JCTC Journal of Chemical Theory and Computation

Nonpolar Solvation Free Energies of Protein−Ligand Complexes

Samuel Genheden,[†] Jacob Kongsted,[‡] Pär Söderhjelm,[§] and Ulf Ryde*,[†]

Department of Theoretical Chemistry, Lund University, Chemical Centre, P.O. Box 124, SE-221 00 Lund, Sweden; Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark; and Department of Chemistry and Applied Biosciences—Computational Science, ETH Zürich, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

Received May 24, 2010

Abstract: Recent investigations have indicated that different solvation methods give qualitatively different results for the nonpolar solvation contribution to ligand-binding affinities. Therefore, we have calculated the nonpolar solvation contribution to the free energy of benzene binding to the T4 lysozyme Leu99Ala mutant using thermodynamic integration (TI) and three approximate methods. The total binding free energy was calculated with TI and then decomposed into contributions from the solvent and the solute. The nonpolar contribution from the solute was compared to approximate methods within the framework of the molecular-mechanics and generalized Born with surface area method (MM/GBSA). First, the nonpolar solvation energy was calculated with a linear relation to the solvent-accessible surface area (SASA). Second, a recent approach that divides the nonpolar solvation energy into cavity and dispersion parts was used, and third, the nonpolar solvation energy was calculated with the polarized continuum model (PCM). Surprisingly, the simple SASA estimate reproduces the TI results best. However, the reason for this is that all continuum methods assume that the benzene cavity is filled with water for the free protein, contrary to both experimental and simulation results. We present a method to avoid this assumption and then, PCM provides results that are closest to the results obtained with TI.

Introduction

Solvation plays a crucial role when a small ligand binds to a protein. Therefore, accurate methods to estimate the solvation free energy, ΔG_{solv}, of protein−ligand complexes are of great importance in drug design.[1,2] The free energy of solvation can in principle be calculated by thermodynamic integration (TI) or free energy perturbation (FEP), but these methods are computationally very expensive and can hardly be used to calculate the solvation energy of a full protein. Therefore, a large number of less computationally demanding and more approximate methods have been developed, based on a dielectric continuum treatment of the solvent, and they have been successfully used in many applications.[3−6]

Continuum solvation methods normally split ΔG_{solv} into a polar and a nonpolar contribution.[7,8] The polar solvation free energy, $\Delta G_{polsolv}$, can be accurately estimated by various methods, e.g., the polarized continuum model (PCM), the generalized Born methods, or by solving the Poisson−Boltzmann equation.[3] The accuracy of such calculations in the context of protein−ligand complexes has been extensively studied elsewhere.[9,10] In this work, we are mainly interested in the nonpolar contribution to the solvation free energy, $\Delta G_{nonpolsolv}$.

* To whom correspondence should be addressed. Tel: +46 - 46 2224502. Fax: +46 - 46 2228648. E-mail: Ulf.Ryde@teokem.lu.se.

[†] Department of Theoretical Chemistry, Lund University, Chemical Centre.

[‡] Department of Physics and Chemistry, University of Southern Denmark.

[§] Department of Chemistry and Applied Biosciences—Computational Science, ETH Zürich.

10.1021/ct100272s
Published on Web 09/28/2010

A commonly used method to estimate $\Delta G_{\text{nonpolsolv}}$ is to relate it to the solvent-accessible surface area (SASA) through a linear relation:

$$\Delta G^{\text{SASA}}_{\text{nonpolsolv}} = \gamma_{\text{SASA}}\text{SASA} + b_{\text{SASA}} \quad (1)$$

where γ_{SASA} and b_{SASA} are taken from a linear regression of the solvation free energy of a set of small apolar molecules in water.[11,12] This is the method used in the MM/GBSA (molecular mechanics with generalized Born and surface-area) approach to calculate ligand-binding free energies. Equation 1 will be referred to as the SASA approach in the following.

More rigorous approaches can be derived by considering the binding process in more detail. At the next level of approximation, the solvent contribution to the nonpolar binding free energy has been divided into two parts.[13,14] First, a cavity in the solvent is created that can accommodate the solute. Then, the apolar solute is introduced into the cavity. The energy of the first step is usually estimated by multiplying some kind of molecular surface (MS) with the surface tension of water, γ_{wat},

$$\Delta G_{\text{cavity}} = \gamma_{\text{wat}} MS \quad (2)$$

Even though this resembles the SASA approach in eq 1, the two γ values are intrinsically different: γ_{SASA} in eq 1 is obtained from a linear fit, whereas γ_{wat} in eq 2 is a measurable physical constant of water. The energy of the second step can be estimated from the mean-field interaction energy between the solute and the solvent. In most force fields, it is modeled by a Lennard-Jones potential, i.e., by an attractive dispersion term and a repulsive exchange-repulsion term,

$$\Delta G_{\text{insertion}} = \Delta E_{\text{disp}} + \Delta E_{\text{rep}} \quad (3)$$

In a recent approach of Tan et al.,[15] ΔE_{rep} was merged into the cavity term, so that the nonpolar solvation free energy was estimated by the following:

$$\Delta G^{\text{CD}}_{\text{nonpolsolv}} = \gamma_{\text{CD}} MS + b_{\text{CD}} + \Delta E_{\text{disp}} = \Delta G_{\text{CR}} + \Delta E_{\text{disp}} \quad (4)$$

Several molecular surfaces and volume estimates (in the latter case, γ_{CD} is a pressure term) were examined in ref 15, but it was found that solvation free energies of small drug-like molecules were indifferent to the choice of the surface or volume (probably because these two measures are strongly correlated for small molecules). The dispersion term was estimated by a volume integral. Methods based on eq 4 are referred to as the cavity−dispersion (CD) method.

Another, still more complicated, approach is used in the polarized continuum model (PCM).[16] In this approach, a separate term for the exchange repulsion is also included, giving three terms for the nonpolar solvation free energy:

$$\Delta G^{\text{PCM}}_{\text{nonpolsolv}} = \Delta G_{\text{cavity}} + \Delta E_{\text{disp}} + \Delta E_{\text{rep}} \quad (5)$$

The cavity term is calculated from expressions of the radius of each atom to the power of 0 to 3,[17] i.e., including both area and volume terms. The two other terms are based on volume integrals.[18]

Recently, we have used PCM in an MM/GBSA framework to estimate ligand-binding affinities, i.e., we replaced the generalized Born and the SASA estimate with PCM calculations.[19] In those calculations, we observed a qualitative difference between the SASA and PCM estimates of the nonpolar solvation free energy. For the binding of a series of seven biotin analogous to avidin, SASA predicted a small and negative nonpolar contribution to the binding affinity, whereas the nonpolar part of PCM was three times larger and positive. Interestingly, the 3D-RISM method gives a nonpolar estimate of the solvation free energy similar to that of PCM for the same test case.[10] Of course, it is a major problem if different continuum solvation approaches give qualitatively different estimates of the same contribution. In this work, we use strict TI calculations of the absolute binding affinity to establish which of the SASA and PCM estimates is more reliable.

