computer simulations,[21] and it was shown that a significant contribution is made by atoms in the interior of the protein (for ubiquitin, a rather small protein, the protein-water van der Waals energy was estimated at ~45 kcal/mol[21]). Other simulation studies have shown similar results.[24] Although water and proteins are arguably similar organic substances, small differences in dispersive energies could rapidly add up as the system size increases, and could conceivably make a measurable contribution in multimeric complexes and assemblies. Such effective attraction may even exist between proteins and small molecules, as shown in μs-length simulations of urea binding to hen lysozyme.[31] The implications for the behavior of crowded, anisotropic media, such as those in the interior of a living cell could be far reaching, possibly affecting subcellular organization and molecular translocation.

The model introduced here corrects the aforementioned limitations of current ISMs, and is designed for large-scale simulations. Parameterization is based on results from dynamic simulations of barnase/barstar and cytochrome C peroxidase/cytochrome C complexes in explicit water, and on calorimetric data of a sub-optimally hydrated protein binding site. Results are presented for a set of binary and ternary complexes of different sizes, and for a series of aliphatic alcohol/MUP-I complexes.

II. Model

The simplest treatment of electrostatic and non-electrostatic forces in a continuum solvent representation is through a local, position-dependent permittivity ε and density ρ.[18,39] The

dependence of these functions at the solute/liquid interface is complicated, but suitable approximations based on classical theories of dielectrics[40] and liquid structure[41] can provide practical expressions for both.[3,18,39] In this case each element of liquid exerts a force on the solute, and the total force is obtained by integration over the space occupied by the liquid.[18] Improvements can be introduced empirically by incorporating non-local and other non-linear effects in ε and ρ.[39,42,43] Most ISMs used in macromolecular simulations assume ρ = 0, so the solute-solute DF tend to be stronger, especially for atoms close to the interface, where uncompensated forces are more apparent. The model proposed here adopts an intermediate approach, in which ρ ≠ 0 only at critical points, thus correcting for solvent DF at a fraction of computational cost.

A solute atom *i* in a liquid is subject to an average dispersion force $\mathbf{f} = |f|\mathbf{r}/r$ towards atoms within a volume *dv* at a position **r** (relative to *i*); an element *dv* at −**r** exerts an average force $\mathbf{f}' = -|f'|\mathbf{r}/r$ in the opposite direction. If both elements of volume are in average occupied by the same kind of substance, both forces cancel out. An imbalance of forces can result from different substances occupying **r** and −**r** or by different water densities in either position. In the absence of explicit water a major imbalance of forces will systematically occur throughout the system, which must be corrected. In addition, when a solute is assembled in solution the (dispersion) work necessary to move an atom *i* from bulk water to its final position in the solute is partially compensated by the work of removing water from the site in the opposite direction. These two independent processes are summarized in the cycle of fig. 1. A solute element *p* is moved from *i*′ in bulk water to *i* in the solute, whereas a water element *w* is moved from *i* to *i*′. The position *j* is occupied by another solute element, and site *k* is occupied by water with an occupancy s_k dependent on the solute conformation ($s \to 1$ in bulk water, and $s \to 0$ within the solute); positions *j*′ and *k*′ are both occupied by bulk water. If the dispersive energy between two elements of volume separated a distance *r* is $V^d = -2\varepsilon\sigma^6/r^6$ (with ε > 0) the energy change upon transfer, $\Delta V_{ij}^d = V_{ij}^d(p \text{ in } i; w \text{ in } i') - V_{ij}^d(w \text{ in } i; p \text{ in } i')$, is given by

$$\Delta V_{ij}^d = -2(\varepsilon_{ij}\sigma_{ij}^6 + \varepsilon_{ik}\sigma_{ik}^6 + \varepsilon_{wj'}\sigma_{wj'}^6 + \varepsilon_{wk'}\sigma_{wk'}^6 - \varepsilon_{wj}\sigma_{wj}^6 - \varepsilon_{wk}\sigma_{wk}^6 - \varepsilon_{i'j'}\sigma_{i'j'}^6 - \varepsilon_{i'k'}\sigma_{i'k'}^6)/r^6 \quad (1)$$

which is probably the simplest way to incorporate the dependence of the solvent dispersion energy on the solute conformation. All the σ's are here set equal to σ_{ij}, so water-exclusion effects are incorporated only through ε's. By setting $\varepsilon_{ik} = \varepsilon_{iw} s_k$, $\varepsilon_{wj} = \varepsilon_{wj} s_i$, and $\varepsilon_{wk} = \varepsilon_{ww} s_k s_i$, and because $\varepsilon_{wj'} = \varepsilon_{wk'} = \varepsilon_{ww}$ and $\varepsilon_{i'j'} = \varepsilon_{i'k'} = \varepsilon_{iw}$, the term in parenthesis in eq 1 can be cast into a single effective dispersive factor $\varepsilon_{ij}^s \sigma_{ij}^6$, where

$$\varepsilon_{ij}^s = \varepsilon_{ij} + \varepsilon'_{ij} = \varepsilon_{ij} + \{-\varepsilon_{wj} + (\varepsilon_{ww} - \varepsilon_{iw})[2 - s_k]\} \quad (2)$$

is the solvent-corrected strength of the direct (ε_{ij}) dispersive interaction between the two solute elements (s_i was set to unity, i.e., the site is assumed to be occupied by water prior to transferring the atom). Similar calculation for *j* in the presence of *i* yields (cf. fig. 1)

$$\varepsilon^{s}_{ji}=\varepsilon_{ji}+\varepsilon'_{ji}=\varepsilon_{ji}+\{-\varepsilon_{iw}+(\varepsilon_{ww}-\varepsilon_{wj})[2-s_q]\} \quad (2')$$

where s_q is the water occupancy at site q. To preserve the symmetry of the potential the effective strength of the pair ($\tilde{\varepsilon}_{ij}$) is obtained as the arithmetic mean of ε^{s}_{ij} and ε^{s}_{ji}, i.e.,

$$\tilde{\varepsilon}_{ij}=\varepsilon_{ij}+\tilde{\varepsilon}'_{ij} \quad (3)$$

where the solvent-corrected interaction is determined by $\tilde{\varepsilon'}_{ij} = [\varepsilon'_{ij} + \varepsilon'_{ij}]/2$. Because $\tilde{\varepsilon}_{ij}$ is symmetrical upon permutation of indexes i and j, the total van der Waals energy of a solute composed of N atoms is

$$E_{vdw}=E^{0}_{vdw}+V_{disp}=\frac{1}{2}\sum_{i\neq j}^{N}\varepsilon_{ij}\{(\sigma_{ij}/r_{ij})^{12}-2(\sigma_{ij}/r_{ij})^{6}\}-\sum_{i\neq j}^{N}\tilde{\varepsilon}'_{ij}(\mathbf{r})(\sigma_{ij}/r_{ij})^{6} \quad (4)$$

where E^{0}_{vdw} is the direct solute-solute van der Waals energy and V_{disp} is the correction to the dispersive energy due to the solvent. The dependence on the solute conformation, represented by $\mathbf{r} \equiv \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$, where $\mathbf{r}_i$ is the position of atom i, is made explicit in the last term.

