HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2017 March 07.

Published in final edited form as:

J Chem Theory Comput. 2015 October 13; 11(10): 4593–4600. doi:10.1021/acs.jctc.5b00684.

Examining the Assumptions Underlying Continuum-Solvent Models

Robert C. Harris and **B. Montgomery Pettitt***

Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Blvd, Galveston, Texas 77555-0304, United States

Abstract

Continuum-solvent models (CSMs) have successfully predicted many quantities, including the solvation-free energies (ΔG) of small molecules, but they have not consistently succeeded at reproducing experimental binding free energies ($\Delta\Delta G$), especially for protein–protein complexes. Several CSMs break ΔG into the free energy (ΔG_{vdw}) of inserting an uncharged molecule into solution and the free energy (ΔG_{el}) gained from charging. Some further divide ΔG_{vdw} into the free energy (ΔG_{rep}) of inserting a nearly hard cavity into solution and the free energy (ΔG_{att}) gained from turning on dispersive interactions between the solute and solvent. We show that for 9 protein–protein complexes neither ΔG_{rep} nor ΔG_{vdw} was linear in the solvent-accessible area A, as assumed in many CSMs, and the corresponding components of $\Delta\Delta G$ were not linear in changes in A. We show that linear response theory (LRT) yielded good estimates of ΔG_{att} and $\Delta\Delta G_{att}$, but estimates of $\Delta\Delta G_{att}$ obtained from either the initial or final configurations of the solvent were not consistent with those from LRT. The LRT estimates of ΔG_{el} differed by more than 100 kcal/mol from the explicit solvent model's (ESM's) predictions, and its estimates of the corresponding component ($\Delta\Delta G_{el}$) of $\Delta\Delta G$ differed by more than 10 kcal/mol. Finally, the Poisson–Boltzmann equation produced estimates of ΔG_{el} that were correlated with those from the ESM, but its estimates of $\Delta\Delta G_{el}$ were much less so. These findings may help explain why many CSMs have not been consistently successful at predicting $\Delta\Delta G$ for many complexes, including protein–protein complexes.

Graphical Abstract

***Corresponding Author**, Phone: 409 7720723. Fax: 409 7720725. mpettitt@utmb.edu.

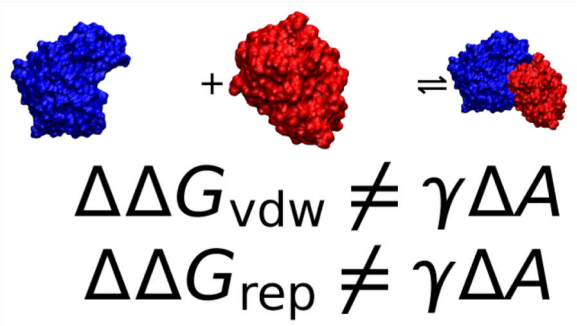
ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00684.

Scatter plots and data analyses (PDF)

The authors declare no competing financial interest.



1. INTRODUCTION

Implicit solvent models, such as the Poisson–Boltzmann equation[1] (PBE), generalized Born (GB) models,[2] proximal distribution approaches,[3–5] and integral equation methods,[6,7] provide estimates of the solvation energies (ΔG) of biomolecules more quickly than explicit solvent models (ESMs). Implicit solvent models are faster than ESMs because they approximately integrate over the degrees of freedom of the aqueous solvent in the partition function. Once estimates of ΔG have been obtained, they can be used to obtain estimates of other free energies, such as binding ($\Delta\Delta G$) and mutation free energies.[8,9] Implicit solvent models may be divided into two classes, continuum solvent models (CSMs),[1,2] such as the PBE and GB models that model water as a high-dielectric continuum, and those that approximate the distribution of the solvent,[3–7] such as integral equations, density functional theories, and other structured approaches. CSMs have been fairly successful at predicting many quantities, such as ΔG, and the salt dependencies of various free energies,[10–14] but they have not been uniformly successful at predicting some other quantities, such as $\Delta\Delta G$.[15–21]

Typically, CSMs break ΔG into the free energy (ΔG_{vdw}) required to insert an uncharged molecule into solution and the free energy (ΔG_{el}) gained by turning on the partial atomic charges.[22–24] Some researchers have then further divided ΔG_{vdw} into the free energy (ΔG_{rep}) required to insert a nearly hard cavity into solution and the free energy (ΔG_{att}) gained from turning on the dispersive interactions between the solute and solvent.[25–33] The Methods contains formal definitions of these and related quantities.