As a test case, we use the binding of benzene to an engineered apolar cavity in T4 lysozyme. This system has previously been used in theoretical studies because the system is experimentally well characterized[20] and the benzene molecule is small so it is possible to compute accurate absolute binding free energies.[21,22]

Methods

TI Calculations. The binding of a ligand to its receptor can be described with the thermodynamic cycle in Figure 1. This double-decoupling method has been employed extensively.[21−25] P is the protein and L is the ligand. L′ denotes an apolar ligand, i.e., a ligand with all the atomic charges set to zero, and *L″* denotes an apolar ligand where also the van der Waals interactions with the surroundings have been turned off. Restraints are applied to the ligand relative to the receptor so that a standard state can be defined. This also improves the convergence of the simulations.[23,25] The restraints are denoted by "restr" in Figure 1. According to the thermodynamic cycle in the figure, the binding free energy can be written as follows:

$$\Delta G_{\text{bind}} = \Delta G^{\text{free}}_{\text{ele}} + \Delta G^{\text{free}}_{\text{vdW}} - (\Delta G^{\text{bound}}_{\text{ele}} + \Delta G^{\text{bound}}_{\text{vdW}}) + \Delta\Delta G_{\text{restr}} \quad (6)$$

where the terms denote the free energy of removing the charges of the unbound ligand in bulk water, the free energy of turning off the van der Waals interaction of the unbound ligand, the free energy of removing the charges of the bound ligand, the free energy of turning off the van der Waals interaction of the ligand with the surrounding protein and water, and the free energy difference of applying and removing the restraints, $\Delta\Delta G_{\text{restr}} = -(\Delta G_{\text{restr+}} + \Delta G_{\text{restr−}})$. We define the nonpolar contribution to the free energy as $\Delta G_{\text{nonpol}} = \Delta G^{\text{free}}_{\text{vdW}} - \Delta G^{\text{bound}}_{\text{vdW}}$ and similar for the polar contribution, $\Delta G_{\text{pol}} = \Delta G^{\text{free}}_{\text{ele}} - \Delta G^{\text{bound}}_{\text{ele}}$.

The free energies associated with removing the charges or the van der Waals interactions were calculated with TI.[26] By TI, the free energy between two states is calculated as follows:

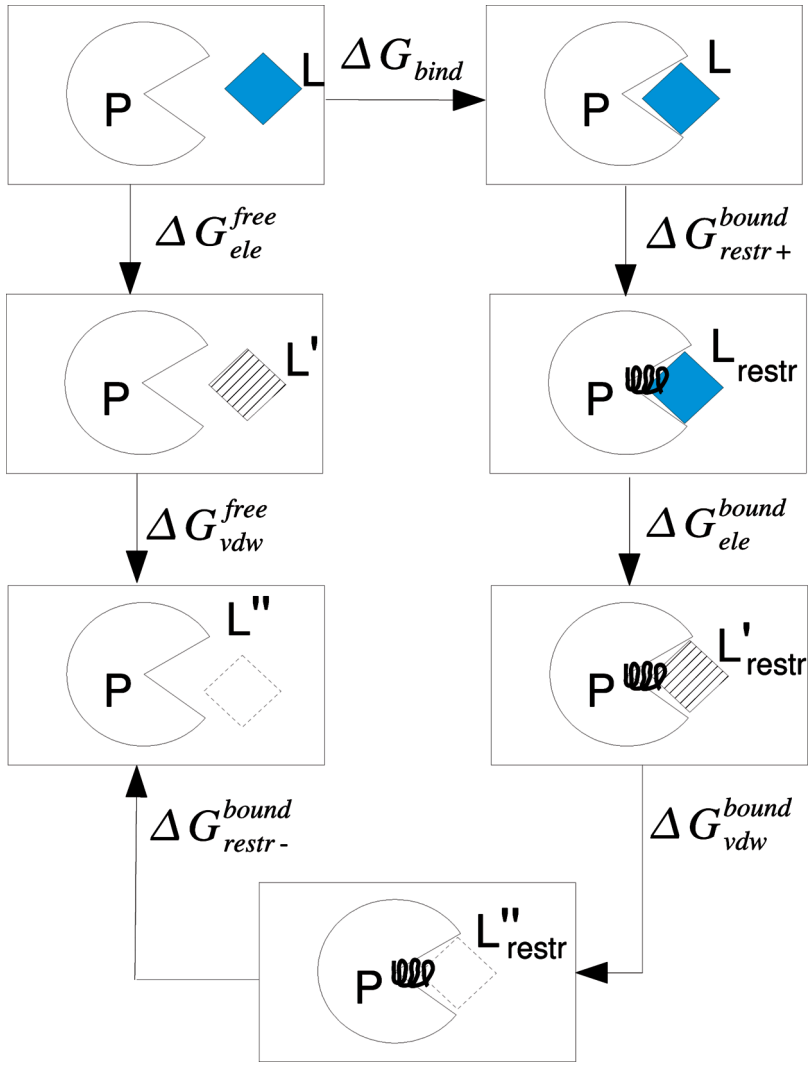

Figure 1. The thermodynamic cycle used to describe the binding of a ligand to its receptor.

$$\Delta G = \int_0^1 \left\langle \frac{\delta V}{\delta \lambda} \right\rangle_\lambda d\lambda \tag{7}$$

where the brackets indicate an ensemble average and V is defined as follows:

$$V(\lambda) = (1 - \lambda)V_0 + \lambda V_1 \tag{8}$$

where V_0 and V_1 are the potential energy of initial and final states, respectively. λ is a coupling parameter, describing the amount of the initial and final states in V. The integral in eq 7 was estimated by simulating the system at 11 distinct values of λ (0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95), followed by a trapezoid integration. The λ values at 0 and 1 where estimated with linear extrapolation from the two nearest points. Additional λ points were added where the formula $\delta V/\delta\lambda$ curve was less smooth or where there was a sharp change in the curvature, but no significant changes were obtained. The overlap between the $\delta V/\delta\lambda$ distribution at the different integration points was also checked. To avoid end-point problems in the van der Waals calculations, soft-core potentials, as implemented in the Amber simulation package, were used.[27]

The value of $\Delta\Delta G_{\text{restr}}$ can in principle be calculated by TI as well, but we used an approximation applicable in the limit of strong restraints.[21,24] The restraints are constructed with respect to three points in the protein (P_1, P_2, P_3) and three points in the ligand (L_1, L_2, L_3). Following previous work on lysozyme,[21] P_1 should be an atom lying close to the center of mass of the protein and was chosen to be the CB atom of Phe104, P_2 is the N atom in Met1, and P_3 is the C atom of Tyr161. For the benzene, three not connected carbon atoms were selected, C1, C3, and C5. One distance, two angles, and three dihedral angles were then defined based on these points, viz. the distance from P_1 to L_1 (r) the angles P_2-P_1-L_1 and P_1-L_1-L_2 (Θ and α), and the dihedral angles P_3-P_2-P_1-L_1, P_2-P_1-L_1-L_2 and P_1-L_1-L_2-L_3 (ϕ, β, and ζ). These geometric parameters were harmonically restrained to their values in the crystal structure with the force constants of 41.84 kJ/mol/Å^2 for the bond and 836.8 kJ/mol/rad^2 for the angles and the dihedral angles.[21] In the limit of strong restraints, the free energy of applying the restraints can be estimated as follows:[24]

$$\exp\left(-\frac{\Delta\Delta G_{\text{rest}}}{kT}\right) = C^\circ(2\pi)^{(3/2)} r_0^2 \sigma_r \sigma_\Theta \sigma_\phi \sin\theta_0 + \frac{\sigma_\alpha \sigma_\beta \sigma_\zeta \sin\alpha_0}{\sqrt{8\pi}} \tag{9}$$

where k is Boltzmann's constant, T the absolute temperature, C° is the standard concentration (1/1661 Å^{-3}), r_0, θ_0, and α_0 are the initial values of r, θ, and α, i.e., the values in the crystal structure, and the six σ values correspond to the standard deviations of the corresponding geometric parameter in an unrestrained simulation. These standard deviations were estimated from the same simulations that were used to compute the MM/GBSA estimates. A symmetry number of 12 for benzene was used for the rotational restraints.[21]