An obvious choice for $s(\mathbf{r})$ in eqs 2 and 2′ is a function proportional to the local water density $\rho(\mathbf{r})$, for which any of the common approximations in liquid theory, including a simple barometric law probably suffices. For computational efficiency the approach used here is based on a contact model similar to that used in the development of the SCP implicit model of electrostatics,[5,44] namely,

$$s_k(\mathbf{r})=1-s_0^{-1}\sum_{l=1}^{N}\exp(-r_{kl}/\lambda) \quad (5)$$

where $r_{kl} \equiv |\mathbf{r}_l + \mathbf{r}_j - 2\mathbf{r}_i|$ is the distance between an atom l of the solute and the off-solute site $\mathbf{r}_k = 2\mathbf{r}_i - \mathbf{r}_j$ determined by the (i, j) pair; a similar equation holds for s_q, with $r_{ql} = |\mathbf{r}_l + \mathbf{r}_i - 2\mathbf{r}_j|$. Thus, $s \to 1$ for a point far from the solute (bulk water), and s_0 is chosen so that $s \to 0$ for a point in the interior of the solute. For proteins, s_0 can be estimated analytically[5] as $s_0 \approx 3(\zeta + 2\zeta^2 + 2\zeta^3)\exp(-1/\zeta)$, where $\zeta = \lambda/p_a$ and $p_a = (3V_p/4\pi N_p)^{1/3}$ ~1.4 Å is the average radius of a protein atom, and V_p ~183 Å^3 and N_P ~15 are the average volume and average number of atoms per residue, respectively, estimated from the PDB. The behavior of eq 5 is illustrated in fig. 2 for a sphere with a radius of 2 nm containing a cylindrical cavity of radius δ along its major axis. The sphere is composed of atoms distributed uniformly in a cubic lattice, with a number density N_P/residue. To illustrate the effects of cavity size on water occupancy, δ is varied from 1.4 Å (radius of a water molecule) to 7 Å (effective diameter of an amino acid). Figure 2 shows a grey-scale representation of $s(\mathbf{r})$ on a plane containing the cavity axis (left panel), along with the values of water occupancy along the axis (right panel) for λ = 3.5 Å. For a narrow cavity only its openings are partially accessible

to water, while the interior remains fully dehydrated ($s \sim 0$). As the cavity size increases the level of hydration increases from sub-optimal ($0 < s < 1$) to optimal ($s \sim 1$). In the current implementation the level of hydration is represented by the single parameter λ, which must be optimized using experimental or simulation data (cf. Section III). The extent of hydration inside the cavity depends not only on the thermodynamic conditions, the size and the topology of the pocket, but also on the chemical properties of the lining atoms. Improvement of the parameterization (not pursued here) would then require at least two λ's, one for non-polar groups and one for polar/charged groups.

Introducing the expression of s_k and s_q into eq 3 yields $\tilde{\varepsilon'}_{ij} = a_{ij} + b_i\Sigma\exp(-r_{kl}/\lambda) + b_j\Sigma\exp(-r_{ql}/\lambda)$ where $a_{ij} = \varepsilon_{ww} - \varepsilon_{iw} - \varepsilon_{wj}$; $b_i = (\varepsilon_{ww} - \varepsilon_{iw})/2s_0$; and $b_j = (\varepsilon_{ww} - \varepsilon_{wj})/2s_0$. For proteins, an additional simplification is introduced in which a single dispersion strength is assigned to all the atoms (represented by a generic index p), i.e., $\varepsilon_{iw} = \varepsilon_{pw}$. Given the known sensitivity of simulation results to Lennard-Jones parameters, as evidenced by the continuous effort in force field parameterizations based on chemical atom types,[45] this simplification may require revision. In addition, water occupancy at a given off-solute site is largely determined by the solute atoms close to the site, so the summation in eq 5 can be restricted to a subset N_k of atoms l within a distance d_c from k; likewise for a subset N_q around q. Finally, the parameter λ is made independent of k, so the final expression for $\tilde{\varepsilon'}_{ij}$ is

$$\tilde{\varepsilon}'_{ij}(\mathbf{r}) = a + b\sum_{l=1}^{N_k}\exp(-r_{kl}/\lambda) + b\sum_{l=1}^{N_q}\exp(-r_{ql}/\lambda) \quad (6)$$

This expression contains three parameters to be determined, either a, b, and λ; or ε_{ww}, ε_{pw}, and λ. In the latter approach, which is followed here, ε_{ww} and ε_{pw} have a physical meaning. In particular, these are positive quantities, and their values are expected to be close to those of LJ parameters in the protein force field. Their relative values determine whether the effective interactions between hydrated solutes are attractive or repulsive, as illustrated in fig. 3. All the potentials show a long-range decay.

The van der Waals force $\mathbf{F}_{vdw,i}$ on an atom i is calculated as $\mathbf{F}_{vdw,i} = -\nabla_i E_{vdw}$, where the gradient is taken with respect to $\mathbf{r}_i$ while keeping all the off-solute positions $\mathbf{r}_k$ fixed in the calculation of the partial derivatives. From eq 4 $\mathbf{F}_{vdw,i} = -\nabla_i E^0_{vdw} + \mathbf{F}_{disp,i}$, where the first term is the direct van der Waals force ($\mathbf{F}^0_{vdw,i}$) and the second term is the correction of the total dispersion force on i due to the solvent, given by

$$\mathbf{F}_{disp,i} = 12\sum_{j\neq i}^{N}\tilde{\varepsilon}'_{ij}(\mathbf{r})r_{ij}^{-1}(\sigma_{ij}/r_{ij})^6\hat{\mathbf{r}}_{ij} + \sum_{k=1}^{N(N-1)}\eta_{ik}\hat{\mathbf{r}}_{ik} \quad (7)$$

where $\hat{\mathbf{r}_{ij}} = (\mathbf{r}_j - \mathbf{r}_i)/|\mathbf{r}_j - \mathbf{r}_i|$ is the unit vector in the direction determined by the pair (i, j), and $\hat{\mathbf{r}_{ik}} = (\mathbf{r}_k - \mathbf{r}_i)/|\mathbf{r}_k - \mathbf{r}_i|$ is the unit vector determined by i and the off-solute site k. The first term in eq 7 has the usual pairwise form of a dispersion force determined by $\tilde{\varepsilon'}_{ij}$, and originates in the presence of solvent at sites k and q. The quantity η in the second sum of eq 7 is given by

$$\eta_{ik}=(2b/\lambda)(\sigma_k/r_k)^6\theta(d_c-r_{ik})\exp(-r_{ik}/\lambda) \quad (8)$$

and originates in the changes of solvent occupancy at site k due to changes in the position of atom i; θ is the Heaviside step function, and the index k is determined by the corresponding (m, n) atom pairs, i.e., $\mathbf{r}_k = 2\mathbf{r}_m - \mathbf{r}_n$, and $\mathbf{r}_{ik} = \mathbf{r}_k - \mathbf{r}_i$, so $\sigma_k \equiv \sigma_{nm}$ and $r_k \equiv r_{nm}$. In practice, the cutoff distance d_c can be made rather small, so most of the $N(N-1)$ terms in eq 7 vanish.

