Continuum Solvent Modeling of Nonpolar Solvation: Improvement by Separating Surface Area Dependent Cavity and Dispersion Contributions

Martin Zacharias*

Theoretical Biophysics, Institute of Molecular Biotechnology, Beutenbergstr. 11, D-07745 Jena, Germany

Received: December 2, 2002

Continuum models are frequently used to calculate hydration properties of organic and biomolecules and to estimate its dependence on conformation and association. Often a single surface tension parameter has been used to estimate nonpolar solvation contributions from the solvent accessible surface of a molecule. The assumption of a uniform surface tension parameter is based on the observation that for linear alkanes free energies of hydration (vacuum to water transfer free energies) increase (approximately) linearly with the solvent accessible surface area (SAS). However, this correlation for example does not hold for the vacuum to water transfer of cyclic alkanes. The transfer of a nonpolar solute from vacuum to water can be formally split into a contribution due to cavity formation which involves a redistribution and reordering of water molecules and changes in water−water interaction and second van der Waals (dispersion) interactions between solute and water. In the present study, the solute solvent dispersion contribution has been calculated using a surface integral continuum approach (Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616−627). Combined with a cavity contribution that has been assumed to be proportional to the solvent accessible surface area calculated hydration free energies for linear, branched and cyclic alkanes are in significantly better agreement with experiment than using a pure SAS model. In addition, the calculated changes of hydration free energies upon alkane conformational changes agree much better with results of explicit solvent simulations compared to a model that employs a single surface tension parameter.

Introduction

Continuum solvent models have been used extensively to study many biological processes such as folding, association, and biomolecule partioning between aqueous and organic phases.[1−12] Because of computational limitations, the calculation of solvation contributions for large organic or biomolecules using explicit solvent simulations is difficult because of the slow convergence of thermodynamic perturbation and integration methods. Many applications that involve large biomolecular structures, generation of many different conformers, or drug design applications which require the evaluation of many putative biomolecule-ligand complexes demand the application of reasonably accurate continuum solvent models to account for solvation.

Based on macroscopic solvation concepts, one of the most common continuum modeling approaches involves splitting the aqueous solvation of a molecule into a nonpolar and polar contribution (for example, see refs 4−8). For calculating the polar solvation contribution, the solute is treated as a low dielectric cavity embedded in a high dielectric (aqueous) continuum. Assuming a charge distribution given by a molecular mechanics force field, the polar solvation contribution can be calculated from a solution of the Poisson equation (or Poisson−Boltzmann equation in the case of salt containing solutions). The nonpolar solvation of a molecule is estimated from the solvent accessible surface area (SAS) of the molecule using a uniform surface tension coefficient. Usually, the surface tension coefficient is obtained from a fit of experimental *n*-alkane vacuum−water transfer (hydration) free energies vs surface area. This yields an estimate for the surface area tension coefficient (termed γ) around $\sim 0.005-0.007$ kcal mol^{-1} Å^{-2}.[4,6−9] Interestingly, with an appropriate parametrization (internal dielectric constant for the Poisson calculations, surface probe diameter, etc.), very reasonable overall correlation of calculated vs experimental solvation free energies can been obtained.[4,7] This includes, however, both nonpolar and polar solutes, and for many small polar organic molecules, the polar solvation contribution dominates. For example, the simple SAS dependent nonpolar solvation model largely underestimates the solvation of cyclic alkanes. In addition, the free energy of hydration depends on the conformation of the molecule. Explicit solvent simulations indicate that the standard SAS hydration model with $\gamma \sim 0.005-0.007$ kcal mol^{-1} Å^{-2} significantly underestimates the change of solvation upon conformational changes of nonpolar molecules.[13,14] However, to be useful to study association events and conformational changes of organic and biomolecules, an accurate continuum calculation especially of conformation dependent contributions to solvation is desirable.

The present study indicates that it is possible to yield a much better correlation between calculated nonpolar solvation and experimental results (including cyclic alkanes) if cavity contribution and dispersion contributions are calculated separately. Only the former is assumed to be directly proportional to the solvent accessible surface area, whereas the latter term is calculated by calculating the dispersion interaction of each solute atom with the surrounding solvent using a surface integral approach by Floris and Tomasi.[15] In addition, predictions for

* To whom correspondence should be addressed. International University Bremen, Dept. of Comput. Biology, Campus Ring 1, D-28759 Bremen, Germany. E-mail: m.zacharias@iu-bremen.de.

