J|A|C|S
ARTICLES
Published on Web 07/11/2003

On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute−Solvent Interaction Energy

Ronald M. Levy,* Linda Y. Zhang, Emilio Gallicchio, and Anthony K. Felts

Contribution from the Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854

Received December 19, 2002; E-mail: ronlevy@lutece.rutgers.edu

Abstract: Implicit solvent hydration free energy models are an important component of most modern computational methods aimed at protein structure prediction, binding affinity prediction, and modeling of conformational equilibria. The nonpolar component of the hydration free energy, consisting of a repulsive cavity term and an attractive van der Waals solute−solvent interaction term, is often modeled using estimators based on the solvent exposed solute surface area. In this paper, we analyze the accuracy of linear surface area models for predicting the van der Waals solute−solvent interaction energies of native and non-native protein conformations, peptides and small molecules, and the desolvation penalty of protein−protein and protein−ligand binding complexes. The target values are obtained from explicit solvent simulations and from a continuum solvent van der Waals interaction energy model. The results indicate that the standard surface area model, while useful on a coarse-grained scale, may not be accurate or transferable enough for high resolution modeling studies of protein folding and binding. The continuum model constructed in the course of this study provides one path for the development of a computationally efficient implicit solvent nonpolar hydration free energy estimator suitable for high-resolution structural and thermodynamic modeling of biological macromolecules.

1. Introduction

Hydration phenomena play an important role in virtually every process occurring in aqueous solution. Hydration has a particularly large effect on the thermodynamics of processes involving the breakage or formation of noncovalent bonds. The accurate description of hydration thermodynamics is therefore essential in the prediction of protein structures, ligand binding free energies, and conformational equilibria.[1−5]

Explicit solvent models provide the most detailed and complete description of hydration phenomena.[6] They are, however, computationally demanding because of the large number of atoms involved and the need to average over many solvent configurations to obtain meaningful thermodynamic parameters. Implicit solvent models[7] offer an attractive alternative to explicit solvent models. They have been shown to be useful for applications including protein decoy recognition,[8,9] small molecule hydration free energy prediction,[10−12] and binding affinity prediction.[13] Early empirical surface area models of hydration based on surface area accessibility[14,15] have been shown to be of limited accuracy.[16] In a typical modern implicit solvent model, the solvation free energy is decomposed into a nonpolar component and an electrostatic component.[7] The nonpolar component corresponds to the free energy of hydration of the uncharged solute and the electrostatic component to the free energy of turning on the solute partial charges. The recent advances in improving the applicability of implicit solvent models have focused on the electrostatic component and have been primarily fueled by the development of accurate and computationally efficient continuum dielectric models.[10,17−19]

Modeling of the nonpolar component has received less attention despite the fact that it is the dominant term whenever hydrophobic interactions[20] are important. The formation of micelles and phospho-lipid membranes and their mechanism

(1) Honig, B.; Yang, A. *Adv. Protein Chem.* **1995**, *46*, 27−58.
(2) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.
(3) Apostolakis, J.; Ferrara, P.; Calfish, A. *J. Chem. Phys.* **1999**, *110*, 2099−2108.
(4) Lazaridis, T.; Karplus, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139−145.
(5) Felts, A.; Gallicchio, E.; Wallqvist, A.; Levy, R. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 404−422.
(6) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531−67.
(7) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1.
(8) T. Lazaridis, M. K. *J. Mol. Biol.* **1999**, *288*, 477−487.
(9) Wallqvist, A.; Gallicchio, E.; Felts, A. K.; Levy, R. M. In *Computational Methods for Protein Folding: A Special Volume of Advances in Chemical Physics*, Vol. 120; Friesner, R., Ed.; John Wiley & Sons: New York, 2002 (I. Prigogine and S. A. Rice, series editors).
(10) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(11) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.
(12) Gallicchio, E.; Zhang, L.; Levy, R. *J. Comput. Chem.* **2002**, *23*, 517−529.
(13) Zhou, R.; Friesner, R.; Ghosh, A.; Rizzo, R.; Jorgensen, W.; Levy, R. *J. Phys. Chem.* **2001**, *105*, 10388−10397.
(14) Ooi, T.; Oobatake, M.; Nemethy, G.; Sheraga, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.
(15) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1993**, *232*, 639−659.
(16) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 262−344.
(17) Zhang, L.; Gallicchio, E.; Friesner, R.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591−607.
(18) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(19) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, C. W. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(20) Ben-Naim, A. *Hydrophobic Interactions;* Plenum Press: New York, 1980.

of interaction with plasma and membrane bound proteins involve the hydration of hydrophobic molecular assemblies.[21] The structure and properties of proteins in water are highly influenced by hydrophobic interactions.[1,2,22−23] Hydrophobic interactions also play a key role in the mechanism of ligand binding to proteins.[24−27]

Empirical surface area models of nonpolar hydration are widely used.[10,11,28−35] They commonly estimate the total nonpolar solvation free energy (cavity formation and dispersion interaction) by a linear relationship between the nonpolar free energy, ΔG_{np}, and the solute surface area, A,

$$\Delta G_{np} = \gamma A + b \tag{1}$$

where A is the solute surface area. The surface tension proportionality constant γ represents the contribution to the solvation free energy per unit surface area. The constant b is the free energy of hydration for a point solute ($A = 0$). Surface area models have been justified on the basis of theoretical considerations[36] and on the experimental observation that vacuum-to-water and oil-to-water transfer free energies of the normal alkanes are linearly related to the surface areas.[37−39] Surface area models have served as a useful first approximation; some of their deficiencies have been discussed previously.[30,40−42] Furthermore, the proposed value of the surface tension proportionality constant in conjunction with nonpolar solvation surface area models vary by more than 1 order of magnitude. They range from 5 cal/mol/Å^2 [10] to 138 cal/mol/Å^2,[43] corresponding to the various definitions of solute surface area (van der Waals surface, molecular surface or solvent accessible surface[44]), the different origins of experimental data to which the model is parametrized[30,45,46] and the applicability range of the model (e.g., small molecule solvation,[11] protein folding,[1] and binding[47]). In this paper, we examine the behavior of surface area models for the nonpolar hydration free energy of proteins in different contexts: native proteins, misfolded and unfolded proteins, and ligand and protein binding.