III. Results

Isothermal titration calorimetry (ITC) of a series of aliphatic alcohols bound to the major urinary protein (MUP-I) showed that the binding enthalpy is determined mainly by ligand-protein dispersive interactions.[26,27] This is an example of suboptimal hydration, hence an ideal system for optimization of λ in eq 5. The crystal structures of the complexes (cf. caption fig. 7) show a narrow binding pocket located deep into MUP-I (fig. 4A) and lined mostly with nonpolar residues. These structures are used here to calculate the average water occupancy within the binding pocket and the binding enthalpies of the complexes. For computational efficiency the summation in eq 5 is restricted to C_α atoms only, so s_0 is determined by p_a = 3.5 Å. This simplification was used previously to treat long-range water-exclusion effects on electrostatics.[5] The protein is immersed into a cubic lattice with a cell length of 1.4 Å, and s calculated with eq 5 at each point of the lattice. The density of water within the binding pocket was estimated[26] at ~0.21 g/cm^3. An average occupancy <s> ~0.2 is well reproduced by eq 5 with λ = 2 Å, although s is heterogeneous throughout the pocket (fig. 4B). In general, increasing λ lowers s. This value of λ was estimated with no cutoffs in eq 5, but <s> remains unchanged with d_c as small as ~12 Å.

The optimized λ is used next to obtain ε_{ww} and ε_{pw} in eq 6. Figure 5 depicts a generalization of the cycle of fig. 1, in which two proteins (1 and 2) are dissociated from their bound state. The van der Waals contribution to the total dissociation energy is $\Delta E_{vdw} = \Delta V + \Delta V_r$, where ΔV is the energy required to separate the proteins without affecting the behavior of water (an ideal process), and ΔV_r is the reorganization energy of water in contact with the protein. A' and B' are regions of bulk water with the same shapes and volumes as proteins 1 and 2, whereas A and B are the same regions of (non-bulk) water in contact with the proteins. After evaluating ΔV and ΔV_r the energy change can be approximated as

$$\Delta E_{vdw}=(E_{BB}-E_{B'B'})+(E_{AA}-E_{A'A'})+E_{1B}+E_{2A}-E_{A'B'}-E_{12} \quad (9)$$

where E_{IJ} is the van der Waals interaction energy between subsystems I and J, and the terms with I = J are self-energies; all the terms corresponding to interactions of the proteins with bulk water (E_{1w} and E_{2w}) cancel out, and the differences ($E_{Aw} - E_{A'w}$) and ($E_{Bw} - E_{B'w}$) are small compared to the other terms in eq 9, thus neglected. A system in which the protein interfaces are well hydrated prior to binding appears to be the barnase/barstar complex, for which the binding enthalpy at 25 °C and pH 7 has been measured[46] at ~19.3 kcal/mol. Each term of eq 9 is here calculated for this complex using molecular dynamic simulations in explicit water, so ΔE_{vdw} and the individual contributions can be estimated directly. The all-atom representation of the CHARMM protein force field[45] (version c35b5) with default

protonation states (pH 7) was used, resulting in an electrically neutral system. The simulations were carried out in the canonical ensemble at 25 °C and 1 atm, using periodic boundary conditions and the CHARMM-optimized TIP3P water model. A cubic simulation cell with a side length of ~9.33 nm was used, large enough to accommodate both bulk and non-bulk (interfacial) water around the complex and the unbound proteins. Long-range electrostatics was treated with PME summations,[47] with specifications reported previously.[16] Bond lengths and angles were constrained with SHAKE, and a 2-fs time step was used (leapfrog integrator). Pressure was maintained with the Langevin piston method,[48] with a mass of 400 amu and a collision frequency of 20 ps^{-1}; temperature was fixed with the Hoover thermostat,[49] using a mass of 10^3 kcal mol^{-1} ps^2. NPT-equilibration of pure water yielded an average density of ~0.993 g/cm^3.

The conformations of the proteins were taken from the crystal structure of the complex (PDB id 1brs). To calculate E_{12} (1 = barnase; 2 = barstar) the complex was initially centered in the simulation cell and water molecules overlapping protein atoms were removed. To calculate E_{1B}, E_{2A}, E_{AA}, and E_{BB} the unbound proteins were positioned within the simulation cell in the same positions they occupied in the simulation of the complex. Throughout the simulations all the C_α atoms were constrained at fixed positions, and the system equilibrated for 5 ns. To calculate $E_{A'A'}$, $E_{B'B'}$ and $E_{A'B'}$ only the dynamics of water was simulated. The regions of non-bulk (A, B) and bulk (A', B') water were defined by the molecular surfaces of the corresponding proteins, using the Lee-Richards algorithm (probe radius set to zero). The error in the estimate of ΔE_{vdw} is chosen to be $\Delta\Delta E_{vdw}$ ~2 kcal/mol, which is the expected accuracy of the force field. Therefore, each term in eq 9 was evaluated with a statistical error $\Delta E = \sigma/n^{1/2}$ ~0.25 kcal/mol, where σ is the standard deviation of the corresponding term E_{IJ} obtained from the fluctuating energy $E_{IJ}(t)$, and n is the number of uncorrelated steps given by $n = \tau/\tau_c$, where τ is the total simulation time and τ_c the correlation time estimated from the autocorrelation function $C(t) \sim \int E_{IJ}(t')E_{IJ}(t+t')dt'$. To decrease the errors to the required values each term in eq 9 required a different simulation length τ, ranging from 10 to 20 ns. For the calculation of the energies from the trajectories PME was abandoned, and all the non-bonded terms were calculated directly without truncation. All the energies reported below are in kcal/mol. The changes in water self-energies upon reorganization were practically identical despite the different amounts of water in each region: $E_{AA} - E_{A'A'} = 12.5 \pm 0.8$ and $E_{BB} - E_{B'B'} = 12.7 \pm 0.8$. Both energies are also positive, implying unfavorable rehydration. The similarity of self-energies appears to be related to the number of water molecules “in contact” with the protein surfaces, which are the ones affected the most by the reorganization at the interfaces. Thus, region A has a number of water molecules in contact with protein 2, which is similar to the number of water molecules in B in contact with 1. However, this similarity is not expected to be a general feature, and larger differences would likely be observed in the presence of more complex surfaces. There is also a slight increase in the average number of water molecules within each region: $n_A = 668 \pm 10$ and $n_{A'} = 663 \pm 10$; and $n_B = 544 \pm 9$ and $n_{B'} = 542 \pm 9$, where Δn are now standard deviations. From a thermodynamic standpoint, these changes are related to the differences in chemical potential (μ) induced by the surfaces. Thus, μ of water in A′ cannot be equal to μ_w of the bulk water, but μ in A becomes equal to μ_w once all the structural rearrangements (including water absorption) set in; similar for B and B′. The

changes in self-energies within each region can be partially explained by an increase in water packing induced by the protein surfaces. The interaction terms are: $E_{2A} = -38.5 \pm 0.1$, $E_{1B} = -41.2 \pm 0.1$, $E_{A'B'} = 7.3 \pm 0.1$, and $E_{12} = -62.9 \pm 0.1$. The resulting van der Waals contribution is thus $\Delta E_{vdw} = 1.1 \pm 2$ kcal/mol, a small attraction but within the statistical error of the calculation. Therefore, in this case where water reorganization at the interface is modest and the proteins are relatively small, $\Delta E_{vdw} \sim 0$ is justified.