10.1021/jp027598c CCC: $25.00
Published on Web 04/01/2003

the conformation dependence of hydration free energies of some alkanes are in much better agreement with explicit solvent simulations compared to the pure SAS model.

Methods

Because the focus of the present study is on calculating nonpolar contributions to solvation, the alkane molecules were treated as neutral without any partial charges on the atoms. For simplicity and to avoid any electrostatic contributions to solvation, a united atom model was used with a carbon−(water)-oxygen dispersion parameter, C_{sw}, taken from the OPLS force field (ref 16, $C_{sw} = 1246.0$ kcal mol^{-1} Å^6). All molecules were generated using the Builder module of the InsightII software package (Accelrys. Inc., Burlington, MA). An extended (trans) conformation was used for the linear alkanes. It is assumed that the solvation free energy of alkane molecules can be separated into a cavity contribution that includes entropic and enthalpic contributions due to reordering and reorientation of water molecules around the solute and second solute−solvent van der Waals interactions. The calculation of the cavity contribution was based on the SAS calculated using the Shrake and Rupley method[17] with a water probe radius of 1.4 Å.

In contrast to solvation models that assume for both contributions a linear surface area dependence, in the present study, the solute−solvent van der Waals interaction was calculated using a continuum method by Floris and Tomasi.[15] In this continuum model, the water distribution around solute atoms (i) is approximated by a continuous density function, $\rho_{iw}(r_{iw})$, depending only on the distance, r_{iw}, between water and solute atoms and a solute atom-water correlation function, $g_{iw}(r_{iw})$

$$\rho_{iw}(r_{iw}) = \rho_w\, g(r_{iw})$$

Assuming that the dispersion interaction between solute and water falls off with the sixth power of the distance (possible contributions that fall off with a different power in the distance have been neglected) and is characterized by a single carbon−water−oxygen dispersion parameter, C_{iw}, the dispersion interaction can be calculated from

$$U_{disp}(\text{solute in water}) = \sum_i^{N_s} \rho_w C_{iw} \int_{V_w} dr_{iw}\, g(r_{iw}) r_{iw}^{-6}$$

The sum is here taken over all solute atoms (N_s) of the molecule and the integration is over the solute excluded volume (V_w). As shown by Floris and Tomasi[15] based on Huron and Claverie,[18] this volume integral can be transformed into an integral over the surface of the solute excluded volume (vectors are in bold)

$$U_{disp}(\text{solute in water}) = \sum_i^{N_s} \rho_w C_{iw} \int_{S_s} d\sigma_S\, \mathbf{A}(\mathbf{r}_{iw})\cdot\mathbf{n}_S$$

with

$$\mathbf{A}(\mathbf{r}_{iw}) = [(1/r_{iw}^3)\int_{r_{iw}}^{\infty} g_{iw}(x)/x^4\, dx]\mathbf{r}_{iw}$$

The solute excluded volume and its surface are determined by appropriate van der Vaals radii of the solute atoms and a solvent probe radius (see below). For the case of $g(\mathbf{r}) = 1$ (uniform solvent density outside solute cavity), the above one-dimensional integral can be solved

TABLE 1: Comparison of Alkane−Water Interactions from Explicit Solvent Simulations and Continuum Model Calculations[a]

alkane molecule	U_{disp} MD (explicit water)	U_{disp} (SASI)
methane	−3.3	−3.2
ethane	−5.4	−5.4
propane	−7.2	−7.3
butane	−9.0	−9.0
pentane	−10.8	−10.7
hexane	−12.4	−12.4
isobutane	−8.9	−8.8
2-methylbutane	−10.1	−10.3
neopentane	−10.4	−10.1
cyclopentane	−10.0	−10.6
cyclohexane	−11.7	−12.1