In previous work, we demonstrated that the methods used to compute ΔG_{rep}, ΔG_{att}, and ΔG_{vdw} in some CSMs were unsatisfactory for alanine and glycine peptides.[24,34] In this work, we extend this analysis to nine protein–protein complexes covering wide ranges of sizes, biological functions, and equilibrium binding constants. By doing so, we try to ensure that our results are generally applicable to protein–protein complexes and not artifacts of our choice of examples. The complexes with their PDB codes are bovine α-chymotrypsin with eglin c complex (1ACB),[35] porcine pancreatic trypsin with soybean trypsin inhibitor complex (1AVX),[36] bovine β-lactoglobulin, which is a homodimer (1BEB),[37] barnase–barstar complex (1BRS),[38] colicin E9 dnase domain with IM9 (1EMV),[39] *Pseudomonas aeruginosa* exos toxin with human rac (1HE1),[40] bovine β-trypsin with CMTI-I (1PPE),[41] and uracil-dna glycosylase with uracil glycosylase inhibitor (1UDI).[42] In the following

sections, we present the theoretical framework followed by the computational methods. We then present our results and discuss their implications for CSMs.

2. THEORY

In principle, ΔG_{rep}, ΔG_{att}, ΔG_{vdw}, and ΔG_{el} could be computed with thermodynamic integration by integrating over λ the derivatives ($\langle \partial U_{rep}(\lambda)/\partial\lambda \rangle_\lambda$, $\langle \partial U_{att}(\lambda)/\partial\lambda \rangle_\lambda$, $\langle \partial U_{vdw}(\lambda)/\partial\lambda \rangle_\lambda$, and $\langle \partial U_{el}(\lambda)/\partial\lambda \rangle_\lambda$) of λ-dependent potentials.[43,44] Plots of $\langle \partial U_{att}(\lambda)/\partial\lambda \rangle_\lambda$ and $\langle \partial U_{el}(\lambda)/\partial\lambda \rangle_\lambda$ for one of the complexes in the present study (1BRS) are shown in the Supporting Information. If these curves represented purely linear functions of λ, then ΔG_{att} and ΔG_{el} could be computed from linear response theory (LRT)

$$\Delta G_{att}^{lrt} = 1/2(\langle U_{att} \rangle_0 + \langle U_{att} \rangle_1 \quad (1)$$

$$\Delta G_{el}^{lrt} = 1/2(\langle U_{el} \rangle_0 + \langle U_{el} \rangle_1 \quad (2)$$

where U_{att} and U_{el} are attractive and electrostatic potential energies between the solute and solvent, respectively, as defined in the Methods, and $\langle \ldots \rangle_0$ and $\langle \ldots \rangle_1$ indicate that expectation values were computed in ensembles where λ = 0 and λ = 1. Some work has then gone farther by assuming that ΔG_{att} can be computed by averaging U_{att} over an approximate solvent distribution.[29,30,33] Natural choices for such a solvent distribution would be either the initial or final solvent configurations leading to single-step perturbation (SSP) estimates

$$\Delta G_{att}^{ssp,\,0} = \langle U_{att} \rangle_0 \quad (3)$$

$$\Delta G_{att}^{ssp,\,1} = \langle U_{att} \rangle_1 \quad (4)$$

of ΔG_{att}. Similarly, most CSMs assume that $\langle U_{el} \rangle_0 = 0$,[5,45] yielding a SSP approximation

$$\Delta G_{el}^{ssp} = 1/2 \langle U_{el} \rangle_0 \quad (5)$$

In contrast, the LRT cannot be used to compute either ΔG_{rep} or ΔG_{vdw} because $\langle \partial U_{rep}/\partial\lambda \rangle_\lambda$ and $\langle \partial U_{vdw}/\partial\lambda \rangle_\lambda$ typically contain large peaks or even poles.[24,46] Instead, CSMs typically use formulas taken from macroscopic liquid theory, such as

$$\Delta G_{vdw} = \gamma_{vdw} A + b \quad (6)$$

$$\Delta G_{\mathrm{rep}}=\gamma_{\mathrm{rep}}A+b \quad (7)$$

where A is the solvent-accessible surface area of the molecule, and γ_{vdw} and γ_{rep} are positive constants analogous to the surface tension of macroscopic liquid interfaces.[9,47–51]