The parameterization requires a second protein complex. To partially incorporate the effect of protein size, the complex is chosen on the basis of the following criteria: One of the proteins must be similar in size to barnase (or barstar) and the second protein must be larger than barstar; in addition, the proteins must be globular and relatively rigid, with a topographically simple protein/protein interface to also expect near-optimal pre-binding hydration. Cytochrome C peroxidase contains the same number of residues than barnase; and its ligand, cytochrome C, is ~3.5 times larger than barstar. The binding enthalpy of this complex (PDB id: 2pcc) has been measured at ~2.6 kcal/mol. Dynamic simulations were performed using the same protocol described above. The van der Waals contribution to dissociation was estimated at $\Delta E_{vdw} = 0.5 \pm 2$ kcal/mol, also weakly attractive but still within the statistical error.

The dispersive contributions to the dissociation energies are now calculated using eqs 4 and 6 with the optimized λ. Robust methods exist to evaluate the relative binding affinities of small ligands for a rigid binding site,[50] including free energy perturbation and thermodynamic integration. Accurate estimates of absolute free energies are more challenging, even for small system, and no reliable methods are yet available for proteins, although progress is being made.[51] The method employed here is simple, yet robust enough for the purpose of parameterization. The conformations of the complexes are first relaxed using canonical MC simulations of the proteins in the bound state at 25 °C. The average energy is calculated[52] as $\langle E_b \rangle = Z^{-1} \Sigma_i E_i \exp(-E_i/kT) \approx \Sigma_i E_i/n$, where E_i and n in the last sum are the energy of the accepted conformation i and the number of accepted conformations, respectively. The energy is given by $E_i = E_{elec,i} + E_{vdw,i} + E_{hydr,i}$, each term corresponding, respectively, to electrostatic, van der Waals, and hydrophobic contributions. Electrostatic interactions are calculated with the SCP model.[5,53] In this model the electrostatic energy E_{elec} of a solute composed of N atoms is given by,

$$E_{elec} = \frac{1}{2}\sum_{i \neq j}^{N} \frac{q_i q_j}{r_{ij} D_{ij}(r_{ij};\mathbf{r})} + \frac{1}{2}\sum_{i=1}^{N} \frac{q_i^2}{R_i(q_i;\mathbf{r})} \left\{ \frac{1}{D_i[R_i(q_i;\mathbf{r});\mathbf{r}]} - 1 \right\} \quad (10)$$

where **r** denotes the conformation of the system, and q are the atomic charges. In the partition of eq 10 the first sum is the interaction energy term, and the second sum is the self-energy term. The dependence of E_{elec} on the conformation is through the screening functions D and the effective radii R, as both of these quantities account for solvent-exclusion effects at different length scales.[5] The effects of SIF on HB energies are introduced[5] through a suitable modification of R, whereas the hydrophobic energy is given by a cavity term $E_{\text{hydr}} = a\gamma$, where γ is the solute solvent-accessible surface-area, and a is a

parameter obtained from hydration energies of small alkanes.[53] The functions *D*, *R*, and γ are all estimated with a contact model similar to eq 5.

Optimization of the model requires adjusting ε_{ww} and ε_{pw} in such a way that the association energy $\Delta E_d = \langle E_b \rangle - E_\infty$ equals the experimental binding enthalpy at same temperature. This procedure would require a fully-converged MC simulation for each parameter set over the course of the optimization. To simplify the computational demand only one structure in the bound state is used, which is chosen from a MC simulation carried out with the non-corrected model ($E_{vdw}=E^0_{vdw}$ in eq 4). A member *k* of the canonical ensemble is selected such that E_k is closest to the average $\langle E_b \rangle$ of the ensemble. The representative structure *k* is subjected to a short minimization to relax it with the dispersion-corrected force field. For each values of ε_{ww} and ε_{pw} the dissociation energy is calculated simply as $\Delta E_d = E_k - E_\infty$, a major simplification but good enough for the optimization. Then, for each (ε_{ww}, ε_{pw}) pair, the van deer Waals dissociation energy $\Delta E_{vdw} = E_{vdw,k} - E_{vdw,\infty}$ is compared to ΔE_{vdw} obtained from the explicit-water simulations. Calculations are also carried out with the all-atom CHARMM force field,[45] using standard protonation states, and no truncations of the non-bonded interactions.

Figure 4C shows the region in the ε_{ww}–ε_{pw} plane within which ΔE_{vdw} (in kcal/mol) equals the dissociation energies. For each complex the iso-energy contour plots are straight lines, and lines corresponding to different complexes intersect, forming an extended rhomboid-shaped region. Any combination of ε_{ww} and ε_{pw} within this region can be chosen as optimized parameters; in particular the values at the center, namely, ε_{ww} ~0.2548 kcal/mol and ε_{pw} ~0.2707 kcal/mol. These values make physical sense as they are within the order of magnitude of the LJ parameters of the CHARMM force field. These values are used to estimate the van der Waals contributions to the binding energies of a set of binary and ternary protein complexes of varying sizes. The procedure is the same as described above for 1brs and 2pcc; for the ternary complex the dissociated state corresponds to the three components separated from one another. Results are shown in fig. 6. The uncorrected energies (solid circles) are large in all cases, but correction by water dispersion destabilize the complexes significantly (open circles). Yet, in all but one of the complexes the corrected van der Waals contributions are attractive. The degree of stabilization does not correlate with complex size (at least within this range of sizes considered here) but appears to be related more closely to the structural details of the protein/protein interfaces. For the binary complexes the contribution is negative ($\Delta E_{vdW} < -2$ kcal/mol) for complexes in which a specificity-conferring residue (Arg or Lys) in one protein is buried in the other protein (1tgs, 2pts, 1brc, 1ppe). The interfaces of the other complexes are less tight, including 1a0o, which shows a rather low degree of complementarity. This reinforces the notion that dispersion in the presence of topographically complex interfaces characterized by crevices and narrow pockets could make important contributions to binding. This is illustrated by the aliphatic-alcohols/MUP-I protein complexes. Binding energies are calculated with the ISM at $T = 27$ °C, using the same protocol used for the protein complexes. Figure 7A shows the components of the binding energy without corrections for water dispersion effects. The hydrophobic attraction increases with chain length, as expected, but it is compensated by an unfavorable electrostatic energy. This leaves only the direct van der Waals interaction (E^0_{vdw}

in eq 4) as the main contribution to the binding energy. This energy, however, is almost twice as large as the experimental enthalpy[27] ΔH_{exp}, a consequence of uncompensated dispersive interaction. The partial weakening of the binding energy is due to the solvent dispersion correction (V_{disp} in eq 4), as shown in fig. 7B. To reproduce the experimental enthalpy within ~2 kcal/mol accuracy the parameter ε_{pw} for the alcohol atoms has to be adjusted to ~0.2120 kcal/mol. Using the same value of ε_{pw} obtained for proteins, the van der Waals contribution is underestimated by ~6 kcal/mol, which demonstrates the sensitivity of the results to the LJ parameters; nonetheless, the linear behavior with the chain length is still reproduced. As mentioned, a single ε_{pw} is unlikely to be optimal for all chemical species, and different values are needed for amino acids, nucleic acids, alcohols, fatty acids, and other basic units. Ultimately, an atom-type-based parameterization would probably be required, in line with the parameterization of the LJ potential in the force field. Another reason to recalibrate ε_{wp} for alcohols is that the C_α-based calculation of s can no longer be used; in this case the sum in eq 5 included all the C atoms of the aliphatic chain, which underestimates the values of s if the same λ is used.