We begin by observing that the solvation properties of hydrophobic species are determined by the volume and shape of the excluded solvent volume and by their attractive van der Waals interactions with the solvent. The hydrophobic nonpolar hydration free energy is then decomposed as

$$\Delta G_{np} = \Delta G_{cav} + \Delta G_{vdW} \tag{2}$$

where ΔG_{cav} is the cavity hydration free energy, defined as the hydration free energy due to excluded volume effects, and ΔG_{vdW} is the free energy for establishing the solute−solvent van der Waals dispersion interactions. This equation implies that the two effects can be studied independently[48] by subdividing the total solvation process into two steps. In the first step, a suitable cavity is created in the solvent; in the second step, the attractive interactions between the solute and the solvent are established.[49,50]

The decomposition of the nonpolar free energy into a cavity term and an attractive dispersion energy term has a long history; the theoretical study of cavity formation in water has been frequently used as a model for studying the hydrophobic effect. Studies of the solvation free energy of a cavity in water have related the free energy change to the cavity volume for small cavities and the surface area for larger cavities;[36,51] recent theory and simulation suggest that the crossover occurs for spheres with radii around 1 nm.[41,52] While theory[42,53] and computer simulations[42,54−56] show that solvation free energy changes associated with certain perturbations of molecular cavities scale as the surface area, the relationship may, in general, be more complex.[41] We expect that the free energy of cavity formation, which depends only on the size and shape of the cavity, will be better described by purely geometrical parameters, such as volume, surface area, and surface curvature, than the solute−solvent dispersion attraction term which depends also on the density, location, and nature of the solute atoms that are placed in the cavity. In this paper, we focus on the solute−solvent dispersion term and analyze the degree of correlation between the solute−solvent dispersion energy of proteins and peptides with their solvent accessible surface area (SASA). As discussed in the following section, deviations of the solute−solvent dispersion energy from the predictions of linear surface area models will be reflected in the total nonpolar hydration free energy.

Gallicchio et al.[42] have shown that the free energy change, ΔG_{vdW}, to establish the solute−solvent dispersion interactions for a set of alkanes of similar size depends mainly on the atomic composition of the solute and not on their surface area. This observation helped to explain the smaller hydration free energies of cyclic alkanes compared to the linear alkanes; it also helped to explain why two very different values of the surface tension

(21) Tanford, C. *The hydrophobic effect: formation of micelles and biological membranes;* Wiley: New York, 1973.
(22) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1−63.
(23) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660−679.
(24) Sturtevant, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 2236−2240.
(25) Williams, D. H.; Searle, M. S.; Mackay, J. P.; Gerhard, U.; Maplestone, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1172−78.
(26) Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−1301.
(27) Siebert, X.; Hummer, G. *Biochemistry* **2002**, *41*, 2965−2961.
(28) Lee, M.; Duan, Y.; Kollman, P. *Proteins* **2000**, *39*, 309−316.
(29) Hünenberger, P.; Helms, V.; Narayana, N.; Taylor, S.; McCammon, J. *Biochemistry* **1999**, *38*, 2358−2366.
(30) Simonson, T.; Brünger, A. *J. Phys. Chem.* **1994**, *98*, 4683−4694.
(31) Still, W. C. Tempczyk, A.; Hawley, R. C.; Hendrikson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(32) Rapp, C. S.; Friesner, R. A. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 173−183.
(33) Fogolari, F.; Esposito, G.; Viglino, P.; Molinari, H. *J. Comput. Chem.* **2001**, *22*, 1830−1842.
(34) Pellegrini, E.; Field, M. J. *J. Phys. Chem. A* **2002**, *106*, 1316−1326.
(35) Curutchet, C.; Cramer, C.; Truhlar, D.; Ruiz-Lòpez, M.; Rinaldi, D.; Orozco, M.; Luque, F. *J. Comput. Chem.* **2003**, *24*, 284−297.
(36) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717−26.
(37) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
(38) Chothia, C. *Nature* **1974**, *248*, 338.
(39) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2925.
(40) Wallqvist, A.; Covell, D. G. *J. Phys. Chem.* **1995**, *99*, 13118−13125.
(41) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.
(42) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271−6285.
(43) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *J. Am. Chem. Soc.* **1999**, *121*, 9243−9244.
(44) Pascual-Ahuir, J.; Silla, E. *J. Comput. Chem.* **1990**, *11*, 1047−1060.
(45) Tanford, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 4175−4176.
(46) Sitkoff, D.; Sharp, K. A.; Honig, B. *Biophys. Chem.* **1994**, *51*, 397−409.
(47) Elcock, A.; Sept, D.; McCammon, J. *J. Phys. Chem. B* **2001**, *105*, 1504−1518.
(48) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683−3704.
(49) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1980**, *73*, 3434−41.
(50) Widom, B. J. *Chem. Phys.* **1982**, *86*, 869.
(51) Irisa, M.; Nagayama, K.; Hirata, F. *Chem. Phys. Lett.* **1993**, *207*, 430−5.
(52) Huang, D. M.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324−8327.
(53) Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. *Phys. Chem. B* **1998**, *102*, 10469−82.
(54) Pohorille, A.; Pratt, L. R. *J. Am. Chem. Soc.* **1990**, *112*, 5066−74.
(55) Wallqvist, A.; Berne, B. J. *J. Phys. Chem.* **1994**, *99*, 2885−2892.
(56) Wallqvist, A.; Berne, B. J. *J. Phys. Chem.* **1995**, *99*, 2893−2899.

parameter are necessary to reproduce the hydration free energies and conformational equilibria of the alkanes.[43,57] A weak correlation between SASA and solute−solvent interaction energy was also observed by Pitarch et al.[58] who reported that the interaction energy between the methane dimer and water is nearly independent of the methane−methane separation distance. These observations have inspired the development of a nonpolar functional form that, together with a parameterization of the generalized Born model, has been shown to reproduce the experimental hydration free energies of a large set of small polar and nonpolar molecules with very high accuracy.[12]

It is expected that the solute−solvent dispersion interaction energy approximately tracks the solvent accessible surface area. The solute−solvent energy increases with increasing solute size, and the surface area grows with solute size; in addition, the atoms on the solute surface interact more strongly with the solvent and contribute the most to the solute−solvent energy. Furthermore, given that the distribution of atoms on the protein surface is roughly homogeneous, it is expected that a single proportionality constant may be sufficient to approximately predict the protein−solvent dispersion energy based on the solvent accessible surface area of the protein. Pitera et al.,[59] however, have recently shown that the protein−water van der Waals interaction energy due to atoms not directly exposed to the solvent is substantial. They have also shown that the contribution due to the loss of protein−water van der Waals interaction energy to protein−protein binding free energies is significantly affected by the inclusion of buried atoms.