Alternatively, some studies have claimed that ΔG_{rep} should increase linearly with the solvent-accessible volume (V) rather than A for sufficiently small cavities and that for larger cavities it should approach eq 7.[52–55] In the present study, all of our proteins are large enough that this model would predict that ΔG_{rep} should approximately obey eq 7. The interested reader is referred to ref 24 and its supporting information for the volume correlations.

For an equilibrium binding process between proteins A and B, given that estimates of ΔG of the components and the complex have been obtained, the desolvation energy ($\Delta\Delta G^{\mathrm{desol}} = \Delta G^{\mathrm{c}} - \Delta G^{\mathrm{a}} - \Delta G^{\mathrm{b}}$, where ΔG^{a}, ΔG^{b}, and ΔG^{c} are the ΔG's of the first component of the complex, the second component, and the complex, respectively) can be defined. In turn, $\Delta\Delta G^{\mathrm{desol}}$ can be broken into repulsive ($\Delta\Delta G^{\mathrm{desol}}_{\mathrm{rep}}$, attractive ($\Delta\Delta G^{\mathrm{desol}}_{\mathrm{att}}$, total van der Waals cavity insertion ($\Delta\Delta G^{\mathrm{desol}}_{\mathrm{vdw}}$, and electrostatic ($\Delta\Delta G^{\mathrm{desol}}_{\mathrm{el}}$ components, as described in the Methods.

Combining the definitions of $\Delta\Delta G^{\mathrm{desol}}_{\mathrm{rep}}$ and $\Delta\Delta G^{\mathrm{desol}}_{\mathrm{vdw}}$ with eqs 6 and 7, we can write

$$\Delta\Delta G^{\mathrm{desol}}_{\mathrm{vdw}}=\gamma_{\mathrm{vdw}}\Delta A \quad (8)$$

$$\Delta\Delta G^{\mathrm{desol}}_{\mathrm{rep}}=\gamma_{\mathrm{rep}}\Delta A \quad (9)$$

where $\Delta A = A^{\mathrm{c}} - A^{\mathrm{b}} - A^{\mathrm{a}}$, where A^{c}, A^{a}, and A^{b} are the areas of the complex and its two components.

Given $\Delta\Delta G^{\mathrm{desol}}$, estimates of $\Delta\Delta G$ can be obtained by adding the binding free energy (ΔG^{vac}) of the complex in vacuum. In turn, $\Delta\Delta G$ can be broken into repulsive ($\Delta\Delta G_{\mathrm{rep}}$), attractive ($\Delta\Delta G_{\mathrm{att}}$), total van der Waals cavity insertion ($\Delta\Delta G_{\mathrm{vdw}}$), and electrostatic ($\Delta\Delta G_{\mathrm{el}}$) components, as described in the Methods.

Many of these assumptions underlying CSMs can be called into question for biomolecules with rough surfaces and charge distributions varying over lengths comparable to the size of a water molecule. We recently found that none of ΔG_{rep}, ΔG_{att}, or ΔG_{vdw} are simple functions of A for alkanes and short peptides, indicating that eqs 6 and 7 are not strictly valid, and that eqs 3 and 4 were not consistent with ESMs.[24,34,46,56,57] Some other studies have examined the validity of eqs 2 and 5.[5,45] Also, $\Delta\Delta G^{\mathrm{desol}}_{\mathrm{att}}$ and $\Delta\Delta G^{\mathrm{desol}}_{\mathrm{el}}$ are typically positive because they account for the loss of favorable solute–solvent interactions upon binding. In contrast,