MM/GBSA Calculations. The binding free energy, ΔG_{bind}, of the benzene to T4 lysozyme was also estimated with the MM/GBSA method[28,29] according to the following:

$$\Delta G_{\text{bind}} = \langle G^{\text{PL}} \rangle - \langle G^{\text{P}} \rangle - \langle G^{\text{L}} \rangle \tag{10}$$

where PL is the protein−ligand complex, P is the protein, and L is the ligand. The brackets indicate ensemble averages over a MD simulation. Each free energy is estimated by the following:

$$G = E_{\text{int}} + E_{\text{ele}} + E_{\text{vdW}} + G_{\text{polsolv}} + G_{\text{nonpolsolv}} - TS_{\text{MM}} \tag{11}$$

where E_{int}, E_{ele}, and E_{vdW} are the internal, electrostatic, and van der Waals interactions, calculated with infinite cutoff. G_{polsolv} is the polar solvation energy, estimated by the generalized Born method of Onufriev et al., model I (GB^{OBCI}),[30] i.e., with $\alpha = 0.8$, $\beta = 0$, and $\gamma = 2.91$. $G_{\text{nonpolsolv}}$ is the nonpolar solvation energy, estimated from the SASA, using eq 1 with $\gamma_{\text{SASA}} = 0.0227$ kJ/mol/Å^2 and $b_{\text{SASA}} = 3.85$ kJ/mol.[29,31] In addition, we also estimated $G_{\text{nonpolsolv}}$ by the CD method,[15] using the σ decomposition and radii optimized by Tan et al.[32] For the cavity term, the solvent-accessible volume was used as the MS,[33] $\gamma_{\text{CD}} = 0.0378$ kJ/mol/Å^2, $b_{\text{CD}} = -0.5692$ kJ/mol, and the probe radius was 1.3 Å. For the dispersion term, a probe radius of 0.557 Å was used and the water density was set to 1.129 kg/L.[15,33] S_{MM} is the entropy, estimated from harmonic frequencies, calculated at the MM level[29] on a truncated and buffered system, as described previously, to improve the statistical precision of the estimate.[34] All the terms in eq 11 were averages over the last snapshot from 20 independent MD simulations. In general, more snapshots are required to obtain a high precision in ΔG_{bind},[35] but the $\Delta G_{\text{nonpolsolv}}$ term is usually very stable and 20 snapshots is enough to obtain converged results for this term. As is customary in MM/GBSA, the same geometry was used for all three reactants (i.e., only the complex was simulated).[29] Then, the E_{int} term cancels in eq 11. All MM/GBSA calculations were performed with the Amber 10 software.[36]

PCM Calculations. We also calculated $\Delta G_{\text{nonpolsolv}}$ with PCM on the same snapshots that were used to estimate MM/GBSA energies. We used the integral-equation formulation of PCM, IEFPCM,[37] which exhibits a better numerical stability than other formulations of PCM and is the default PCM method in the Gaussian software.[38] Owing to the large size of the molecular systems, the PCM induced charges were obtained using a direct inversion of the iterative subspace procedure,[39] as implemented in the GAMESS software.[40] Thus, no explicit matrix inversion is needed. We employed the UAKS radii as implemented in Gaussian[38] and a scaling factor for the polar part of 1.15.[19]

Explicit Calculations of Ligand-Solvent van der Waals Interaction. The van der Waals interaction energy between the benzene molecule and the surrounding water molecules was estimated using 20 independent simulations of benzene free in bulk water or bound to T4 lysozyme. The interaction energy was calculated every 5 ps using infinite cutoff, but with no correction for long-range interactions. The total van der Waals interaction energy was decomposed into repulsive and attractive contributions using three schemes as described in.[15] These were 12−6, which just separates the r^{-12} and r^{-6} terms in the Lennard-Jones potential, the Weeks−Chandler−Anderson (WCA) approach,[41] which separates the potential at the point where the force changes sign, and the σ approach,[15] which separates the potential at the point where the energy changes sign.

System Preparation. The calculations were based on the crystal structure 181L of the Leu99Ala mutant of T4 lysozyme in complex with benzene[42] and the protein was prepared as described previously.[43] All Asp and Glu residues were assumed to be negatively charged and all Lys and Arg residues were positively charged. The single histidine residue was assumed to be protonated on the ND1 atom, based on the local surroundings and the hydrogen-bond network. The protein was described by the Amber99SB force field[44] and the benzene molecule with the generalized Amber force field[45] with charges derived from a RESP calculation,[43] using electrostatic potentials calculated at the HF/6-31G* level in points sampled according to the Merz−Kollman scheme.[46] The protein−ligand system was immersed in a truncated octahedral box of TIP4P-Ewald waters,[47] extending at least 8 Å from the protein (7774 water molecules and 33711 atoms in total for the simulations of the complex).

MD Simulations. All MD simulations were run by the sander module of Amber 10.[36] The SHAKE algorithm[48] was used to constrain bonds involving hydrogen atoms so that a time step of 2 fs could be used. The temperature was kept constant at 300 K using a Langevin thermostat[49] with a collision frequency of 2.0 ps^{-1}. The pressure was kept constant at 1 atm using a weak-coupling isotropic algorithm[50] with a relaxation time of 1 ps. Particle-mesh Ewald summation,[51] with a fourth-order B spline interpolation and a tolerance of 10^{-5} was used to handle long-range electrostatics. The nonbonded cutoff was 8 Å and the nonbonded pair list was updated every 50 fs.

Unconstrained protein−ligand simulations, used in the MM/GBSA analysis, were generated in the following way: A single system was first energy minimized for 100 cycles of steepest descent, with all atoms, except water molecules and hydrogen atoms, restrained to their start position with a force constant of 418 kJ/mol/Å^2. This was followed by a 20 ps NPT simulation with the same restraints, and a 100 ps unconstrained NPT equilibration. From this equilibrated structure, 20 independent simulations were initiated by assigning random starting velocities. These simulations were further equilibrated for 50 ps in the NPT ensemble and then, a 200 ps production run were performed, also in the NPT

Table 1. ΔG_{nonpol} as Calculated by TI in kJ/mol

	total	solvent	solute
$\Delta G_{\text{vdW}}^{\text{bound}}$	38.7 ± 1.6	2.2 ± 0.0	36.5 ± 1.6
$\Delta G_{\text{vdW}}^{\text{free}}$	−5.0 ± 0.7	−5.0 ± 0.7	0
ΔG_{nonpol}[a]	−43.7 ± 1.7	−7.3 ± 0.7	−36.5 ± 1.6

[a] $\Delta G_{nonpol} = \Delta G_{\text{vdW}}^{\text{free}} - \Delta G_{\text{vdW}}^{\text{bound}}$.

ensemble. The same protocol was also used to simulate an apolar benzene molecule, both in the bulk and bound to T4 lysozyme.

The TI simulations were performed as follows: 10 independent simulations were initiated from the crystal structure by assigning random starting velocities. These were then simulated for 20 ps in a NPT ensemble with restraints to all atoms, except the hydrogen atoms and water molecules, as described above. This was followed by a 100-ps equilibration run and a 200-ps production run, both in the NPT ensemble. Snapshots were saved every 5 ps for analysis of $\delta V/\delta\lambda$. Restraints were applied in the bound simulations, as described above. The reported TI results are averages over the 10 simulations. A similar protocol has been used before.[21] For several perturbations, we tested to use 1 ns production simulations, but this did not change the results significantly.

Results

Nonpolar Solvation Energies. As detailed above, we have estimated the binding free energy of benzene to T4 lysozyme using TI and three approximate methods. This binding free energy has then been decomposed into various terms according to the MM/GBSA formalism to allow for the investigation of the nonpolar contribution to the solvation free energy.