In the following section, we construct a continuum solvent model for the protein−solvent van der Waals dispersion energy which is able to accurately reproduce the results of explicit solvent simulations. We first present the solvation free energy decomposition scheme we employ. We then proceed to calculate solute−water dispersion energies from explicit solvent calculations of peptides and proteins in water. The data gathered from the simulations is used to optimize and test this continuum solvent model for the solute−solvent van der Waals energies. The continuum solvent model is then used to calculate the solute−water dispersion energies of a large number of folded and misfolded conformations of peptides and proteins, as well as the solute−water dispersion energy contribution to the free energy of binding of various protein−ligand and protein−protein complexes. For comparison, the corresponding accessible surface area models are analyzed. We examine why surface area models appear to describe nonpolar solvation energies in a coarse grained sense but exhibit poor transferability and are not accurate enough for high-resolution studies.

2. Methods

2.1. Hydration Free Energy Decomposition. The solvation free energy of a molecule is defined as the free energy change for transferring a molecule from the gas phase to the solution. The transfer process can be decomposed into a series of steps: first, the atomic partial charges and van der Waals interactions of the solute are removed, and then the resulting solute cavity is transferred into aqueous solution. Then, the solute's van der Waals interactions and partial charges are restored.[10,12,30,31,60] This is illustrated in terms of a thermodynamic cycle (see Figure 1) whereby the total solvation free energy can be expressed as

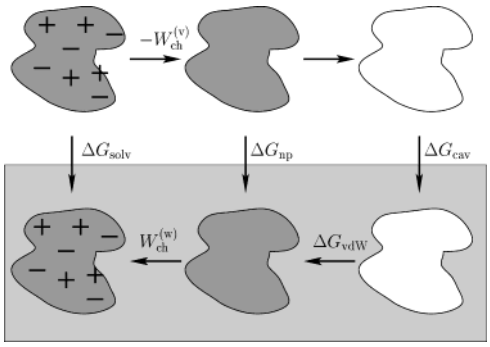

Figure 1. Thermodynamic cycle depicting the decomposition of the solvation free energy into electrostatic and nonpolar components. The sequence of steps starting from the upper left and moving clockwise is as follows: uncharging of the solute in vacuum, removal of the solute−solvent van der Waals interaction in vacuum (there is no free energy change associated with this step because of the lack of solvent molecules), hydration of the solute cavity, establishment of the solute−solvent van der Waals interactions in water, and charging of the solute in water.

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} = \Delta G_{elec} + \Delta G_{cav} + \Delta G_{vdW}, \quad (3)$$

where ΔG_{elec} is the electrostatic contribution to the solvation free energy, which is the difference between the work of charging the solute in solution, $W_{chg}^{(w)}$, and the work of charging the solute in a vacuum, $W_{chg}^{(v)}$, ΔG_{cav} is the cavity hydration free energy, and ΔG_{vdW} is the free energy for establishing the solute−solvent van der Waals dispersion interactions. In this paper, we discuss the proper modeling of the nonpolar term ΔG_{vdW}.

The solute−solvent site−site dispersion interaction energy is modeled with the standard 12-6 Lennard-Jones pair potential, u_{LJ}. The Lennard-Jones pair potential is decomposed into repulsive and attractive components using the Weeks−Chandler−Andersen (WCA) decomposition scheme[61,62] setting

$$u_{LJ}(r) = u_{rep}(r) + u_{vdW}(r), \quad (4)$$

where

$$u_{rep}(r) = \begin{cases} u_{LJ}(r) + \epsilon & r \le 2^{1/6}\sigma \\ 0 & r > 2^{1/6}\sigma \end{cases} \quad (5)$$

$$u_{vdW}(r) = \begin{cases} -\epsilon & r \le 2^{1/6}\sigma \\ u_{LJ}(r) & r > 2^{1/6}\sigma \end{cases}, \quad (6)$$

where σ and ϵ are the radius and well-depth parameters of the Lennard-Jones potential, $u_{rep}(r)$ is the short-ranged repulsive portion of the standard LJ potential, and $u_{vdW}(r)$ is the attractive van der Waals dispersion interaction potential.

The nonpolar solvation process is then naturally divided into two steps: the formation of the repulsive solute cavity in which the solute atoms interact with the solvent by way of the repulsive potential $u_{rep}(r)$, followed by the addition of the attractive van der Waals interactions $u_{vdW}(r)$. Explicit solvent simulations of the solvation thermodynamics of small alkanes using the WCA decomposition of the Lennard-Jones potential have shown[42] that the free energy change, ΔG_{cav}, in the first

(57) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *Biophys. J.* **1998**, *75*, 755−768.
(58) Pitarch, J.; Moliner, V.; Pascual-Ahuir, J.; Silla, A.; Tuñón, I. *J. Phys. Chem.* **1996**, *100*, 9955−9959.
(59) Pitera, J. W.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 3163−3164.
(60) McCammon, J. A.; Straatsma, T. P. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.
(61) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237−47.
(62) Chandler, D.; Weeks, J. D.; Andersen, H. C. *Science* **1983**, *220*, 787−794.

step is approximately proportional to the solvent accessible surface area, whereas the free energy change, ΔG_{vdW}, in the second step depends on the number, location, and nature of the atomic interaction centers of the solute. For a set of hexane rotamers, it was shown that ΔG_{vdW} is in fact independent of the solute accessible surface area.[42,43]

We have shown in earlier work[42] that, for small alkanes, the solute−solvent van der Waals interaction energy accurately reproduces the free energy to "turn on" the solute−solvent dispersion interactions (i.e., to convert a soft cavity into an alkane) as predicted by Pratt and Chandler.[49] In the recent theory of hydrophobicity at multiple length scales by Lum, Chandler, and Weeks,[41] the solute−solvent attractions play a similar perturbative role for larger solutes as well;[63] that is, the free energy contribution of adding the dispersion interactions may be approximated by the interaction energy in the presence of the dispersion term. The structural basis of this is contained in the study by Wallqvist et al.[64] Therefore, we focus on the protein−solvent van der Waals interactions as a surrogate for the corresponding part of the protein solvation free energy. In this study, we approximate the free energy change ΔG_{vdW} by the corresponding energy term U_{vdW}. In particular, the free energy change for the process of adding the solute−solvent van der Waals dispersion interactions to the solute cavity is modeled by the average solute−solvent WCA attractive potential energy when the solute interacts with the solvent including the full set of Lennard−Jones interactions. None of the conclusions of this work are significantly affected by this approximation; if the average van der Waals solute−solvent interaction energy is poorly correlated with the accessible surface area, so are the free energy changes ΔG_{vdW} and ΔG_{np}.