[a] Column 2 gives the solute solvent interactions (in kcal mol^{-1}) calculated by Gallicchio et al.[14] for a set of alkane molecules (column 1) using thermodynamic perturbation simulations including explicit solvent (termed ΔU_{vw} in Table 2 of ref 14). Values in column 3 are obtained from the surface integral continuum method by Floris and Tomasi[15] using $r_{probe} = 1.8$ Å for defining the solute solvent boundary.

$$\mathbf{A}(\mathbf{r}_{iw}) = \mathbf{r}_{iw}/(3r_{iw}^6)$$

such that the dispersion interaction can be estimated from a discrete summation over surface area elements, ΔS_k, around the molecule:[15]

$$U_{disp}(\text{solute in water}) = \sum_i^{N_s} \rho_w C_{iw} \sum_k \Delta S_k \mathbf{A}(\mathbf{r}_{ik})\cdot\mathbf{n}_{S_k}$$

Here, $\mathbf{r}_{ik}$ points from atom i to the surface element k, and $\mathbf{n}_{S_k}$ is the normal vector associated with surface element ΔS_k. $\mathbf{A}(\mathbf{r}_{ik})$ is evaluated as defined above. The surface area elements (ΔS_k) were calculated using the Shrake and Rupley method[17] which generates a set of surface points representing uniformly sized surface elements around the molecule. The normal vector, $\mathbf{n}_{S_k}$, associated with each surface element was calculated as the normalized vector pointing from the atom center to the associated surface point. For the calculation of the dispersion contribution, a second surface was defined using the Shrake and Rupley method[17] and a slightly larger probe radius of 1.8 Å compared to a probe radius of 1.4 Å used to calculate the solvent accessible surface area (see above). This probe radius together with a united atom carbon van der Waals radius of 1.9 Å and the OPLS parameter for the water−solute dispersion interaction gave good agreement with results of explicit solvent simulations by Gallicchio et al. (ref 14, see below). At a minimum distance of ∼3.7 Å of each surface element from any carbon atom, the repulsive part of the LJ potential is much smaller than the attractive dispersion contribution and therefore only the attractive dispersion interactions have been included.

Results and Discussion

The solute−solvent dispersion interactions calculated with the continuum model[15] were compared to solute−solvent interaction energies for a set of alkanes studied by Gallicchio et al.[14] using explicit solvent simulations (Table 1). It should be noted that Gallicchio et al. used an all-atom (OPLS) model in the explicit solvent simulations, whereas in the present study, for simplicity, a united atom model with a single water−carbon dispersion coefficient was used. Therefore, perfect agreement is not expected; nevertheless, with an appropriate choice for $r_{probe} = 1.8$ Å (probe radius to define the surface for calculating average solute solvent dispersion interactions) U_{disp} from the continuum model correlates quite well with the results from

explicit solvent simulations.[15] It is interesting to compare for example butane and cyclopentane (or pentane and cyclohexane). Both pairs of molecules have approximately the same surface area, yet, their experimental hydration free energies differ by ~1 kcal mol^{-1} in favor of cyclopentane (or cyclohexane). This clearly indicates that a model which simply relates hydration free energies to the amount of solvent accessible surface area is insufficient for these pairs of molecules. Both the explicit solvent studies by Gallicchio et al.[14] as well as the continuum solute−solvent dispersion model indicate that the van der Waals interaction between cyclopentane and water is ~1−1.5 kcal mol^{-1} more favorable then the butane−water interaction (similar for the cyclohexane/pentane pair). A simple physical interpretation for this result is that the solvent accessible surface or envelope of cyclopentane/cyclohexane encompasses one more atom (or more electrons that cause dispersion interactions) than butane/pentane, respectively, which in turn enhances the dispersion interaction with the solvent. This is also in line with explicit solvent simulation studies on nonpolar solvation that indicate that dispersion interactions are not necessarily simply proportional to the accessible surface area of a molecule.[13−14,19−20]