We employed the TI method with double decoupling as a computational benchmark. This method estimated ΔG_{bind} to be −24 ± 2 kJ/mol (see below), which is in good agreement with the experimental value of −22 ± 1 kJ/mol.[20] This shows that the TI results are reliable and can be used as the benchmark for more approximate methods. Several other groups have studied the same system with free-energy methods, yielding results of −15 to −38 kJ/mol,[52] −36 to −38 kJ/mol,[53] −19.1 ± 0.8 kJ/mol,[22] and −24.9 ± 0.8 kJ/mol.[21] Thus, our results are among the best available.

All nonbonded terms are pairwise additive, so $\delta V/\delta\lambda$ can be decomposed on a residue-wise basis, making it possible to investigate the contributions from the solvent and the solute to the calculated free energies. Admittedly, these contributions may depend on how the decomposition is performed and therefore are not fully unambiguous. All contributions involving solvent molecules were assigned to the solvent and all the rest to the solute. Strictly, the TI free energies calculated for the bound state depend on the applied restraints, but it is assumed that these restraints have a negligible effect on the solvent molecules. In Table 1, the calculated ΔG_{nonpol}, as well as its components, $\Delta G_{\text{nonpol}} = \Delta G_{\text{vdW}}^{\text{free}} - \Delta G_{\text{vdW}}^{\text{bound}}$(cf. Figure 1), and the solvent and solute contributions are shown. Overall, $\Delta G_{\text{nonpol}} = -44$ kJ/mol and only −7 kJ/mol of these comes from the solvent molecules. The solvent contribution is favorable for the free ligand, whereas it is unfavorable in the complex (−5 and 2 kJ/mol; cf. Table 1). The $\delta V/\delta\lambda$ curves for the bound and free simulations are shown in Figure 2. It can be seen that the $\delta V/\delta\lambda$ curves are rather smooth, so the trapezoid integration is reliable.

Next, we estimated the binding free energy of benzene to T4 lysozyme within a MM/GBSA framework. As specified in the Methods section, we used three continuum methods to estimate the solvation energy. In the standard MM/GBSA approach, $\Delta G_{\text{polsolv}}$ is calculated by the generalized Born

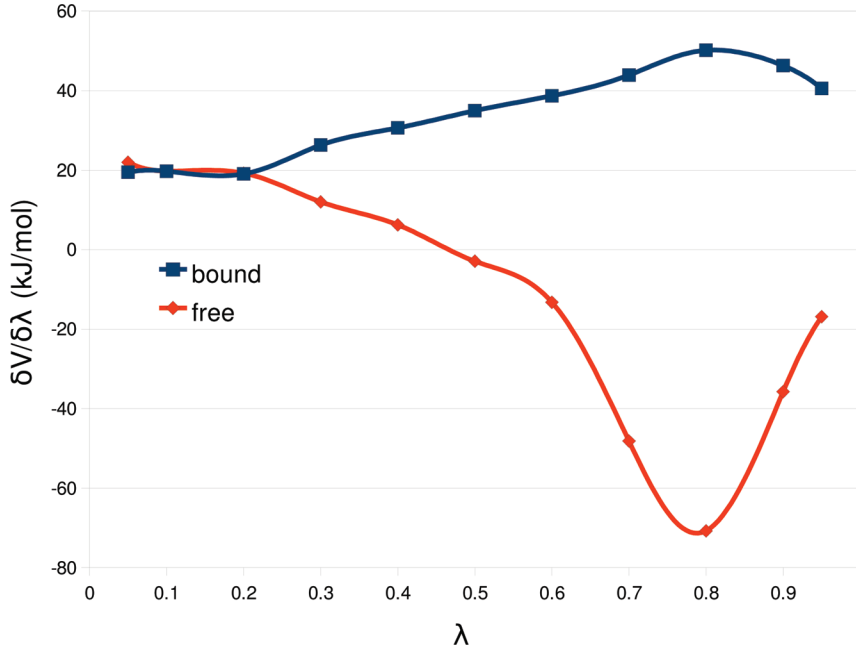

Figure 2. The derivative of the potential with respect to the coupling parameter ($\delta V/\delta\lambda$) in kJ/mol. The curves show $\Delta G_{\text{vdw}}^{\text{bound}}$ and $\Delta G_{\text{vdw}}^{\text{free}}$. The state $\lambda = 0$ contains a benzene molecule without charges, whereas in state $\lambda = 1$, the Lennard-Jones parameters of benzene have been turned off also.

Table 2. $\Delta G_{\text{nonpolsolv}}$ Calculated with Various Methods (kJ/mol)

	SASA	SASA(P0)[d]	CD	CD(P0)[d]	PCM	PCM(P0)[d]	TI
bound[a]	1.3 ± 0.6	0.0 ± 0.0	−27.8 ± 0.4	2.1 ± 0.1	−44.9 ± 1.0	1.1 ± 0.2	2.2 ± 0.0
free[b]	−9.4 ± 0.0	−9.4 ± 0.0	−8.5 ± 0.1	−8.5 ± 0.1	−4.5 ± 0.0	−4.5 ± 0.0	−5.0 ± 0.7
$\Delta G_{\text{nonpolsolv}}$[c]	−10.7 ± 0.6	−9.4 ± 0.0	19.2 ± 0.4	−10.7 ± 0.1	40.4 ± 1.0	−5.6 ± 0.2	−7.3 ± 0.7

[a] For TI this is $\Delta G^{\text{bound}}_{\text{vdWsolv}}$, for the other method it is the difference $\langle G_{\text{nonpolsolv}}(P)\rangle - \langle G_{\text{nonpolsolv}}(PL)\rangle$. [b] For TI this is $\Delta G^{\text{free}}_{\text{vdWsolv}}$, for the other methods it is $-\langle G_{\text{nonpolsolv}}(L)\rangle$. [c] $\Delta G_{\text{nonpolsolv}} = \Delta G^{\text{free}}_{\text{vdWsolv}} - \Delta G^{\text{bound}}_{\text{vdWsolv}}$. [d] Here, we assume that the ligand cavity in the free-protein calculations is filled with a dummy benzene molecule with zeroed charges, dispersion, and exchange-repulsion parameters.

Table 3. Components of the PCM and SASA Non-Polar Energy, As Well As Area and Volume Estimates for Lysozyme with (PL) and without (P) the Ligand (L)[a]

nonpolar energies (kJ/mol)	PL	P	L	P−PL	PL−P−L	P0	P0−PL	PL−P0−L
$\Delta G^{\text{PCM}}_{\text{cavity}}$	9319	9282	53	−38	−16	9319	0	−53
$\Delta G^{\text{PCM}}_{\text{disp}}$	−3325	−3335	−60	−10	70	−3324	1	58
$\Delta G^{\text{PCM}}_{\text{rep}}$	817	820	11	3	−13	817	0	−11
$\Delta G^{\text{PCM}}_{\text{nonpolsolv}}$	6811	6766	4	−45	40	6812	1	−6
$\Delta G^{\text{CD}}_{\text{CH}}$	5453	5452	55	−1	−54	5453	0	−55
$\Delta E^{\text{CD}}_{\text{disp}}$	−3281	−3307	−46	−26	73	−3278	2	44
$\Delta G^{\text{CD}}_{\text{nonpolsolv}}$	2173	2145	9	−28	19	2175	2	−11
$\Delta G^{\text{SASA}}_{\text{nonpolsolv}}$	209	210	9	1	−11	209	0	−9
area (Å^2)								
vdWS	16855	16765	91	−91	0	16855	0	−91
SES	7809	7975	90	166	−255	7809	0	−90
SAS	8590	8665	211	75	−286	8590	0	−211
volume (Å^3)								
vdWS	14056	13985	71	−71	0	14056	0	−71
SES	20627	20453	77	−174	97	20627	0	−77
SAS	31162	31134	276	−28	−248	31162	0	−276