2.2. Explicit Solvent Simulations. Explicit solvent molecular dynamics (MD) simulations were carried out with the IMPACT program[65] using the OPLS all-atom force field[66] and the TIP4P[67] water model. The solute molecules were kept rigid and their atomic partial charges set to zero. A water box was constructed around the solute. The size of the water box was adjusted according to solute size by ensuring that the solute was surrounded by at least three layers of water molecules along any direction. Periodic boundary conditions were applied. The spherical cutoff for nonbonded interactions was set to 15.5 Å. The simulations were performed in the constant pressure and constant temperature ensemble. The temperature and pressure were set to 298.15 K and 1 atm, respectively. Full Lennard-Jones solute−solvent pair potentials were employed to generate system conformations. The solute−water box was first equilibrated for at least 54 ps using a 1 fs time step for the initial 24 ps and a 2 fs time step for the remaining 20 ps. A 2 fs MD time step was used during the data collection run. Trajectories were collected every 200 MD steps for at least 60 ps. The average solute−solvent van der Waals energies were calculated by energy analysis of the saved MD trajectories.

2.3. Continuum Solute−Solvent van der Waals Energy Model. Explicit solvent simulations can be used to obtain the average solute−solvent van der Waals potential energy, but they are also computationally very demanding. To obtain data for a large number of systems, we have developed a fast continuum solvent model and parametrized it against explicit solvent results in water.

The model assumes that the average water oxygen number density outside the solute volume is constant. The average van der Waals solute−water energy, $U_{vdW}(i)$ of atom i located at $\mathbf{r}_i$ is therefore written as the integral of the van der Waals solute atom−water oxygen pair potential over the solvent region

$$U_{vdW}(i) = \int_{solvent} \rho_w u_{vdW}^{(i)}(|\mathbf{r} - \mathbf{r}_i|) d^3\mathbf{r} \quad (7)$$

where ρ_w is the water bulk number density (at standard conditions $\rho_w = 0.0336$ Å^{-3}), $u_{vdW}^{(i)}$ is the solute−solvent van der Waals pair potential for atom i defined by eq 6 with $\sigma = \sigma_{iw}$ and $\epsilon = \epsilon_{iw}$, where σ_{iw} and ϵ_{iw} are respectively the diameter and well depth parameters of the solute atom−water oxygen Lennard-Jones pair potential. The van der Waals interactions between solute atoms and water hydrogen atoms are neglected as the TIP4P[67] water model used in the explicit solvent simulations prescribes. The total average van der Waals solute−solvent interaction energy U_{vdW} is obtained by summing over the atoms of the solute

$$U_{vdW} = \sum_{i=1}^{n} U_{vdW}^{(i)} \quad (8)$$

where n is the number of atoms of the solute.

For the convenience of numerical computation, the integration domain of eq 7 is converted from the unbounded solvent region to the bounded solute region obtaining[68]

$$U_{vdW}^{(i)} = U_{vdW}^{(i)}(\text{isolated}) - \rho_w \int_{solute} \theta(|\mathbf{r} - \mathbf{r}_i| - R_i) u_{vdW}^{(i)}(|\mathbf{r} - \mathbf{r}_i|) d^3\mathbf{r} \quad (9)$$

where the first term is the solute−solvent van der Waals energy when the solute is composed solely of atom i. This term is obtained in analytic form by direct integration of eq 7 for a single atom i with radius R_i. The second term in eq 9 is the integral over the solute volume outside atom i. This term represents the effects of the displacement of the solvent around atom i due to the presence of the other solute atoms. Its effect is to reduce the solute−solvent van der Waals energy of the atom from the value corresponding to the isolated atom. The second term in eq 9 is evaluated numerically using a spherical adaptive grid of points centered around atom i inside the solute volume but outside atom i.

The parameters ϵ_{iw} and σ_{iw} are obtained from the OPLS all-atom force field[66] using the combination rules $\sigma_{iw} = \sqrt{\sigma_i \sigma_w}$ and $\epsilon_{iw} = \sqrt{\epsilon_i \epsilon_w}$, where σ_i and ϵ_i are the OPLS σ and ϵ Lennard-Jones parameters of atom i and $\sigma_w = 3.153\,65$ Å and $\epsilon_w = 0.155$ kcal/mol are the OPLS Lennard-Jones parameters for the oxygen atom of the TIP4P[67] water model.

The location of the boundary between the solute region and the solvent region is a required parameter in most implicit continuum solvent models.[10,12,17,31,68] In our model, the solute region is the region enclosed by a set of spheres of radii R_i centered on each solute atom. The nucleii of the water oxygen atoms are excluded from the solute region in analogy with the definition of the solvent accessible surface. The density of water molecules in the solvent region is assumed uniform. The radius R_i of atom i is set as $\sigma_i/2$, where σ_i is the OPLS σ Lennard-Jones parameter for atom i, plus an adjustable water probe radius r_w.

The solvent probe radius r_w is the only adjustable parameter of the model. The best value for the adjustable water probe radius was determined by minimizing the root-mean-square deviation between the continuum solute−solvent van der Waals energy model predictions and the solute−solvent van der Waals energies obtained from explicit solvent simulations. To this end, explicit solvent simulations were performed for nine native protein conformations (1fc2, 2ovo, 4pti, 1fas, 1shg, 2cro, 1tul, 1lz1, and 2aza(a)), two misfolded conformations of 1ctf (a1977 and d8787[69]), and 6 conformations of the octadecapeptide Ace-GEWTYDDATKTFTVTE-Nme from the β-hairpin C-terminal fragment of the B1 domain of protein G (PDB id 1gb1).[70] The value

(63) Huang, D.; Chandler, D. *J. Phys. Chem. B* **2002**, *106*, 2047−2053.
(64) Wallqvist, A.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 6745−6753.
(65) Kitchen, D. B.; Hirata, F.; Kofke, D. A.; Westbrook, J. D.; Yormush, M.; Levy, R. M. *J. Comput. Chem.* **1990**, *11*, 1169−1180.
(66) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(67) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.
(68) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.
(69) Park, B.; Levitt, M. *J. Mol. Biol.* **1996**, *258*, 367−392.
(70) García, A.; Sanbomatsu, K. *Proteins* **2001**, *42*, 345−354.