Based on the observation that for linear alkanes the hydration free energy increases (approximately) linearly with the number of carbon atoms and the alkane surface area nonpolar hydration free energies of organic or biomolecules are frequently calculated using[4−9]

$$G_{\mathrm{hydr}}(\mathrm{SAS}) = \beta + \gamma\ \mathrm{SAS} \qquad (1)$$

with $\beta = 0.92$ kcal mol^{-1} and $\gamma = 0.0055$ kcal mol^{-1}Å^{-2},[7] the correlation between experimental and calculated hydration free energies is quite reasonable for the linear and branched alkanes (correlation coefficient: $r = 0.98$) but poor for cyclic alkanes (Figure 1a, inclusion of cyclic alkanes gives a correlation coefficient: $r = 0.33$). The possibility to calculate free energies of hydration as a sum of cavity contributions (SAS part) and solute solvent dispersion interactions obtained from the surface integral (SI) continuum approach[15] was exploited by using

$$G_{\mathrm{hydr}}(\mathrm{SASI}) = a + b\ \mathrm{SAS} + c\ U_{\mathrm{disp}}(\mathrm{SI}) \qquad (2)$$

The parameters *a*, *b*, and *c* have been obtained from a fit of calculated vs experimental hydration free energies (SASI indicates surface area cavity + surface integral dispersion contributions to solvation). All linear, branched, and cyclic alkanes given in the legend of Figure 1 have been included for the fit. Similar to the pure SAS model (eq 1), parameter *a* is required because in the model a solute with zero radius still has a solvent accessible surface and a solute solvent dispersion interaction. Parameters *b* and *c* relate the surface area dependent cavity contribution and the solvent−solute dispersion interaction, respectively, to the calculated hydration free energy of linear, branched, and cyclic alkanes. Optimal correlation between calculated and experimental hydration free energies was obtained for $a = -2.0$ kcal mol^{-1}, $b = 0.0435$ kcal mol^{-1} Å^{-2} and $c = 0.62$. With this set of parameters, the correlation coefficient for comparing experimental and calculated solvation free energies of all alkanes given in the legend of Figure 1 is $r = 0.83$ (compared to 0.33 in case of the SAS model, see above). If one includes only linear and branched alkanes, the correlation coefficient for the SASI model is similar to the value obtained with the SAS model ($r = 0.98$). This indicates that the SASI model performs equally well as the SAS model for predicting linear and branched alkane solvation and is a significant improvement in the case of including cyclic alkanes. Parameter

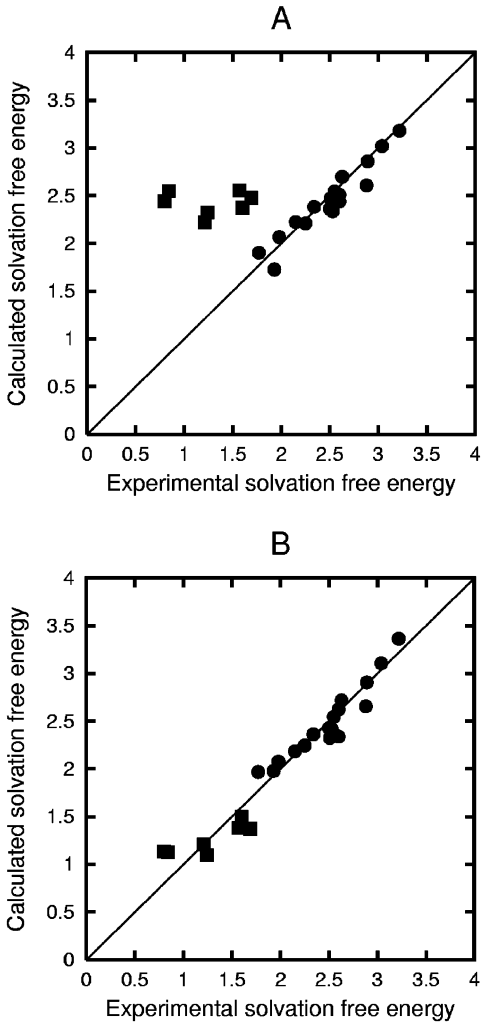

Figure 1. Calculated hydration free energy for cyclic (■), linear, and branched (●) alkanes vs experimental hydration free energies for the SAS continuum solvent model (A) and the SASI model (B). The list of 24 alkane molecules includes cyclopentane, cyclohexane, cycloheptane, methyl-cyclopentane, methylcyclohexane, dimethylcyclohexane, cylooctane, 2-methylpropane, 2-methylbutane, 2,2-dimethylpropane, 2-methylpentane, 3-methylpentane, 2,2-dimethylbutane, 2,4-dimethylpentane, 2,2,5-trimethylhexane, and all linear *n*-alkanes starting from methane through *n*-octane. Experimental hydration free energies are from Ben-Naim and Marcus.[21] All values are in kcal mol^{-1}.