[a] P0 is the free protein in the cavity of the complex. The areas were estimated by the molsurf program,[36] whereas the volumes were estimated by a local software with a grid spacing of 0.3 Å. In both cases, Bondi radii were used and a probe radius of 1.4 Å.

method and $\Delta G_{\text{nonpolsolv}}$ from the SASA, according to eq 1. Second, $\Delta G_{\text{nonpolsolv}}$ was estimated by the CD method, which divides $\Delta G_{\text{nonpolsolv}}$ into a cavity and dispersion term (eq 4). Third, we replaced the entire solvation free energy estimate with that obtained by PCM as implemented in the GAMESS software.[39] The three $\Delta G_{\text{nonpolsolv}}$ estimates are compared to the result of the TI calculations in Table 2. To make the analysis easier, we divided the free energy into contributions from the bound and free states. For TI, this is $\Delta G^{\text{free}}_{\text{vdWsolv}}$ and $\Delta G^{\text{bound}}_{\text{vdWsolv}}$(cf. Figure 1, divided into solvent and solute contributions), and for the other methods it is $\langle G_{\text{nonpolsolv}}(P)\rangle - \langle G_{\text{nonpolsolv}}(PL)\rangle$ and $-\langle G_{\text{nonpolsolv}}(L)\rangle$, respectively.

From Table 2, it can be seen that SASA gives almost the same result as TI for the nonpolar solvation energy in the bound state, 1 kJ/mol compared to 2 kJ/mol. In contrast, the CD method gives the wrong sign; it estimates a negative free energy of −28 kJ/mol. PCM also gives the wrong sign; it assigns a negative free energy of −45 kJ/mol to the process.

To gain some further understanding of this discrepancy, the PCM nonpolar solvation energy was divided into its three contributions, viz. the cavitation, the dispersion, and exchange-repulsion energies (Table 3). These are −38, −10, and 3 kJ/mol, respectively. Thus, the PCM nonpolar solvation energy is dominated by the cavitation term. This gives us a clue to the cause of the problem: PCM predicts an appreciable difference in the size of the complex and the free protein, i.e., it predicts that the benzene cavity in lysozyme is filled with solvent in the free protein. This is also confirmed by the other two PCM terms: Both the dispersion and exchange repulsion are larger in the free protein than in the complex (i.e., the dispersion term is more negative and the exchange-repulsion term more positive; cf. Table 3). Considering that the binding site in lysozyme is completely hidden inside the protein (cf. Figure 3), this shows that PCM considers the cavity to be filled with water in the free protein.

On the other hand, the MD simulations do not show any water molecules in the cavity in the unbound state. Likewise, Figure 4 shows that $\delta V/\delta\lambda$ for $\Delta G^{\text{bound}}_{\text{vdWsolv}}$ is flat, indicating that the water molecules do not go through any large changes as we remove the benzene molecule. This is also supported by experiments, which show that the cavity in the protein is empty at normal pressure[54] and that there is no change in the general shape of the protein when benzene binds.

However, this is not necessarily an erroneous behavior of PCM. On the contrary, large protein cavities are typically filled with water molecules, and also smaller cavities may contain water.[55] Therefore, it cannot be known beforehand whether a cavity is filled with water molecules or not. It is quite natural to let the continuum models assume that all cavities are water-filled, although it would be better if this could be affected as an input option.

In Table 3, we also provide estimates of the area and volume of lysozyme (from the crystal structure). Unfortunately, there are several definitions of the surface of a molecule. The van der Waals surface (vdWS) is the union of the van der Waals surfaces of all atoms in the molecule. The solvent-excluded surface (SES) is the surface accessible to a solvent probe, i.e., the vdWS, but excluding crevices and cavities that are not large enough to room a solvent molecule. Finally, the solvent-accessible surface (SAS) is

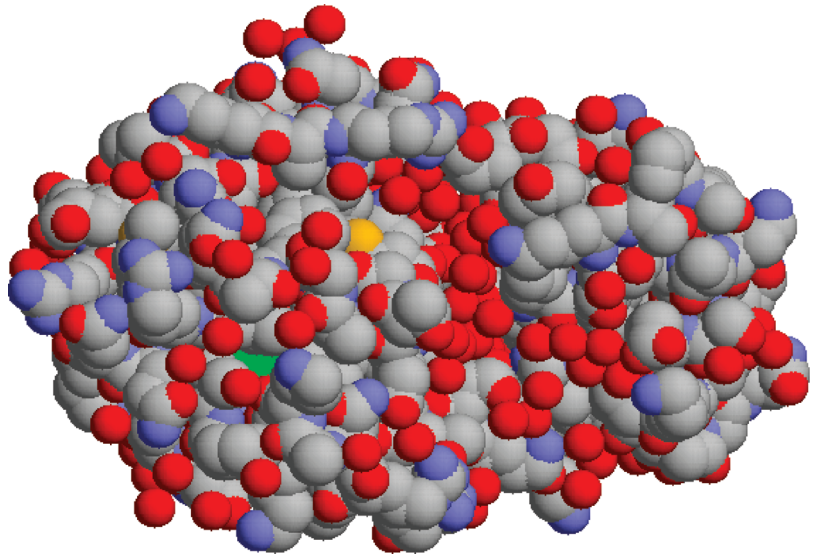

Figure 3. Space-filling model of lysozyme, showing that the benzene molecule (light green) is almost completely buried in the protein.

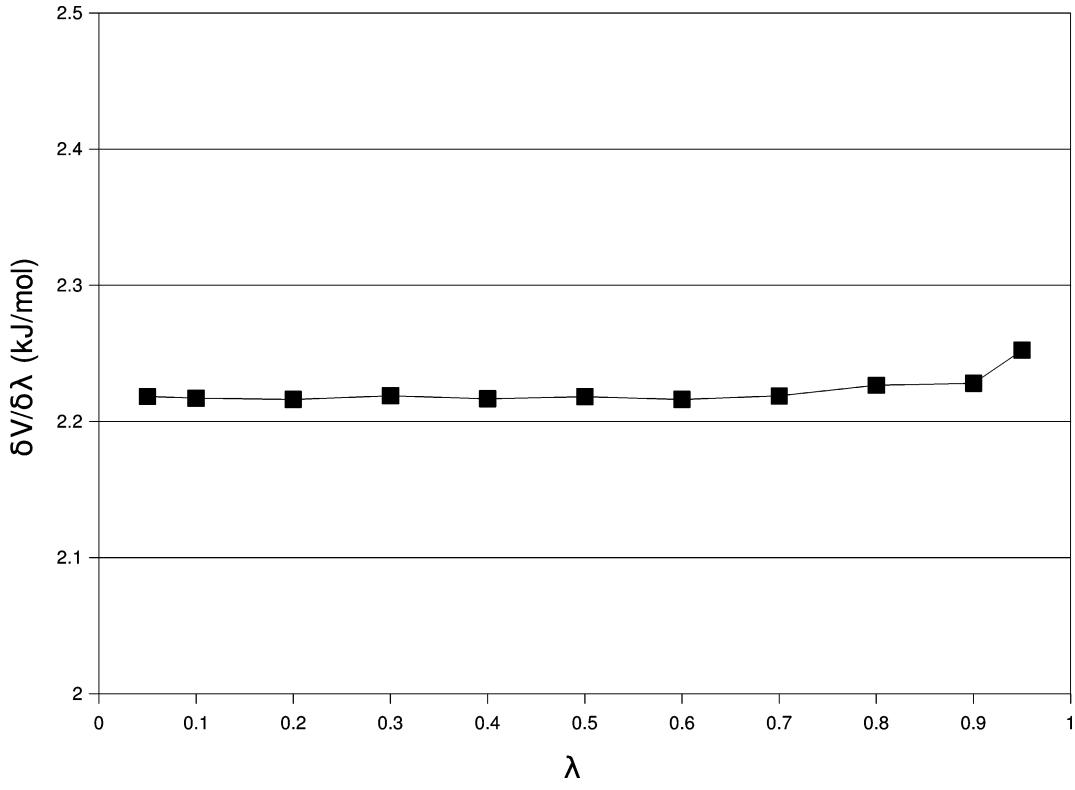

Figure 4. The derivative of the potential with respect to the coupling parameter in kJ/mol for the $\Delta G^{\text{bound}}_{\text{vdWsolv}}$ process.