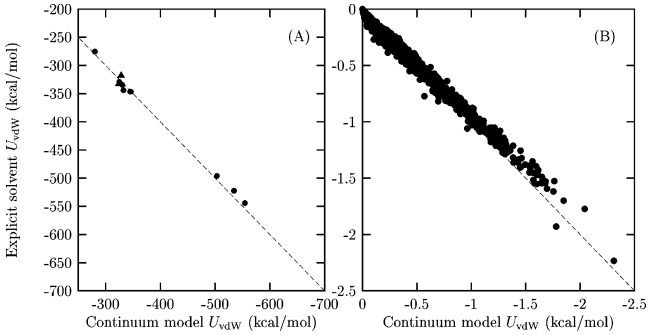

Figure 2. (A) Comparison between explicit and continuum solvent solute−solvent van der Waals energies of a set of native (●) and misfolded (▲) protein conformations. (B) Comparison between the explicit and continuum solvent solute−solvent van der Waals energies for individual atoms of protein 2aza(a). The dashed lines of unit slope indicate perfect correlation.

$r_w = 0.85$ was found to give the best agreement between the continuum and explicit solvent models.

3. Results and Discussion

3.1. Accuracy of the Continuum Model. The explicit and continuum solvent values of the solute−solvent van der Waals energies of the native and misfolded protein conformations are presented in Figure 2A. The explicit solvent simulations and the continuum van der Waals model results are in very good agreement (0.995 correlation coefficient). In addition to the good accuracy achieved, several orders of magnitude of time saving is achieved by using the continuum solvent model instead of the explicit solvent model. The explicit solvent calculations for the proteins took on average one to two weeks of computer time to converge compared to the minutes required by the corresponding continuum solvent calculations.

An important property of the continuum solvent model we developed is that it predicts accurately not only the total solute−solvent van der Waals energy but also the solute−solvent van der Waals energy of individual solute atoms and residues. As an example, Figure 2B compares the continuum solute−solvent van der Waals energy of each individual atom of chain A of the protein Azurin (PDB id 2aza) with the corresponding values obtained from the explicit solvent simulation. The agreement (0.999 correlation coefficient) is quantitative across the wide range of solute−solvent van der Waals energies, from the large ones corresponding to atoms exposed to the solvent to the very small ones corresponding to atoms buried in the protein core. The continuum model is therefore capable of characterizing not only large scale conformational rearrangements, such as when comparing folded and misfolded protein conformations, but also small conformational changes involving the motion of only a few atoms or, for example, when predicting the effects of mutations. In contrast, we observe that the solute−solvent van der Waals interaction energies of individual solute atoms do not correlate well with the accessible surface area of the atoms. Indeed, we estimate that the buried atoms of a protein (defined as having less than 1 Å^2 of solvent accessible surface area) comprise approximately 50% of the atoms of the protein and account for up to 40% of the total van der Waals protein−water interaction energy (this is further considered in section 3.2).

3.2. Native and Extended Protein Conformations, Peptides, and Misfolded Structures. In this section, we present the analysis of solute−solvent van der Waals interaction energies

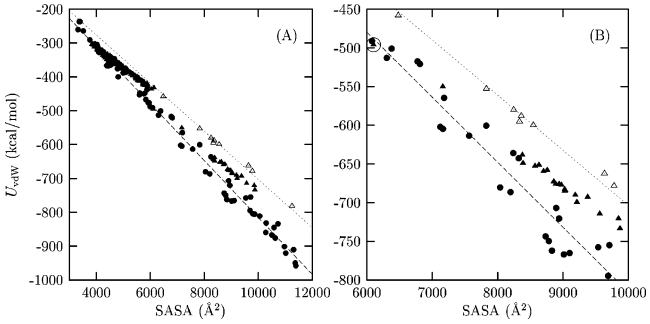

Figure 3. Continuum solvent solute−solvent van der Waals interaction energies of various peptide and protein conformations (see below) plotted against their accessible surface area. (A) Data with accessible surface area between 3000 and 12 000 Å^2. (B) Data with accessible surface area between 6000 and 10 000 Å^2; circles denote native protein conformations, filled triangles denote decoy conformations of protein lz1 (the native conformation of lz1 is circled), and open triangles denote extended protein conformations. The dashed line is the linear least-squares fit to all native protein conformations examined; the dotted line is the linear least-squares fit to all extended protein conformations examined. Native protein structures (PDB id, chain designation in parentheses): 1a9m, 1a9m(a), 1a9m(b), 1ay7, 1ay7(a), 1ay7(b), 1azg, 1azg(a), 1azg(b), 1bhh, 1bhh(a), 1bhh(b), 1bun, 1bun(a), 1bun(b), 1c1y, 1c1y(a), 1c1y(b), 1c5y, 1c5y(a), 1c5y(b), 1d0d, 1d0d(a), 1d0d(b), 1d0q, 1d0q(a), 1d0q(b), 1fpr, 1fpr(a), 1fpr(b), 1fxt, 1fxt(a), 1fxt(b), 1fyn, 1fyn(a), 1fyn(b), 1gdn, 1gdn(a), 1gdn(b), 1i8h, 1i8h(a), 1i8h(b), 1j4l, 1j4l(a), 1j4l(b), 1j4q, 1j4q(a), 1j4q(b), 1k8r, 1k8r(a), 1k8r(b), 1kwa, 1kwa(a), 1kwa(b), 1qix, 1qix(a), 1qix(b), 1qwf, 1qwf(a), 1qwf(b), 1sph, 1sph(a), 1sph(b), 1taw, 1taw(a), 1taw(b), 1zii, 1zii(a), 1zii(b), 2cyh, 2cyh(a), 2cyh(b), 2phk, 2phk(a), 2phk(b), 2pld, 2pld(a), 2pld(b), 1vac(a), 1vac(p), 1vac(a+p), 2clr, 1bkm, 1dwc, 1aq7, 2bpx, 1ctf, 1fas, 1fc2, 1lz1, 1shg, 1tul, 1ubi, 2aza(a), 2cro, 2ovo, 4pti, and 6pti. Extended protein structures of the following: 1ctf, 1fas, 1fc2, 1lz1, 1shg, 1tul, 1ubi, 2aza(a), 2cro, 2ovo, 4pti, and 6pti. A series of 34 low energy conformations of the Ace-GEWTYDDATKTFTVTE-Nme octadecapeptide including the fully extended conformation and the β-hairpin conformation from the C-terminal fragment of 1gb1. Protein decoys of the 1ctf, 2cro, and 1lz1 proteins from the 4-state-reduced[69] and ROSETTA[71] all-atom protein decoy sets.