$\Delta G_{\mathrm{att}}^{\mathrm{vac}}$ and $\Delta G_{\mathrm{el}}^{\mathrm{vac}}$ are typically negative because $U_{\mathrm{att}}^{\mathrm{ij}}<0$ and because binding partners usually have complementary charges at the binding interface. Therefore, $\Delta\Delta G_{\mathrm{att}}$ and $\Delta\Delta G_{\mathrm{el}}$ can be much smaller than the ΔG_{att} and ΔG_{el} of the complex and its components. Because of this cancellation of free energies, theories that generate estimates of ΔG that are correlated with experimental data are not guaranteed to produce similarly accurate estimates of $\Delta\Delta G$.[58]

In the present study, we test the validity of eqs 1–9 for 9 protein–protein complexes[35–42,59] and also whether the PBE[1] produces estimates of ΔG_{el}, $\Delta\Delta G_{\mathrm{el}}^{\mathrm{desol}}$, and $\Delta\Delta G_{\mathrm{el}}$ consistent with those obtained from ESMs for these complexes.

3. METHODS

All MD simulations were run with a modified version of NAMD 2.9.[60] SHAKE was used to constrain the hydrogens. All simulations used the TIP3P water model[61] modified for use with the CHARMM force field,[62] a constant temperature of 300 K, a constant pressure of 1 atm, periodic boundary conditions, particle mesh Ewald for the electrostatics, and a 2 fs time step. In all simulations, all protein atoms were fixed. All *A, V,* and their derivatives with respect to the atomic coordinates were computed with the DAlphaBall program.[63] The solvent-accessible surface was used[64] with the van der Waals radii taken from the CHARMM36 force field.[62,65,66] A probe radius of 1.7682 Å was used rather than the normal choice of 1.4 Å because it corresponds to the vdW radius of the oxygen atom in the water model, which for a chargeless protein is the only interaction with solvent via the Lennard-Jones force. This choice has been used in previous works.[24,34]

All PBE calculations were run with the Adaptive Poisson– Boltzmann Solver (APBS)[67] with a temperature of 300 K, an interior dielectric constant of 1 (because the CHARMM36 force field is not polarizable), an exterior dielectric constant of 96.7 (to match the dielectric constant of TIP3P water[68]), the solvent-excluded surface defined with a probe radius of 1.4 Å, no salt, autofocusing with a fine grid 20 Å larger than the molecule in each dimension, a coarse grid 1.7 times the size of the molecule in each dimension, and a fine-grid spacing of either 0.5 or 0.55 Å (to check convergence with respect to grid spacing).

3.1. Structure Preparation

The coordinates of the 9 protein–protein complexes[35–42,59] were taken from the RCSB protein databank[69] with no minimization or equilibration of these initial structures. These atomic coordinates remained fixed through all of the remaining calculations. All crystal waters and nonprotein atoms were removed. The chains from these structure files used in each calculation are shown in Table 1.

3.2. Potential and Free Energy Components

We defined ΔG_{vdw} to be the free energy required to move from an ensemble where the solute and solvent did not interact to one where the interaction potential between an atom *i* in the solute and an atom *j* in the solvent was given by the Lennard-Jones potential

$$U_{\mathrm{vdw}}^{\mathrm{ij}}=\varepsilon_{\mathrm{ij}}\left[\left(\frac{r_{\mathrm{ij}}^{\min}}{r_{\mathrm{ij}}}\right)^{12}-2\left(\frac{r_{\mathrm{ij}}^{\min}}{r_{\mathrm{ij}}}\right)^{6}\right] \quad (10)$$

where ε_{ij} is the well depth, and $r_{\mathrm{ij}}^{\min}$ is the location of the minimum of $U_{\mathrm{vdw}}^{\mathrm{ij}}$. Our definitions of ΔG_{rep} and ΔG_{att} followed the Weeks–Chandler–Andersen breakdown,[25,26] where ΔG_{rep} was the free energy required to move from an ensemble where the solute and solvent did not interact to one where the interaction potential between an atom *i* in the solute and an atom *j* in the solvent was given by

$$U_{\mathrm{rep}}^{\mathrm{ij}}=\begin{cases}\varepsilon_{\mathrm{ij}}\left[\left(\frac{r_{\mathrm{ij}}^{\min}}{r_{\mathrm{ij}}}\right)^{12}-2\left(\frac{r_{\mathrm{ij}}^{\min}}{r_{\mathrm{ij}}}\right)^{6}+1\right] & \text{if } r_{\mathrm{ij}}<r_{\mathrm{ij}}^{\min}\\ 0 & \text{otherwise}\end{cases} \quad (11)$$