IV. Discussion and Conclusions

Dispersion forces have been shown to play a role in protein-ligand interactions,[25–28] the conformational stability of proteins,[29,30] and the preferential hydration/binding mechanism underlying denaturation by cosolutes.[31] The importance of incorporating DF in an implicit solvent model has been established.[20,22,23] Progress has been made in recent years, and models have been proposed that incorporate solvent DF empirically.[32,33] The treatment of long-range effects and the dependence on the solute conformation are important physical features that still need careful consideration.

In this paper a model was proposed which incorporates solvent-corrected DF through effective interactions between the solute atoms. The model captures the short- and long-range effects of water-exclusion in conditions of partial and anisotropic hydration. It was developed in the context of a classical (molecular mechanics) force field, where electrostatics and van der Waals interactions are treated independently[36] (although both are electrostatic in origin). The solvent correction is introduced through the LJ coefficients ε (eq 6). The extent of water exclusion is represented in ε through a conformation-dependent water-occupancy function $s(\mathbf{r})$ (eq 5), using a contact model similar to that used earlier in the treatment of water-exclusion in electrostatics.[5,44] In the current implementation ε are made independent of temperature; such dependence can eventually be incorporated through s. The model is simple yet general, as its formulation does not depend on the nature of the solute. In combination with the implicit electrostatic model developed previously[5,44,53] (which includes effects of SIF) it forms a complete ISM, as it captures the main physical effects of partial, anisotropic hydration common in real biological systems. Improvements are probably needed both in the treatment of s and in the parameterization. The current implementation uses only three parameters, one of which (λ) represents the degree of hydration and two others that represent the strength of water-water (ε_{ww}) and solute-water (ε_{pw}) dispersive interactions. These parameters were optimized for proteins, using hydration data of a sub-optimally hydrated binding site (MUP-I), and results from explicit-water MD

simulations of two medium-sized complexes (barnase/barstar and cytochrome C peroxidase/cytochrome C).

The model was applied to a series of aliphatic-alcohol/MUP-I complexes for which ITC data suggest dispersive interactions to be the major contributions to the binding enthalpy. The model reproduces the linear increase of affinity with chain length. It was found, however, that (favorable) hydrophobic and (unfavorable) electrostatic energies are not negligible, but compensate each other for all the chain lengths, thus leaving dispersion as the main contribution. The model was also used to estimate the binding energies of several protein complexes of various sizes, demonstrating that a significant reduction of the dispersive attraction is achieved with the current implementation. Need for improvements can only be determined by MD simulations in explicit water, and accomplished through atom-type-based parameterization of ε_{pw}.

To account for long-range interactions no cutoffs were used to evaluate either eq 4 or eq 5, so the computational cost of the current implementation makes the model more suitable for peptides and small proteins. However, several strategies can be used to speed up computation of the long-range contributions, in line with the treatment of long-range electrostatics in the SCP model.[5] These include suitable truncations of eqs 4 and 5; the use of Hamaker-like potentials between distant residues; infrequent updates of eq 3 during Langevin dynamics; and neglect of the 3-body terms in eq. 7. Depending on how these approximations are implemented the CPU time can be improved substantially, to a maximum of about ~2–4 times the computational cost required in the gas phase; these approximations, however, require careful re-parameterization, and will be the topic of a future study.

In general, aqueous interfaces display non-bulk behavior that can extend up to a few nanometers into the bulk, depending on the surface size, topography, and charge distribution.[5] In particular, the density of interfacial water can vary from mild over-hydration,[34] to partial dewetting,[35] to significant dehydration in crevices and narrow pockets.[26] The imbalance of DF introduced by density variations at the interface could affect the association/dissociation mechanism. Furthermore, dispersive interactions have a long-range component that may lead to effective attraction/repulsion between proteins in large complexes and assemblies. Small differences in the strength of protein-protein and protein-water DF could make significant contributions in weak and ultra-weak protein association.[52,54] The implications in crowded, anisotropic environments, such as those in the interior of a living cell, can be far reaching. Evidence suggests that a large proportion of biologically active molecules in the cell do not diffuse freely in the bulk medium despite their high solubility, but are transiently bound to one another, to membranes, or to the cytoskeleton.[55] In addition to electrostatics, long-range uncompensated DF may play a role in such non-specific interactions, possibly affecting subcellular organization and molecular translocation.

Acknowledgments

This study utilized the high-performance computer capabilities of the Biowulf PC/Linux cluster at the NIH, and was supported by the NIH Intramural Research Program through CIT.

References

1. Gilson MK, Davis ME, Luty BA, McCammon JA. Computation of Electrostatic Forces on Solvated Molecules Using the Poisson-Boltzmann Equation. J Phys Chem. 1993; 97:3591.
2. Cramer CJ, Truhlar DG. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. Chem Revs. 1999; 1999:2161. [PubMed: 11849023]
3. Tomasi J, Mennucci B, Cammi R. Quantum mechanical continuum solvation models. Chem Rev. 2005; 105:2999. [PubMed: 16092826]
4. Feig M, Brooks CL. Recent Advances in the Development and Application of Implicit Solvent models in Biomolecule Simulations. Curr Op Struc Biol. 2004; 14:217.
5. Hassan SA, Steinbach PJ. Water-Exclusion and Liquid-Structure Forces in Implicit Solvation. J Phys Chem B. 2011; 115:14668. [PubMed: 22007697]
6. Ben-Naim A. Solvent-Induced Forces in Protein Folding. J Phys Chem. 1990; 94:6893.
7. Bruge F, Fornilli SL, Malenkov GG, Palma-Vittorelli MB, Palma MU. Solvent-Induced Forces on a Molecular Scale: Non-Additivity, Modulation and Causal Relation to Hydration. Chem Phys Lett. 1996; 254:283.
8. Durell SR, Brooks BR, Ben-Naim A. Solvent-Induced Forces Between Two Hydrophilic Groups. J Phys Chem. 1994; 98:2198.
9. Hassan SA. Amino Acid Side Chain Interactions in the Presence of Salts. J Phys Chem B. 2005; 109:21989. [PubMed: 16479276]
10. Lum K, Chandler D, Weeks JD. Hydrophobicity at Small and Large Length Scales. J Phys Chem B. 1999; 103:4570.
11. Chandler D. Interfaces and the Driving Force of Hydrophobic Assembly. Nature. 2005; 437:640. [PubMed: 16193038]
12. Pratt LR. Molecular theory of Hydrophobic Effects: She is too mean to have her name repeated. Annu Rev Phys Chem. 2002; 53:409. [PubMed: 11972014]
13. Hummer G, Garde S, Garcia AE, Pratt EA. New Perspectives on Hydrophobic Effects. Chem Phys. 2000; 258:349.
14. Tan ML, Cendagorta JR, Ichiye T. Effects of Microcomplexity on Hydrophobic hydration in Amphiphiles. J Amer Chem Soc. 2013; 135:4918. [PubMed: 23506339]
15. Giovambattista N, Lopez CF, Rossky PJ, Debenedetti PG. Hydrophobicity of protein surfaces: Separating geometry from chemistry. Proc Nat Acad Sci (USA). 2008; 105:2274. [PubMed: 18268339]
16. Hassan SA. Intermolecular Potentials of Mean Force of Amino Acid Side Chain Interactions in Aqueous Medium. J Phys Chem B. 2004; 108:19501.
17. Hassan SA, Guarnieri F, Mehler EL. Characterization of Hydrogen Bonding in a Continuum Solvent Model. J Phys Chem B. 2000; 104:6490.
18. Hassan SA. Liquid-structure Forces and Electrostatic Modulation of Biomolecular Interactions in Solution. J Phys Chem B. 2007; 111:227. [PubMed: 17201447]
19. Parsegian, VA. Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists. Cambridge University Press; New York: 2005.
20. Floris F, Tomasi J. Evaluation of the Dispersion Contribution to the Solvation Energy: a Simple Computational Model in the Continuum Approximation. J Comput Chem. 1989; 10:616.
21. Pitera JW, van Gunsteren WF. The importance of solute-solvent van der Waals Interactions with interior atoms of biopolymers. J Amer Chem Soc. 2001; 123:3163. [PubMed: 11457039]
22. Zacharias M. Continuum Solvent Modeling of Nonpolar Solvation: Improvement by separating Surface Area dependent Cavity and Dispersion Contributions. J Phys Chem A. 2003; 107:3000.
23. Wagoner JA, Baker NA. Assessing Implicit Models for Nonpolar Mean Solvation Forces: the Importance of Dispersion and Volume Terms. Proc Nat Acad Sci (USA). 2006; 103:8331. [PubMed: 16709675]
24. Levy RM, Zhang LY, Gallicchio E, Felts AK. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. J Amer Chem Soc. 2003; 125:9523. [PubMed: 12889983]