b is necessary to scale the calculated solute−solvent dispersion interactions to obtain optimal agreement of calculated and experimental hydration free energies (using an $r_{\mathrm{probe}} = 1.8$ Å). An alternative way to adjust the calculated solute solvent dispersion interaction would be to appropriately adjust the carbon−(water)oxygen dispersion parameter or the probe radius used to define the solute−solvent boundary. The correlation between calculated and experimental hydration free energies is much better in particular for the cyclic alkanes compared to the pure SAS model (Figure 1b). The standard deviation for the fit is <0.16 kcal mol^{-1}. The three parameters in eq 2 were further validated by systematically omitting single molecules from the set used for the fit to obtain *a*, *b*, and *c* yielding standard deviations of 0.06 kcal mol^{-1} for parameter *a*, 0.0008 kcal mol^{-1} Å^{-2} for *b*, and 0.04 for parameter *c*. This means that omitting any data point from the fit to obtain the parameters in eq 2 leads to parameter variation of ~3%, ~2%, and ~6% for parameters *a*, *b*, and *c*, respectively. For each case, the parameters were used to predict the solvation free energy of the molecule that has been omitted during the fit. This gave a

TABLE 2: Conformation Dependence of Calculated Hydration Free Energy Changes[a]

hexane rotamers	G_{hydTP}	ΔG_{hydTP}	$G_{hydSASI}$	$\Delta G_{hydSASI}$	G_{hydSAS}	ΔG_{hydSAS}
ttt	3.78	0.00	2.57	0.00	2.54	0.00
tgt	3.39	−0.39	2.50	−0.07	2.51	−0.03
tgg	2.60	−1.18	2.12	−0.45	2.45	−0.09
ggg	1.11	−2.67	1.25	−1.32	2.34	−0.2
butane						
trans	2.15	0.0	2.23	0.0	2.23	0.0
cis	1.8	−0.35	1.96	−0.27	2.19	−0.04

[a] Column 1 indicates the hexane or butane conformation (in case of hexane, a t means trans conformation around the selected C−C bond). Note that the ggg hexane conformation (all three C−C bond torsions in a gauche state) is identical to the cyclohexane conformation. G_{hydTP} corresponds to the calculated hydration free energy using free energy perturbation in explicit water (data for hexane is from Table 5 in ref 14; for cis and trans butane from Ashbaugh et al., ref 13). $G_{hydSASI}$ and G_{hydSAS} are the hydration free energies calculated using the SASI or pure SAS method, respectively. Columns 3, 5, and 7 give relative values with respect to the ttt hexane or trans butane conformations, respectively. All values are in kcal mol^{-1}.

standard deviation of 0.18 kcal mol^{-1} and a maximal deviation of the calculated from the experimental solvation free energies of 0.4 kcal mol^{-1}. This deviation is much smaller than the deviation of the calculated vs experimental solvation free energy of cyclic alkanes when using the SAS model which can be larger than 2 kcal mol^{-1} (see Figure 1a).

It is interesting to note that for aromatic nonpolar compounds the calculated hydration free energies are also closer to experiment than for the SAS model even without adjusting the carbon−oxygen dispersion parameter or accounting for any electrostatic contributions to solvation. For example, the calculated hydration free energy for benzene is 0.45 kcal mol^{-1} using the SASI approach compared to the experimental value of −0.883 kcal mol^{-1} by Ben-Naim and Marcus[21] and ∼2.2 kcal mol^{-1} using the SAS model. Accounting for the small polar character of the C−H bond in benzene and/or use of a more appropriate solute−solvent dispersion parameter (increased dispersion between water and aromatic vs aliphatic carbons) may account for the residual discrepancy between calculated and experimental numbers of ∼1.3 kcal mol^{-1}.