the surface followed by the center of the solvent probe when rolling on the vdWS. In practice, the SAS is equivalent to a vdWS for which all atomic van der Waals radii have been enhanced by the solvent-probe radius.

The various solvation terms employ different size measures: $\Delta G^{\text{SASA}}_{\text{nonpolsolv}}$ is of course based on the SAS area, as are the PCM dispersion and repulsion terms. However, the polar PCM solvation energy is based on the SES area, whereas the PCM cavity energy is based on the vdWS, but with both area and volume terms.[56] As an effect, the PCM cavity term shows a different behavior than the other two PCM terms: As mentioned above, the absolute value of the dispersion and exchange-repulsion terms is larger for the free protein than for the complex, in agreement with the SAS area (cf. Table 3). On the other hand, the cavity term is larger in the complex, in agreement with both the vdWS area term and all volume terms. It is not clear whether the area or volume terms dominate the cavity term, but it seems questionable to use the vdWS in a protein, which will include crevices and cavities that are too small to room a water molecules (in fact, all atoms in lysozyme contributes to the vdWS, whereas only 59% of the atoms contribute to the SAS). As can be seen from Table 3, the vdWS gives the opposite sign for the prediction of the difference between

Table 4. ΔG_{solv} of Benzene (in kJ/mol) Estimated by Four Different Methods[a]

Method	ΔG_{solv}
GB/SASA	-8.3 ± 0.0
GB/CD	-9.1 ± 0.2
PCM	-4.6 ± 0.0
TI	-0.1 ± 0.7
experiment[12]	−3.6

[a] For TI, this energy is $-(\Delta G^{free}_{vdWsolv} + \Delta G^{free}_{elesolv})$, whereas for the other methods it is $<G_{nonpolsolv}(L) + G_{polsolv}(L)>$. A small gas-phase contribution has been neglected.

the complex and the protein. PCM was calibrated for small molecules, for which there is little difference between the vdWS and SAS areas, but for large molecules, they are widely different, as can be seen in Table 3.

Thus, we can conclude that the prime reason for the poor result of the PCM solvation model is the fact that it predicts that the lysozyme cavity is full with water in the free protein. A simple way to avoid this problem is to recalculate the PCM solvation energy for the free protein in the cavity of the complex (in practice, this is obtained by using a dummy benzene molecule with all charges, as well as all dispersion and exchange-repulsion coefficients zeroed). Thereby, we prevent the cavity from being filled with water molecules. As a consequence, the cavitation energy will cancel between the complex and the free protein, whereas the other terms will change somewhat, as shown in Table 3 (the polar solvation energy also changes by 3 kJ/mol). The net bound $\Delta G_{nonpolsolv}$ is shown in Table 2 (column PCM(P0)). It can be seen that now PCM gives a result close to that of TI, 1 kJ/mol, compared to 2 kJ/mol. This supports our suggestion that the main problem with PCM is that it assumes that the cavity in lysozyme is filled with water.

Interestingly, the SASA and CD models also assume that the cavity in lysozyme is filled with solvent. This can for example be seen from the fact that the SASA in Table 3 changes between the free protein (P) and the complex (PL), giving a positive sign of the bound SASA term in Table 2, which shows that the SASA of the free protein is larger than that of the complex. However, the SASA term is much smaller than the corresponding PCM terms, so that it does not cause any problems.

Therefore, we can cure the CD and SASA results in the same way as for PCM. This changes $G_{nonpolsolv}$ for CD from −28 to 2 kJ/mol, i.e., to a close agreement with the TI result (Table 2, column CD(P0)). Thus, this simple approach also cures the poor results of the CD approach. However, the polar GB solvation energy for the free protein changes by 8 kJ/mol. For SASA, the result hardly changes, because $\Delta G_{nonpolsolv}$ for the bound state was only 1 kJ/mol (Table 2, column SASA(P0).

For the free ligand, TI predicts a decrease in $\Delta G_{nonpolsolv}$ of −5 kJ/mol. This value is closely reproduced by PCM (−4 kJ/mol),whereas the SASA and CD results are slightly more negative (−9 kJ/mol).

These results can be combined with the polar solvation energy of the free ligand to give the solvation free energy of benzene (neglecting a small gas-phase correction), which experimentally is −4 kJ/mol.[12] From Table 4, it can be seen that PCM obtains a value close to experiments (−5 kJ/mol). The GB/SASA and GB/CD estimates are 4−5 kJ/mol too negative, whereas the TI estimate is 3 kJ/mol too positive. Thus all three continuum methods give similar results for the solvation free energy of the free ligand, which is expected, because they were calibrated to give accurate solvation energies for small ligand. Major differences are observed only for large ligands and macromolecules, for which experimental solvation energies are missing, and for which the various area and volume definitions give differing results.

An interesting consequence of the differing surface definitions for the nonpolar terms is that PCM and SASA give the opposite sign for $\Delta G_{nonpol\ solv}$, which was also observed for the binding of seven biotin analogues to avidin.[19] In general, both the vdWS area and volume terms nearly cancel for the net PL−P−L term, as can be seen in Table 3. Therefore, the net PCM cavity term also nearly cancels. Among the remaining PCM terms, the dispersion term is larger than the repulsion term in all cases we have studied, and therefore the net PCM nonpolar solvation energy is positive (the SAS area is larger for P than for PL, and this difference is enhanced by the ligand, giving a negative result for PL−P−L; then the dispersion is always negative, i.e., attractive). However, the SASA term is always positive (repulsive) and consequently, the nonpolar PCM and SASA energies always have the opposite sign for a buried cavity. Alternatively, if the cavity is empty, then all PCM terms are nearly identical for PL and P0, so the net nonpolar solvation energy is close to the negative of that of the ligand, which can have any sign, depending on the relative size of the dispersion and the sum of the repulsion and cavity terms.

Dispersion and Repulsion Terms. The dispersive and repulsive parts of PCM and the dispersion part in the CD method can be further analyzed by relating them, in a mean-field approximation, to the average van der Waals interaction between the apolar benzene molecule and the surrounding water molecules.[15] Therefore, explicit simulations of an apolar benzene molecule both bound to T4 lysozyme and free in bulk water were performed. For such a comparison of explicit simulations with the approximate methods, it is necessary to decompose the Lennard-Jones potential into attractive and repulsive parts.[14] We have used three common schemes for such a decomposition, as described in the Methods section. The explicit simulations are compared to PCM and CD in Table 5. The van der Waals interactions in the bound simulations are small and all three decomposition schemes give the same results (because all water molecules are far away from the benzene molecules). The small value is consistent with the TI results as discussed above. However, the PCM and CD continuum models give much larger values with the opposite sign. The reason is that they assume a water-filled cavity, as discussed above. If we force the cavity to be empty, both the PCM and CD results (PCM(P0) and CD(P0) in Table 5) closely reproduce the results of the explicit simulation.