25. Chapman KT, Still WC. A Remarkable Effect of Solvent Size on the Stability of a Molecular Complex. J Amer Chem Soc. 1989; 111:3075.
26. Barrat E, Bingham RJ, Warner DJ, Laughton CA, Phillips SEV, Homans SW. Van der Waals interactions Dominate Ligand-Protein Association in a Protein Binding Site Occluded from Solvent Water. J Amer Chem Soc. 2005; 127:11827. [PubMed: 16104761]
27. Malham R, Johnstone S, Bingham RJ, Barratt E, Phillips SEV, Laughton CA, Homans SW. Strong Solute-Solute Dispersive Interactions in a Protein-Ligand Complex. J Amer Chem Soc. 2005; 127:17061. [PubMed: 16316253]
28. Janshoff A, Neitzert M, Oberdorfer Y, Fuchs H. Force Spectroscopy of Molecular Systems. Single Molecule Spectroscopy of Polymers and Biomolecules. Angew Chem Int Ed. 2000; 39:3212.
29. Vondrasek J, Kubaz T, Jenney FE, Adams MWW, Kozisek M, Cerny J, Sklenar V, Hobza P. Dispersion Interactions Govern the Strong Thermal Stability of a Protein. Chem Eur J. 2007; 13:9022. [PubMed: 17696186]
30. Berka K, Hobza P, Vondrasek J. Analysis of Energy Stabilization inside the Hydrophobic Core of Rubredoxin. Chem Phys Chem. 2009; 10:543. [PubMed: 19170065]
31. Hua L, Zhou R, Thirumalai D, Berne BJ. Urea Denaturation by Stronger Dispersion Interactions with Proteins than Water Implies a 2-Stage Unfolding. Proc Nat Acad Sci (USA). 2008; 105:16928. [PubMed: 18957546]
32. Gallicchio E, Levy RM. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. J Comp Chem. 2004; 25:479. [PubMed: 14735568]
33. Dzubiella J, Swanson JMJ, McCammon JA. Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models. Phys Rev Lett. 2006; 96:087802. [PubMed: 16606226]
34. Merzel F, Smith JC. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? Proc Nat Acad Sci (USA). 2002; 99:5378. [PubMed: 11959992]
35. Jensen TR, Ostergaard M, Reitzel N, Balashev K, Peters GH, Kjaer K, Bjornholm T. Water in Contact with Extended Hydrophobic Surfaces: Direct Evidence of Weak Dewetting. Phys Rev Lett. 2003; 90:086101. [PubMed: 12633443]
36. Israelachvili, JN. Intermolecular and Surfaces Forces. 3. Academic Press; San Diego, CA: 2011.
37. Hamaker HC. The London-van der Waals attraction between spherical particles. Physica. 1937; 10:1058.
38. Hunter CA. Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. Angew Chem Int Ed. 2004; 43:5310.
39. Hassan SA. Self-consistent treatment of the local dielectric permittivity and electrostatic potential in solution for polarizable macromolecular force fields. J Chem Phys. 2012; 137:074102. [PubMed: 22920098]
40. Bottcher, CJF. Theory of Electric Polarisation: Dielectrics in Static Fields. 2. Vol. 1. Elsevier; 1973.
41. Hansen, J-P.; McDonald, IR. Theory of Simple Liquids. 2. Elsevier; London: 1986.
42. Liszi J, Meszaros L, Ruff I. The field dependence of the Kirkwood factor and the nonlinear dielectric behavior of some liquids. J Chem phys. 1981; 74:6896.
43. Rubinstein A, Sherman S. Influence of the Solvent Structure on the Electrostatic Interactions in Proteins. Biophys J. 2004; 87:1544. [PubMed: 15345535]
44. Hassan SA, Mehler EL, Zhang D, Weinstein H. Molecular Dynamics Simulations of Peptides and Proteins with a Continuum Electrostatic Model Based on Screened Coulomb Potentials. Proteins. 2003; 51:109. [PubMed: 12596268]
45. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The Biomolecular Simulation Program. J Comp Chem. 2009; 30:1545. [PubMed: 19444816]

46. Frisch C, Schreiber G, Johnson CM, Fersht AR. Thermodynamics of the interaction of Barnase and Barstar: changes in free energy versus changes in enthalpy on mutation. J Mol Biol. 1997; 267:696. [PubMed: 9126847]
47. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. J Chem Phys. 1995; 103
48. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant Pressure Molecular Dynamics Simulations: the Langevin Piston Method. J Chem Phys. 1995; 103:4613.
49. Hoover WG. Canonical Dynamics: Equilibrium Phase-Space Distributions. Phys Rev A. 1985; 31:1695. [PubMed: 9895674]
50. Gilson MK, Given JA, Bush BL, McCammon JA. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. Biophys J. 1997; 72:1.
51. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct. 2007; 36:21. [PubMed: 17201676]
52. Cardone A, Pant H, Hassan SA. Specific and non-specific protein association in solution: computation of solvent effects and prediction of first-encounter modes for efficient configurational bias Monte Carlo simulations. J Phys Chem B. 2013; 117:12360. [PubMed: 24044772]
53. Hassan SA, Guarnieri F, Mehler EL. A General Treatment of Solvent Effects Based on Screened Coulomb Potentials. J Phys Chem B. 2000; 104:6478.
54. Clore GM, Tang C, Iwahara J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. Curr Op Struc Biol. 2007; 17:603.
55. Luby-Phelps K. Cytoarchitecture and Physical Properties of Cytoplasm: Volume, Viscosity, Diffusion, Intracellular Surface Area. Int Rev Cytol. 2000; 192:189. [PubMed: 10553280]

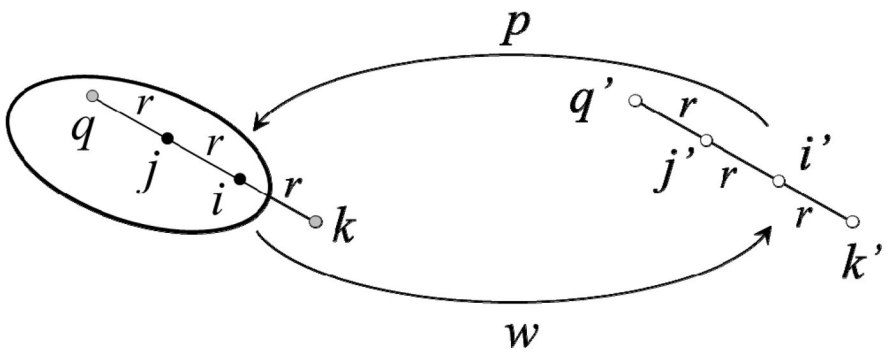

Figure 1.