It is interesting to note that the surface tension parameter in eq 2 is much larger than the corresponding parameter in eq 1. It is more in the regime of tension parameters obtained for organic liquid (oil)−water interfaces or in the regime of surface tension parameters that have been used to describe hydrophobic contributions to macromolecular association or folding.[8,22−23]

This larger surface tension parameter in eq 2 is also the reason for a much larger predicted conformation dependence of nonpolar hydration if one compares the continuum model represented by eq 2 vs 1. The dependence of the calculated free energies of hydration on alkane conformation has been investigated by Ashbaugh et al.[13] for butane in the trans vs cis conformation and by Gallicchio et al.[14] for hexane rotamers using thermodynamic perturbation (TP) methods in explicit solvent simulations. Both the pure SAS model as well as the SASI model represented by eq 2 were applied to calculate the hydration free energies of butane and hexane rotamers (Table 2). Although the calculated changes in hydration free energy upon alkane conformational change are smaller than the results of explicit solvent simulations, the agreement is much better than using the pure SAS model (eq 1). The predicted conformation dependence of hydration free energies of the SAS model for butane or hexane is only 7−10% of the results obtained by explicit solvent simulations compared to between 50 and 70% in case of the SASI model.

Note that, in case of the SASI model, the parameters that give the best agreement with experiment were used (not parameters that give best agreement with explicit solvent simulations) and that the explicit solvent TP simulations are not necessarily in perfect agreement with experiment.[14] For example, one of the hexane rotamers (the ggg hexane state) has the same conformation as cyclohexane, and the SASI model predicts a hydration free energy relative to ttt hexane that is very close to the experiment (∼1.3 kcal mol^{-1}), whereas the TP calculations overestimate the hydration free energy difference >1 kcal mol^{-1}.

Conclusions

A computationally relatively inexpensive continuum model for nonpolar solvation has been presented that is based on a separate calculation of cavity contribution and dispersion contribution to hydration of nonpolar solutes in water. Although both contributions are calculated from the accessible surface area of the molecule, the dispersion contribution in the current model is not linearly related or simply proportional to the accessible surface area but depends on distances and number of solute atoms that can interact with the accessible surface.

The present continuum treatment of nonpolar solvation shows a considerably improved correlation between the calculated and the experimental vacuum to water hydration free energies compared to a model with a single surface area tension parameter that accounts simultaneously for cavity and dispersion contributions to hydration. In addition, the predicted conformation dependence of hydration free energies is in much better agreement with explicit solvent simulation studies than a single surface tension parameter approach. This result may have significance for molecular mechanics simulation studies that employ an implicit solvent description. Although the model requires about twice the CPU time to calculate the solute−solvent dispersion contribution than to only calculate the surface area, it can still easily be applied to large macromolecules and macromolecular complexes.

Acknowledgment. This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG).

References and Notes

(1) Tanford, C. *The hydrophobic effect: Formation of micelles and biological membranes*, 2nd ed.; Wiley: New York, 1980.
(2) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
(3) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Soc.* **1987**, *84*, 3086.
(4) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrikson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(5) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.
(6) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101.
(7) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(8) Sitkoff, D.; Ben-Tal, N.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 2744.
(9) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465.
(10) Dill, K. A. *J. Biol. Chem.* **1999**, *272*, 701.
(11) Brem, R.; Chan, H. S.; Dill, K. A. *J. Phys. Chem. B* **2000**, *104*, 7471.
(12) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521.
(13) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *Biophys. J.* **1998**, *75*, 755.
(14) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271.
(15) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616.

(16) Jorgensen, W. J.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(17) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351.
(18) Huron, M.-J.; Claverie, P. *J. Phys. Chem.* **1972**, *76*, 2123.
(19) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *J. Am. Chem. Soc.* **1999**, *121*, 9243.
(20) Ashbaugh, H. S.; Garde, S.; Hummer, G.; Kaler, E. W.; Paulaitis, M. E. J. *Biophys. J.* **1999**, *77*, 645.
(21) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(22) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 447.
(23) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.