The van der Waals interactions in the simulation of the free ligand are larger and negative. In this case, different decomposition methods give different contributions for

Table 5. Average Benzene−Water van der Waals Interaction Energies (kJ/mol) in Explicit Simulations (Using Three Different Decomposition Schemes) Compared to the PCM and CD Estimates[a]

	free			bound		
method	attractive	repulsive	total	attractive	repulsive	total
explicit (12−6)	−81.8 ± 0.4	38.4 ± 0.4	−43.4 ± 0.2	1.3 ± 0.0	0.0 ± 0.0	1.3 ± 0.0
explicit (WCA)	−52.1 ± 0.1	8.7 ± 0.2	−43.5 ± 0.2	1.3 ± 0.0	0.0 ± 0.0	1.3 ± 0.0
explicit (σ)	−48.8 ± 0.1	5.3 ± 0.1	−43.5 ± 0.2	1.3 ± 0.0	0.0 ± 0.0	1.3 ± 0.0
PCM	−59.5 ± 0.0	10.7 ± 0.0	−48.8 ± 0.0	−10.0 ± 0.5	2.8 ± 0.2	−7.3 ± 0.3
PCM(P0)[b]	−59.5 ± 0.0	10.7 ± 0.0	−48.8 ± 0.0	1.1 ± 0.2	0.0 ± 0.0	1.1 ± 0.2
CD	−46.3 ± 0.1		−46.3 ± 0.1	−26.5 ± 0.4		−26.5 ± 0.4
CD(P0)[b]	−46.3 ± 0.1		−46.3 ± 0.1	2.1 ± 0.1		2.1 ± 0.1

[a] The results for the free state are taken from the simulations with the free ligand in water. The results for the bound state are the negative interaction energy between the ligand and the solvent in the explicit simulations, and the difference between simulations of the free protein and of the complex for the continuum methods. [b] Here, we assume that the ligand cavity in the free-protein calculations is filled with a dummy benzene molecule with zeroed charges, dispersion, and exchange-repulsion parameters.

Table 6. Contributions to ΔG_{bind} (kJ/mol) for Benzene Binding to Lysozyme, Calculated with Four Variants of MM/GBSA and with TI[a]

	MM/GBSA		MM/GBCD		MM/PCM			
Contribution	P	P0[b]	P	P0[b]	P	P0[b]	TI	Exp.
Polar, solute	−6.6 ± 0.4	−6.6 ± 0.4	−6.6 ± 0.4	−6.6 ± 0.4	−6.6 ± 0.4	−6.6 ± 0.4	0.5 ± 0.3	
Polar, solvent	26.3 ± 0.3	26.3 ± 0.3	26.3 ± 0.3	18.1 ± 0.3	11.8 ± 0.3	8.8 ± 0.3	5.3 ± 0.2	
Nonpolar, solute	−11.0 ± 3.2	−11.0 ± 3.2	−11.0 ± 3.2	−11.0 ± 3.2	−11.0 ± 3.2	−11.0 ± 3.2	−23.0 ± 1.7	
Nonpolar, solvent	−10.7 ± 0.1	−9.4 ± 0.0	19.2 ± 0.5	−10.7 ± 0.1	40.4 ± 1.0	−5.6 ± 0.2	−7.3 ± 0.7	
Sum	−1.9 ± 3.1	−8.9 ± 3.0	27.9 ± 3.0	−10.2 ± 3.0	34.6 ± 3.2	−14.4 ± 3.0	−24.5 ± 2.1	−21.7

[a] For the continuum solvation methods, the terms (polar solute, polar solvent, nonpolar solute, and nonpolar solvent) are $\langle\Delta E_{\text{ele}}\rangle$, $\langle\Delta G_{\text{polsolv}}\rangle$, $\langle\Delta E_{\text{vdW}}\rangle - \langle T\Delta S\rangle$, and $\langle\Delta G_{\text{nonpolsolv}}\rangle$, respectively. For TI, the terms are $\Delta G^{\text{free}}_{\text{elesolu}} - \Delta G^{\text{bound}}_{\text{elesolu}}$, $\Delta G^{\text{free}}_{\text{elesolv}} - \Delta G^{\text{bound}}_{\text{elesolv}}$, $\Delta G^{\text{free}}_{\text{vdWsolu}} - \Delta G^{\text{bound}}_{\text{vdWsolu}} + \Delta\Delta G_{\text{rest}}$, and $\Delta G^{\text{free}}_{\text{vdWsolv}} - \Delta G^{\text{bound}}_{\text{vdWsolv}}$, respectively. [b] Here, we assume that the ligand cavity in the free-protein calculations is filled with a dummy benzene molecule with zeroed charges, dispersion, and exchange-repulsion parameters.

dispersion and repulsion. The CD method should be comparable to the σ decomposition of the explicit simulations and in this case, the two results agree within 3 kJ/mol. PCM gives results that resemble the WCA decomposition, although the original implementation used a 12−6 decomposition.[57]

Binding Free Energies. Finally, we have collected in Table 6 all contributions to the binding free energy of benzene to T4 lysozyme, divided into polar and nonpolar contributions, as well as solvent and solute contributions. The entropic term in MM/GBSA is added to the nonpolar contribution of the solute and the contribution from releasing the TI restraints (17 kJ/mol) is assigned to the same contribution, because this is mainly an entropic term. We have tested this approximation by calculating $\Delta G^{\text{bound}}_{\text{ele}}$ without applying any restraints: Comparing the simulation with and without restraints, $\Delta G^{\text{free}}_{\text{ele}} - \Delta G^{\text{bound}}_{\text{ele}}$ increases by 0.6 kJ/mol and the solute part by 0.2 kJ/mol.

It can be seen from Table 6 that all four contributions have sizable errors, compared to the TI simulations. The nonpolar solvent contribution has already been discussed. The polar contribution of the solute is the same for all continuum approaches ($\langle\Delta E_{\text{ele}}\rangle$), providing a 7 kJ/mol too negative estimate. Likewise, the nonpolar contribution of the solute is the same for all continuum methods ($\langle\Delta E_{\text{vdW}}\rangle - \langle T\Delta S\rangle$), −11 kJ/mol, which is 12 kJ/mol less negative than the TI estimate. Thus, the errors in these two terms partly cancel. However, the polar solvation free energy depends on whether we use GB or PCM. Both methods give too positive values, 26 kJ/mol for GB and 12 kJ/mol for PCM, compared to 5 kJ/mol for TI. Consequently, the conventional MM/GBSA approach estimates a binding free energy that is ∼20 kJ/mol too unfavorable, as reported previously.[43] Replacing $\Delta G_{\text{nonpolsolv}}$ from SASA with the CD estimate, results in a slightly better agreement with the experimental ΔG_{bind}, but only if we assume that the cavity in lysozyme is not filled with waters (MM/GBCD(P0)). Otherwise, this method gives a positive ΔG_{bind}. Replacing GBSA with PCM results in a binding free energy that is 60 kJ/mol too unfavorable. However, this result is strongly improved if we assume that the cavity in lysozyme is not filled with water molecules (MM/PCM(P0) in Table 6), which gives an estimate that is 10 kJ/mol too positive. Likewise, the MM/GBSA result is improved to −9 kJ/mol if the cavity is not filled with water molecules, mainly owing to the GB solvation energy. Altogether, PCM without a water-filled cavity gives the best results, with a maximum error compared to the TI components of 12 kJ/mol (for the nonpolar solute term) and a mean absolute deviation of 6 kJ/mol for the four terms, whereas the SASA(P0) and CD(P0) methods give maximum component errors of 13 kJ/mol (polar solvent) and mean absolute deviations of 8−9 kJ/mol. Of course, it is likely that some of the differences between TI and MM/GBSA comes from effects other than the solvation model, e.g., that TI uses alchemical transformations, whereas MM/GBSA only uses end-point simulations.