Formation of a solute in water. A solute element p is transferred from bulk water (site i') to its final position (i) in the growing solute, while water w is moved from i to i'. Site j at a distance r from i is occupied by solute; sites k and k' contain water with occupancies $s_k \leq 1$ and $s_{k'} = 1$; likewise for q and q'; s_i and s_j are assumed to be unity.

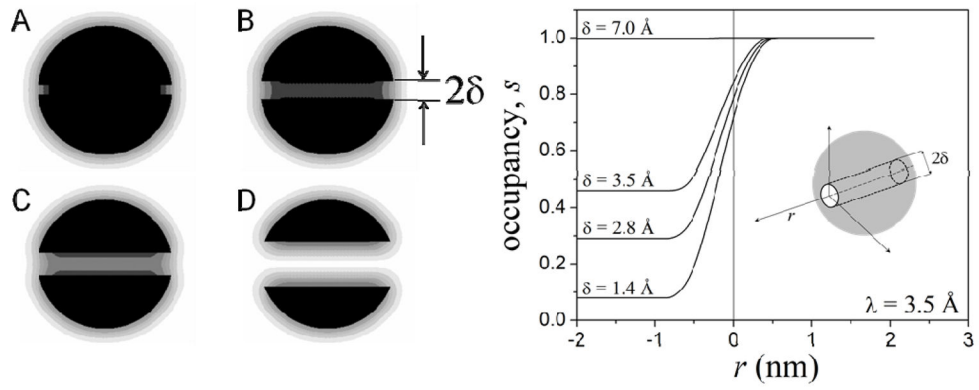

Figure 2.

Left panel: gray-scale representation of water occupancy *s* (eq 5) in a sphere containing a cylindrical cavity of radius δ along one of its major axes (black: *s* = 0; white: *s* = 1). The system mimics a typical globular protein with a diameter of ~4 nm, containing a binding site with varying degrees of water accessibility. The sphere is formed by identical atoms distributed in a cubic lattice with a constant of 1.4 Å; λ in eq 5 is 3.5 Å. (A) δ = 1.4 Å (radius of a water molecule); only the openings of the cavity are accessible to water; (B) δ = 2.8 Å; hydration at the center of the cavity remains significantly reduced; (C) δ = 3.5 Å (effective radius of an amino acid in a protein); hydration reaches half the value of bulk water; (D) δ = 7 Å; near optimal hydration. Right panel: water occupancy as a function of the distance *r* along the cavity axis for different cavity radii δ; *r* = 0 is at the entrance of the cavity (inset).

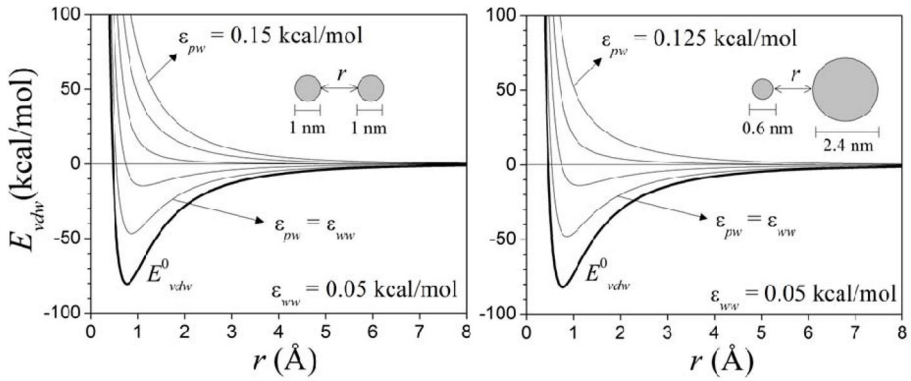

Figure 3.

van der Waals interaction energy E_{vdw} (eq 4) between two spheres in water, as a function of the distance r between their surfaces, and for different values ε_{ww} and ε_{pw} (eq 6). The spheres are composed of identical atoms distributed uniformly on a cubic lattice with a constant of 1.4 Å. Lennard-Jones parameters are ε_{ij} = 0.15 kcal/mol and σ_{ij} = 2.5 Å (LB mixing rules used). The energies at equilibrium are similar (within ~2 kcal/mol) in both systems for the same values of ε_{ww} and ε_{pw}. The direct potentials (E^0_{vdw}; for $\varepsilon_{ww} = \varepsilon_{pw} = 0$) corresponding to the uncorrected interactions are also shown.

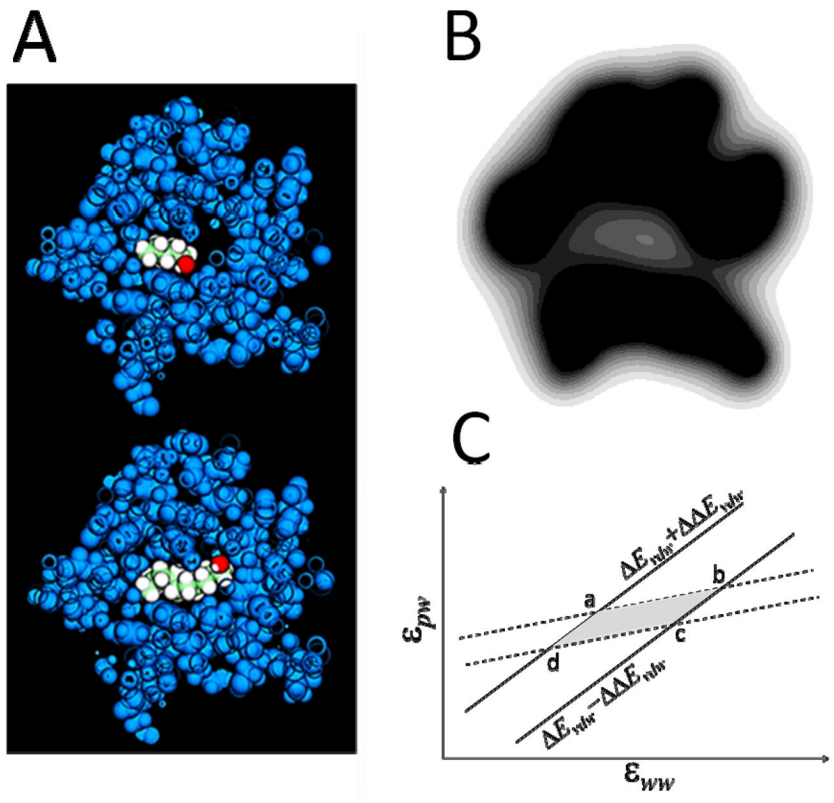

Figure 4.