Next, we defined ΔG_{att} to be the free energy gained by moving from an ensemble where the solute–solvent potential was given by $U_{\mathrm{rep}}^{\mathrm{ij}}$ to one where it was given by $U_{\mathrm{vdw}}^{\mathrm{ij}}$. This process could also be described as turning on the attractive part of $U_{\mathrm{vdw}}^{\mathrm{ij}}$

$$U_{\mathrm{att}}^{\mathrm{ij}}=\begin{cases}-\varepsilon_{\mathrm{ij}} & \text{if } r_{\mathrm{ij}}<r_{\mathrm{ij}}^{\min}\\ \varepsilon_{\mathrm{ij}}\left[\left(\frac{r_{\mathrm{ij}}^{\min}}{r_{\mathrm{ij}}}\right)^{12}-2\left(\frac{r_{\mathrm{ij}}^{\min}}{r_{\mathrm{ij}}}\right)^{6}\right] & \text{otherwise}\end{cases} \quad (12)$$

Finally, we defined ΔG_{el} to be the free energy gained by moving from an ensemble where the solute–solvent potential was given by $U_{\mathrm{vdw}}^{\mathrm{ij}}$ to one where it was given by

$$U^{\mathrm{ij}}=U_{\mathrm{vdw}}^{\mathrm{ij}}+U_{\mathrm{el}}^{\mathrm{ij}}$$

$$=U_{\mathrm{vdw}}^{\mathrm{ij}}+q_iq_j/4\pi\varepsilon_0r_{\mathrm{ij}} \quad (13)$$

where q_i and q_j are the charges on atoms *i* and *j*, respectively, and ε_0 is the permittivity of free space.

We then defined $U_{\mathrm{rep}}=\sum_{\mathrm{ij}}U_{\mathrm{rep}}^{\mathrm{ij}}$, $U_{\mathrm{att}}=\sum_{\mathrm{ij}}U_{\mathrm{att}}^{\mathrm{ij}}$, $U_{\mathrm{vdw}}=\sum_{\mathrm{ij}}U_{\mathrm{vdw}}^{\mathrm{ij}}$, and $U_{\mathrm{el}}=\sum_{\mathrm{ij}}U_{\mathrm{el}}^{\mathrm{ij}}$, where the summations were taken over all solute–solvent atom pairs.

Similarly to the definition of $\Delta\Delta G^{\mathrm{desol}}$ in the Introduction, $\Delta\Delta G_{\mathrm{rep}}^{\mathrm{desol}}$, $\Delta\Delta G_{\mathrm{att}}^{\mathrm{desol}}$, and $\Delta\Delta G_{\mathrm{el}}^{\mathrm{desol}}$ can be defined by

$$\Delta\Delta G_{\mathrm{rep}}^{\mathrm{desol}}=\Delta G_{\mathrm{rep}}^{\mathrm{c}}-\Delta G_{\mathrm{rep}}^{\mathrm{b}}-\Delta G_{\mathrm{rep}}^{\mathrm{a}} \quad (14)$$

$$\Delta\Delta G_{\mathrm{att}}^{\mathrm{desol}}=\Delta G_{\mathrm{att}}^{\mathrm{c}}-\Delta G_{\mathrm{att}}^{\mathrm{b}}-\Delta G_{\mathrm{att}}^{\mathrm{a}} \quad (15)$$

$$\Delta\Delta G_{\mathrm{vdw}}^{\mathrm{desol}}=\Delta G_{\mathrm{vdw}}^{\mathrm{c}}-\Delta G_{\mathrm{vdw}}^{\mathrm{b}}-\Delta G_{\mathrm{vdw}}^{\mathrm{a}},\mathrm{and} \quad (16)$$