(U_{vdW}), using the continuum solvent model of a series of peptides and protein conformations. The surface areas of the molecules analyzed span a very large range, from 50 to 20 000 Å^2. The data collected for proteins with surface areas ranging between 3000 Å^2 to 12 000 Å^2 are shown in Figure 3A; Figure 3B shows a subset of the results in an expanded region between 6000 and 10 000 Å^2.

On the very coarse grained scale of Figure 3A, the solute−solvent energies of proteins in their native conformations (filled circles Figure 3A) lie along one line, while the corresponding results for extended conformations of these proteins lie along another (open triangles Figure 3A). Least-squares fitting of the surface area model to the solute−solvent van der Waals energies of the native protein conformations yields

$$\Delta U_{vdW} = \gamma_{vdW} A + b_{vdW} \qquad (10)$$

where A is the SASA, $\gamma_{vdW} = -84$ cal/mol/Å^2, and $b_{vdW} = 25.2$ kcal/mol for the surface tension and intercept of the line fit to the native conformations and $\gamma_{vdW} = -71$ cal/mol/Å^2 and $b_{vdW} = 8.97$ or the surface tension and intercept of the line fit to the extended conformations. The least-squares fits of the native and extended solute−solvent energies correspond to the dashed and dotted lines in Figure 3. Native and extended proteins represent extreme examples of chain compactness; the native states are the most compact, and the extended conformations, the least compact. We have also analyzed the solute−solvent van der Waals interaction energy as a function of SASA

for protein structures with intermediate compactness representing models for protein folding intermediates and misfolded structures. Protein "decoys" used for macromolecular force field validation provide models for these structures.[5] The solute−solvent van der Waals energies for three sets of decoys (1ctf, 2cro, 1lz1[69,71]), two sets (1ctf and 2cro) with surface areas in the range 4000 Å^2 to 6000 Å^2 and the other (1lz1) with surface areas in the range 8000 to 10 000 Å^2, are also shown in Figure 3. The solute−solvent energies for these misfolded structures fall approximately along two lines corresponding to the progressive unfolding of these proteins (the data for 1ctf and 2cro lie along the same line). The effective surface tension obtained fitting the protein decoy data is approximately $\gamma_{vdW} = -60$ cal/mol/Å^2. We have also calculated (data not shown in Figure 3) the continuum solvent estimates of the solute−solvent van der Waals energies of a set of low energy conformations of the Ace-GEWTYDDATKTFTVTE-Nme octadecapeptide. The best fit to these data yields an effective surface tension close to the values obtained for the protein decoys.

It is apparent from these results that the coarse grained linear fits of the protein−solvent attractive dispersion energy to the solvent accessible surface area are not transferable between these different types of data sets. The surface area models parametrized on the peptides and misfolded proteins yield a much smaller surface tension coefficient ($\gamma_{vdW} = -60$ cal/mol/Å^2) than the value extracted by fitting the native conformations ($\gamma_{vdW} = -84$ cal/mol/Å^2). Why do the three different data sets of (1) native proteins, (2) misfolded proteins, and (3) extended proteins have different coarse grained surface tension parameters? The data of Figure 3 consist of the total energy and total surface area for each structure; these quantities are obtained by summing over atomic contributions. As we illustrate below, the differences in the coarse grained surface tension parameters reflect differences in the underlying atomic distributions of U_{vdW} as a function of SASA.

Figure 4 shows distributions of U_{vdW} as a function of SASA for the heavy atoms of protein 2aza (Figure 4 parts A,B are derived from the native state conformation, while Figure 4 parts C,D are derived from the extended state conformation). There are two points to be made. First, on an atomic scale, the correlation between accessible surface area and solute−solvent interactions is weak. About 40% of the total solute−solvent interaction energy comes from atoms of the protein which are completely buried just under the surface. Second, the atomic distributions (U_{vdW} vs SASA) shown in Figure 4 corresponding to the native and extended conformations of the protein are quite different from each other. The atomic distribution U_{vdW} versus SASA, shown for the native state in Figure 4A,B, determines the ratio, $\gamma_{vdW} = U_{vdW}/A$, of the total van der Waals solute−solvent energy to the total exposed surface area. Because the atomic distribution for any native protein is similar to 2aza shown in Figure 4, there is a coarse grained linear relation between total U_{vdW} and total surface area for native proteins. The corresponding atomic distributions for extended conformations of proteins, while qualitatively different from native conformations, are all similar to the distribution shown for the extended conformation of 2aza shown in Figure 4C,D. This underlies the difference between coarse grained surface tension for extended proteins and native proteins.

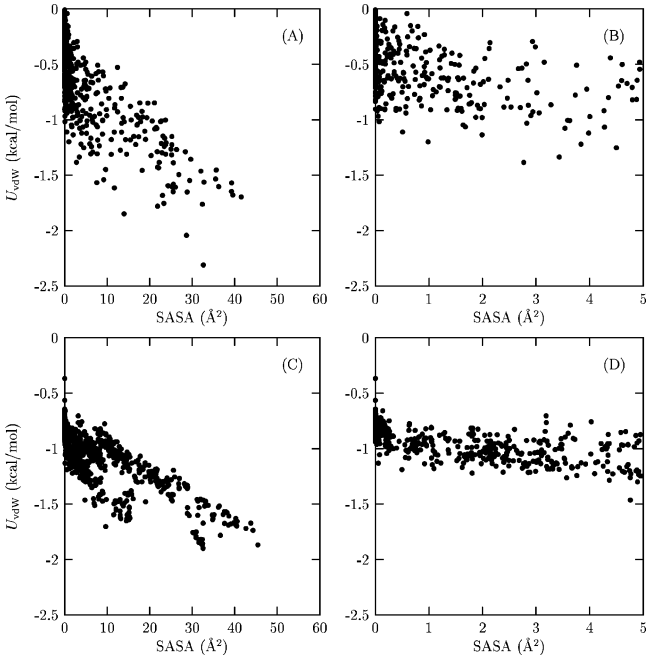

Figure 4. Continuum solvent solute−solvent van der Waals energy of the heavy atoms of protein 2aza(a) vs their accessible surface area. (A) Native conformation. (B) Native conformation, atoms with surface areas less than 5 Å^2. (C) Extended conformation. (D) Extended conformation, atoms with surface areas less than 5 Å^2.