(A) Cross section of the pentan-1-ol/MUP-I (upper) and decan-1-ol/MUP-I (lower) complexes, showing the buried cavity and the alcohol molecules filling the space (PDB 1znd used). (B) Gray-scale representation of water occupancy *s* on a (arbitrary) plane across the sub-optimally hydrated binding cavity. To be compared with fig. 2. Water occupancy was calculated with $\lambda = 2$ Å in eq 5, which yields $\langle s \rangle \sim 0.2$ inside the binding pocket, based on reported results.[27] (C) Iso-energy lines of the van der Waals dissociation energies of barnase/barstar (dashed lines) and cytochrome C peroxidase/cytochrome C (solid) complexes in the ε_{ww}–ε_{pw} plane, determined by the statistical errors (~2 kcal/mol) estimated from MD simulation in explicit water. The overlapping section is defined by the following (ε_{ww}, ε_{pw}) pairs: a = (0.2252, 0.2501); b = (0.4313, 0.4074); c = (0.2824, 0.2898); d = (0.0763, 0.1325); all in kcal/mol.

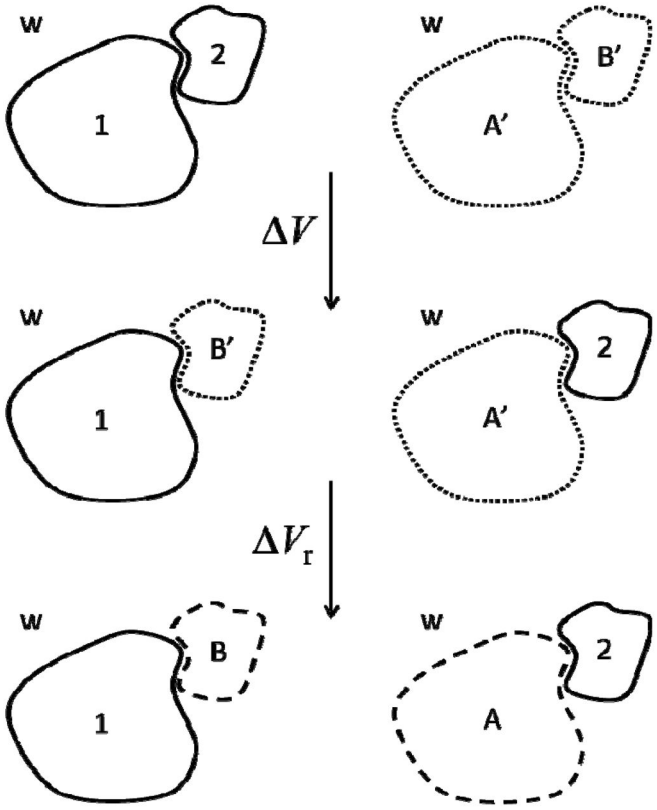

Figure 5.

Thermodynamic cycle used to estimate the total van der Waals contribution to protein-protein binding in water used in the molecular dynamics simulations. *A*′ and *B*′ are regions of bulk water with the same sizes and shapes of proteins 1 and 2, respectively; whereas *A* and *B* are the same regions of water in contact with the proteins 2 and 1 upon structural and dynamic reorganization (see text).

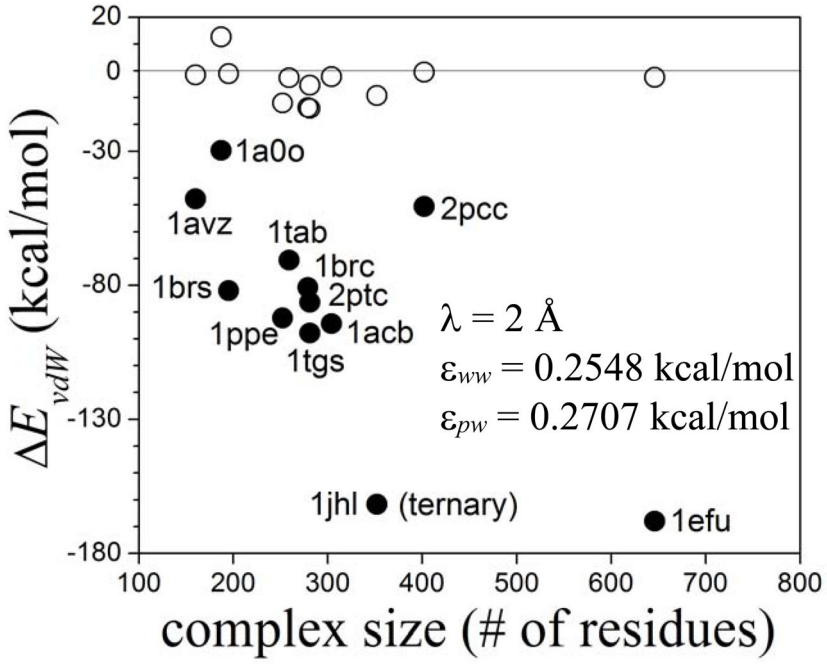

Figure 6.

Water-dispersion effects of the binding energies of a number of binary and ternary protein complexes of varying sizes (identified by PDB id) as estimated with the ISM. Van der Waals energy of binding (ΔE_{vdw}); solid symbols: uncorrected dispersive interactions ($E_{vdw}=E^0_{vdw}$ in eq 4); open: corrected dispersion ($E_{vdw}=E^0_{vdw}+V_{disp}$). Barnase/barstar (1brs) and cytochrome C peroxidase/cytochrome C (2pcc) used in the optimization are also shown. The proteins include enzymes bound to substrates, inhibitors, or activators; and an antigen-antibody complex.

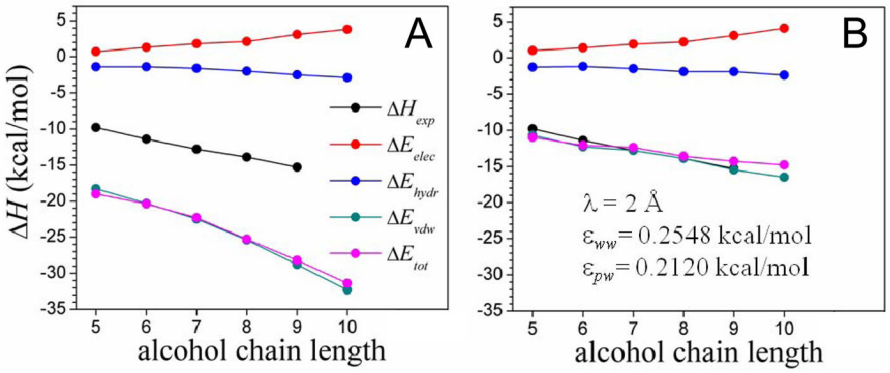

Figure 7.

Energy contributions to the binding energy of the MUP-I protein to a series of aliphatic alcohols as a function of the chain length, calculated with the implicit solvent model (E_{elec} given by eq 10; E_{vdw} by eq 4, and E_{hydr} is the cavity-formation energy). (A) Uncorrected dispersion energy ($E_{vdw}=E^{0}_{vdw}$ in eq 4); and (B) solvent-dispersion corrected energy ($E_{vdw}=E^{0}_{vdw}+V_{disp}$). The structures of the complexes are taken from the PDB; accession codes 1znd, 1zne, 1zng, 1znh, 1znk, 1znl, for pentan-, hexan-, heptan-, octan-, nonan-, and decan-1-ol, respectively. The experimental binding enthalpy[26] (ΔH_{exp}) is also shown (ITC data for decan-1-ol not available). Binding is dominated by dispersive forces, as determined by the extent of sub-optimal hydration of the binding pocket.