Conclusions

In this work, we explain why the SASA and PCM continuum solvation methods give qualitatively different estimates of

the nonpolar solvation energy. Using the binding of benzene to the Leu99Ala T4 lysozyme mutant as a test case and absolute binding free energies estimated with TI and a double-decoupling scheme with restraints as a benchmark, we show that the simple SASA approach gives an appreciably more accurate estimate of the nonpolar solvation energy than the PCM approach. However, the reason for the poor performance of PCM for this case is that it assumes that the benzene cavity in the free protein is filled with water. In general, this is a proper assumption of a continuum solvation method, but in the present case, both calculations and experiments indicate that the cavity is water-free when the ligand is not bound.[54] We suggest a simple way to correct this problem, and then PCM gives the best results among the continuum methods, providing better estimates of each component than the other methods. Similar approaches with inserted dummy atoms have been used before to avoid ligand cavities to be filled with continuum water.[5,58,59] However, PCM uses the van der Waals surface for the calculation of the cavitation energy, which is questionable for a protein, because it leads to numerous small crevices and cavities inside the protein, which are too small to house a solvent molecule, and it also leads to opposite trends regarding the change in surface area of the protein upon ligand binding.

We have also included in the investigation a recent two-term expression consisting of a cavitation and a dispersion energy (CD method). This method also assumes a water-filled cavity for the free protein and the results are strongly improved if this problem is cured.

Interestingly, the SASA method makes the same assumption of a water-filled cavity, but the SASA terms are much smaller, so it does not significantly affect the nonpolar solvation energy (however, the polar solvation energy is improved, giving a better net result). In fact, for this test case, the SASA and CD methods give similar results, and that with PCM is only slightly better. The reason for this is most likely that the SASA terms are smaller and therefore more stable. As can be seen from Table 3, the three terms involved in the nonpolar solvation energies are large but of opposite signs. Therefore, it is a formidable task to obtain a sum that is stable and accurate. It is easier to obtain stable results for the sum, which is much smaller in magnitude. The integration over Lennard-Jones terms as used for the dispersion term in both CD and PCM, and for the exchange-repulsion term in PCM are probably as accurate as you can come with a continuum approximation at the MM level. The main problem is most likely the cavity term, which is not a plain average from a simulation (it involves the entropy).

Of course, our results are valid strictly for the present lysozyme test case only, with its small and buried ligand-binding cavity that is void of water molecules in the apo state. However, the problem with the empty cavity in the free protein is general and emphasizes how important it is to know the hydration state of the ligand cavity. Even if the cavity contains water in the free state, it is most likely that these water molecules are not bulk-like. This can be expected to give major problems for continuum models, but also for TI and FEP methods if the cavity is buried in the protein so that the equilibration of waters between the cavity and bulk is slow.

Acknowledgment. This investigation has been supported by grants from the Swedish Research Council and from the Research School in Pharmaceutical Science. It has also been supported by computer resources of Lunarc at Lund University, C3SE at Chalmers University of Technology and HPC2N at Umeå University. J.K. thanks the Danish Research Council for financial support.

References

(1) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 4–16.

(2) Huang, S.-Y.; Zou, X. *J. Chem. Inf. Model.* **2010**, *50*, 262–273.

(3) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187–4225.

(4) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.

(5) Michael, J.; Verdonk, M. L.; Essex, J. W. *J. Med. Chem.* **2006**, *49*, 7427–7439.

(6) Chen, J.; Brooks, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18* (18), 140–148.

(7) Wagoner, J. A.; Baker, N. A. *Proc. Natl. Acad. Sci.* **2006**, *103*, 8331–8336.

(8) Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919–930.

(9) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2003**, *25*, 238–250.

(10) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. *J. Phys. Chem. B* **2010**, *114*, 8505–8516.

(11) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754–2759.

(12) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(13) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517–529.

(14) Westergren, J.; Lindfors, L.; Höglund, T.; Lüder, K.; Nordholm, S.; Kjellander, R. *J. Phys. Chem. B* **2007**, *111*, 1872–1882.

(15) Tan, C.; Tan, Y.-H.; Luo, R. *J. Phys. Chem. B* **2007**, *111*, 12263–12274.

(16) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210–3221.

(17) Cossi, M.; Tomasi, J.; Cammi, R. *Int. J. Quant. Chem. Quant. Chem. Symp.* **1995**, *29*, 695–702.

(18) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616–627.

(19) Söderhjelm, P.; Kongsted, J.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6*, 1726–1737.

(20) Morton, A.; Baase, W. A.; Matthews, B. W. *Biochem.* **1995**, *34*, 8564–8575.

(21) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

(22) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

(23) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.

(24) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798–2814.

(25) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B.* **2003**, *107*, 9535–9551.

(26) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–314.

(27) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108–2141021.

(28) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *37*, 9401–9809.

(29) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham III, T. E. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(30) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.

(31) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786–3791.

(32) Tan, C. H.; Yang, L. J.; Luo, R. *J. Phys. Chem. B* **2006**, *110*, 18680–18687.

(33) Wang, J.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Tan, C.; Luo, R. *Amber Tools User's Manual, Version 1.4* **2010**, 143–150.

(34) Kongsted, J.; Ryde, U. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 63–71.

(35) Genheden, S.; Ryde, U. *J. Comput. Chem.* **2010**, *31*, 837–846.

(36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 10*, University of California: San Francisco, 2008.

(37) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(38) Frisch, A. E.; Frisch, M. J.; Trucks, G. W. *Gaussian 03 User's Reference*; Gaussian, Inc.: Wallingford, CT, USA, 2003, p. 205.

(39) Li, H.; Pomelli, C. S.; Jensen, J. H. *Theor. Chem. Acc.* **2003**, *109*, 71–84.

(40) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(41) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(42) Morton, A.; Matthews, B. W. *Biochem.* **1995**, *34*, 8576–8588.

(43) Genheden, S.; Ryde, U. *J. Comput. Chem.* 2010 available online, DOI: 10.1002/jcc.21546.

(44) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinform.* **2006**, *65*, 712–725.

(45) Wang, J. M.; Wolf, R. M.; Caldwell, K. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(46) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.

(47) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(48) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(49) Wu, X.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512–518.

(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(51) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(52) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

(53) Rodinger, T.; Howell, P. L.; Romes, R. *J. Chem. Phys.* **2008**, *129*, 155102–155113.

(54) Collins, M. D.; Hummer, G.; Quillin, M. L.; Matthews, B. W.; Gruner, S. M. *Proc. Natl. Acad. Sci.U.S.A.* **2005**, *102*, 16668–16671.

(55) Qvist, J.; Davidovic, M.; Hamelberg, D.; Halle, B. *Proc. Natl. Ac. Sci. U.S.A.* **2008**, *105*, 6296–6301.

(56) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1990**, *255*, 327–335.

(57) Floris, F. M.; Tomasi, J. *J. Comput. Chem.* **1991**, *12*, 784–791.

(58) Zou, X.; Sun, Y.; Kuntz, I. D. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.

(59) Lie, H.-Y.; Kuntz, I. D.; Zou, X. *J. Phys. Chem. B* **2004**, *108*, 5453–5462.

CT100272S