Why is the coarse grained effective surface tension ($\gamma_{vdW} = -84$ cal/mol/Å^2) measured for the native protein conformations larger in magnitude than the corresponding value for extended conformations ($\gamma_{vdW} = -71$ cal/mol/Å^2) and even larger still than the values extracted from the peptide and misfolded protein conformations ($\gamma_{vdW} = -60$ cal/mol/Å^2)? The difference between the native and extended conformations' effective surface tensions can be rationalized in terms of the different protein−solvent van der Waals energies per unit surface area, given by the ratio $\gamma_{vdW} = U_{vdW}/A$, which characterizes the native and extended distributions shown in Figure 4. A comparison of the distributions of atomic solvation energies for the native and extended conformations of protein (compare the mean value along each axis of Figure 4A,B with that of 4C,D) shows that there are two competing effects. The atoms of the native protein which are proximal to the surface but which have no surface area exposed make a substantial contribution (about 40% of the total) to the van der Waals interaction energy between the native protein and the solvent. They increase the effective γ_{vdW} of the native conformation. The extended structures do not contain this large contribution to the solute−solvent energy from buried atoms. This effect more than offsets the somewhat increased U_{vdW} at fixed surface area for exposed atoms in the extended conformation relative to the native.

The effective surface tensions for the native conformations cannot be directly compared to the surface tension coefficients obtained from fitting the surface area model to the protein decoys and octadecapeptide data because they represent two different physical quantities. The first measures the rate of change of the protein−solvent van der Waals energy with respect to solvent accessible surface area for the process of going

(71) Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins: Struct., Funct., Genet.* **1999**, *S3*, 171−176.
(72) http://www.bmm.icnet.uk/docking.
(73) Rizzo, R.; Tirado-Rives, J.; Jorgensen, W. *J. Med. Chem.* **2001**, *44*, 145−154.

from one protein to another protein with a different number of residues without changing the distribution of energies and surface areas which characterizes the native state (see Figure 4A); as a result the native effective surface tension is equal to the native protein−solvent van der Waals energy per unit surface area U_{vdW}/A. In contrast, the effective surface tension fitted to the protein decoy conformations measures the rate of change of the protein−solvent van der Waals energy with respect to solvent accessible surface area as the protein unfolds; a process that corresponds to continuously changing the native solute−solvent interaction energy−surface area distribution into the distribution corresponding to the extended conformations distribution keeping the number of atoms constant. The two processes are not equivalent; indeed, the values of U_{vdW}/A for the decoy conformations, ranging between the native (−84 cal/mol/Å^2) and the extended (−71 cal/mol/Å^2) effective surface tensions, are much larger in magnitude than the slope of the lines connecting the protein decoys ($\simeq$ −60 cal/mol/Å^2). The value of the effective surface tension for the protein decoys is smaller in magnitude than the native surface tension because, on a scale set by the native γ_{vdW}, the surface area per atom increases more rapidly on average than the interaction energy as the atoms of the protein change their environment from native to extended. A similar effect was observed previously when studying the solvation of small molecules.[42]

While the atomic distributions shown in Figure 4 help to rationalize the different coarse grained nonpolar surface tensions for different kinds of protein structures shown in Figure 3 (native, misfolded, extended), at high resolution, there is a large variation of total solute−solvent interaction energies for individual proteins from values predicted by the coarse grained surface tension. This can best be seen on the expanded scale shown in Figure 3B. Note for example the spread of native U_{vdW} values for the cluster of proteins with surface areas around 9000 Å^2. In this region, deviations of as much as 50 kcal/mol from the surface area model are observed and the energy differences between proteins with similar surface area can be 1 order of magnitude larger than that predicted by the surface area model. Conversely, examples can be found of proteins with similar solute−solvent van der Waals energies but differing by as much as 1000 Å^2 in surface area. The effect of these deviations from the coarse grained predictions is greatly magnified when modeling binding as discussed in the following section.

3.3. Protein−Protein and Protein−Ligand Binding Energies. An important application of solvation free energy models is to compute the solvation contribution to the binding free energy.[13,47] We have tested the ability of surface area models to predict the solute−solvent van der Waals energy component of the free energy of binding for a series of protein and ligand−protein complexes computed using the continuum solvent model described in section 2.3.

The solute−solvent van der Waals binding energies are calculated as the difference between the solute−solvent energy of the complex and the sum of the solute−solvent energies of the isolated monomers. The protein−protein complexes analyzed, listed in the caption of Figure 5, include native and nonnative binding decoy conformations. We have also analyzed the solute−solvent van der Waals binding energies of the ligand−protein complexes listed in the caption of Figure 5. Figure 5 shows the solute−solvent van der Waals binding energies, ΔU_{vdw}, as a function of the change of SASA upon binding, ΔA, in the range of ΔA between −1500 and −500 Å^2. The solute−solvent van der Waals binding energies are positive reflecting the fact that solute−solvent energy is lost upon binding (desolvation effect). They are also of the same order of magnitude or larger than typical binding free energies, underlining the fact that nonelectrostatic desolvation effects are important in binding affinity prediction.

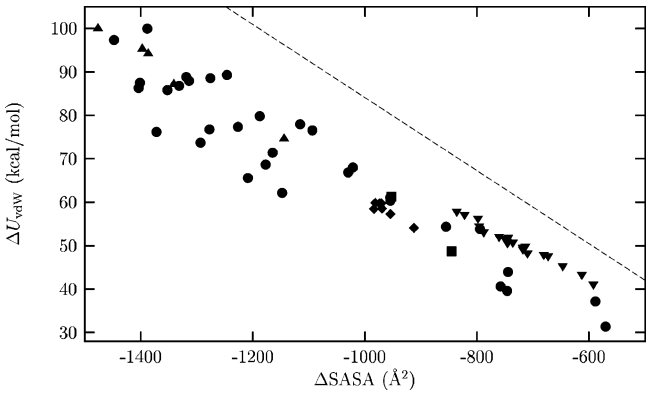

Figure 5. Solute−solvent van der Waals energy of binding, ΔU_{vdW}, plotted vs change of solvent accessible surface area for various protein−protein and ligand−protein complexes. (●) protein−protein complexes, (▲) HIV protease complexes, (▼) HIV reverse transcriptase complexes, (◆) sulfonamide−thrombin complexes, (■) other complexes. The dashed line indicates the prediction from the surface area model fit to the native protein conformations (see Figure 3). The protein−protein complexes analyzed are those proteins composed of two chains listed in the caption of Figure 3 and the native and two decoy conformations[72] of each of the complexes 1brc, 1bgs, 1avz, 1cgi, 2kai, and 1ugh. The ligand−protein complexes analyzed are 1bkm, 2clr, 6 argotroban sulfonamide analogues complexed with thrombin (1dwc), 1aq7, 20 HEPT analogues complexed with HIV reverse transcriptase (1rt1) from reference 73, 2bpx, 1hpv, 1htg, and 1hvj.

The scatter of the binding energies with respect to the surface area loss is, in a relative sense, larger than the scatter of the absolute solute−solvent van der Waals energies of the monomers around the corresponding best surface area model (see Figure 3). Because the values of $\gamma_{vdW} = \Delta U_{vdW}/\Delta A$ obtained from the solute−solvent binding energy data range from −50 to −79 cal/mol/Å^2, a surface area model with a single effective surface tension parameter does not accurately reproduce all of the binding energies. Subsets of the protein−ligand binding energies data (see for example the binding energies for the HIV reverse transcriptase complexes, downward pointing triangles in Figure 5) are found to be linearly correlated with the surface area loss; however, the corresponding proportionality constants obtained by a least-squares fit are often significantly system dependent. For instance, the best γ_{vdW} for the thrombin and sulfonamide complexes alone is −78 cal/mol/Å^2 but for the HIV reverse transcriptase and HEPT analogue complexes is −65 cal/mol/Å^2. The larger scatter with respect to the surface area model observed for the binding energies is due to the fact that the binding energies depend on the desolvation of only those relatively few atoms involved in binding, rather than reflecting the solute−solvent interaction energy averaged over the entire protein. We observed in the previous section that although an effective surface tension γ_{vdW} can be used to describe the approximately linear relation between solute−solvent van der Waals energy and surface area on a coarse grained scale, in general it does not reproduce well the solute−solvent interaction energies of individual atoms. The scatter of the binding energies

with respect to surface area also reflects the wide variety of binding geometries observed. For example, although the contact area between the monomers in a protein−ligand and protein−protein complex are sometimes similar, a small ligand, given its size, shields the binding partner from the solvent to a smaller degree.

The data presented in Figure 5 also show a systematic discrepancy between the calculated binding energies and the surface area model for the native proteins obtained in the previous section. The solute−solvent van der Waals binding energy of a complex formed by monomers *a* and *b* can be written as

$$\Delta U_{\mathrm{vdW}} = [U'_{\mathrm{vdW}}(a) - U_{\mathrm{vdW}}(a)] + [U'_{\mathrm{vdW}}(b) - U_{\mathrm{vdW}}(b)] \quad (11)$$

where U_{vdW} is the solute−solvent interaction energy of the isolated monomer and U'_{vdW} is the solute−solvent interaction energy of the monomer in the complex. When the surface area model for the native protein conformations (eq 10 with $\gamma_{\mathrm{vdW}} = -84$ cal/mol/Å^2) is applied to calculate each term in eq 11, the result, shown as a dashed line in Figure 5, systematically overestimates the calculated binding energies. The value of the effective surface tension for the native protein conformations (−84 cal/mol/Å^2) is inconsistent with the slope (−62 cal/mol/Å^2) of the line that best approximates in a coarse grained sense the continuum solvent solute−solvent van der Waals binding energies of the protein−protein complexes shown in Figure 5. Analysis of the individual desolvation energies $U'_{\mathrm{vdW}} - U_{\mathrm{vdW}}$ of each monomer reveals that atoms buried upon binding still interact to some degree with the solvent (see Figure 4B); therefore, the surface area model, which instead assumes that those atoms no longer interact with the solvent, overestimates the desolvation penalty. We have also observed that, because the interface region of a protein−protein complex is in general more corrugated and interacts less strongly with the solvent than the protein surface not involved in binding, often a smaller than expected desolvation penalty is obtained by burying the interface surface in the interior of the protein complex.

4. Conclusions

The hydration free energy is decomposable into electrostatic and nonpolar components. Implicit solvent models often estimate the nonpolar component by means of a term proportional to the solvent exposed surface area (surface area model). The nonpolar component is further decomposable into a cavity formation component and a solute−solvent van der Waals dispersion attraction component. The solute−solvent dispersion attraction can be accurately approximated by the average solute−solvent van der Waals interaction energy. Implicit solvent models must account for this term in order to accurately predict solvation free energies.

In this work, we show that a continuum solvent van der Waals interaction energy model, based on the integral over the solvent volume of the attractive component of the solute−solvent interaction potential of each solute atom, yields solute−solvent van der Waals energies in good agreement with explicit solvent calculations. We also show that linear surface area models are unable to predict as accurately the average solute−solvent van der Waals interaction energies of ligands and various conformations of peptides and proteins obtained from explicit solvent simulations and from the continuum solvent van der Waals interaction energy model.

The major drawback we have observed with surface area models is their poor transferability. A single surface area model does not reproduce the solute−solvent van der Waals energies for any of the systems examined. In fact, for proteins we found that a particular parametrization of the surface area model is only applicable to the subset of protein conformations included in the parametrization. For instance, the surface area model that approximately reproduces the solute−solvent van der Waals energies of native protein conformations does not reproduce the energies of extended protein conformations. Moreover, this surface area model is found inappropriate for describing the relative solute−solvent energies of a set of misfolded conformations (protein decoys). We also found substantial scatter of the solute−solvent van der Waals energies around the corresponding effective surface tensions that best approximates these datasets. We found that surface area models are even less accurate in reproducing the van der Waals desolvation energy component of the free energy of binding of protein−protein and protein−ligand complexes. The surface area model parametrized for the native protein conformations complexes is found to systematically overestimate the binding desolvation energies, and the scatter around the effective surface tension specifically parametrized to best reproduce the binding desolvation penalty is found to be large in relation to the magnitudes of the desolvation energies. Surface area models are also found to be grossly inaccurate in reproducing the solute−solvent van der Waals energies of individual solute atoms. In proteins, we find many buried atoms (zero solvent accessible surface area) that contribute significantly to the total solute−solvent van der Waals interaction energy.

Acknowledgment. This work was supported in part by a grant from the National Institute of Health (GM30580).

JA029833A