$$\Delta\Delta G_{\mathrm{el}}^{\mathrm{desol}}=\Delta G_{\mathrm{el}}^{\mathrm{c}}-\Delta G_{\mathrm{el}}^{\mathrm{b}}-\Delta G_{\mathrm{el}}^{\mathrm{a}} \quad (17)$$

where $\Delta G_{\mathrm{rep}}^{\mathrm{c}}, \Delta G_{\mathrm{rep}}^{\mathrm{a}}$, and$\Delta G_{\mathrm{rep}}^{\mathrm{b}}$ are the ΔG_{rep} of the complex and its first and second components, $\Delta G_{\mathrm{att}}^{\mathrm{c}}, \Delta G_{\mathrm{att}}^{\mathrm{a}}$, and $\Delta G_{\mathrm{att}}^{\mathrm{b}}$ are the ΔG_{att} of the complex and its first and second components, $\Delta G_{\mathrm{vdw}}^{\mathrm{c}}, \Delta G_{\mathrm{vdw}}^{\mathrm{a}}$, and $\Delta G_{\mathrm{vdw}}^{\mathrm{b}}$ are the ΔG_{vdw} of the complex and its first and second components, and $\Delta G_{\mathrm{el}}^{\mathrm{c}}, \Delta G_{\mathrm{el}}^{\mathrm{a}}$, and $\Delta G_{\mathrm{el}}^{\mathrm{b}}$ are the ΔG_{el} of the complex and its first and second components, respectively.

$\Delta\Delta G$ and its components were defined as follows

$$\Delta\Delta G=\Delta\Delta G^{\mathrm{desol}}+\Delta G^{\mathrm{vac}} \quad (18)$$

$$\Delta\Delta G_{\mathrm{rep}}=\Delta\Delta G_{\mathrm{rep}}^{\mathrm{desol}}+\Delta G_{\mathrm{rep}}^{\mathrm{vac}} \quad (19)$$

$$\Delta\Delta G_{\mathrm{att}}=\Delta\Delta G_{\mathrm{att}}^{\mathrm{desol}}+\Delta G_{\mathrm{att}}^{\mathrm{vac}} \quad (20)$$

$$\Delta\Delta G_{\mathrm{vdw}}=\Delta\Delta G_{\mathrm{vdw}}^{\mathrm{desol}}+\Delta G_{\mathrm{vdw}}^{\mathrm{vac}} \quad (21)$$

$$\Delta\Delta G_{\mathrm{el}}=\Delta\Delta G_{\mathrm{el}}^{\mathrm{desol}}+\Delta G_{\mathrm{el}}^{\mathrm{vac}} \quad (22)$$

where $\Delta G_{\mathrm{vac}}=\sum_{\mathrm{ij}} U^{\mathrm{ij}}$, $\Delta G_{\mathrm{rep}}^{\mathrm{vac}}=\sum_{\mathrm{ij}} U_{\mathrm{rep}}^{\mathrm{ij}}$, $\Delta G_{\mathrm{att}}^{\mathrm{vac}}=\sum_{\mathrm{ij}} U_{\mathrm{att}}^{\mathrm{ij}}$, $\Delta G_{\mathrm{vdw}}^{\mathrm{vac}}=\sum_{\mathrm{ij}} U_{\mathrm{vdw}}^{\mathrm{ij}}$ and $\Delta G_{\mathrm{el}}^{\mathrm{vac}}=\sum_{\mathrm{ij}} U_{el}^{ij}$, where these summations were taken over atom pairs with one atom in each component of the complex.

3.3. Free Energy Calculations

For the 1BRS complex, we computed ΔG_{att} and ΔG_{el} by backward and forward free energy perturbation (FEP) and thermodynamic integration (TI).[43,44] To do so, we defined the λ-dependent potentials

$$U_{\mathrm{att}}^{\mathrm{ij}}(\lambda=U_{\mathrm{rep}}^{\mathrm{ij}}+\lambda U_{\mathrm{att}}^{\mathrm{ij}} \quad (23)$$

$$U_{\mathrm{el}}^{\mathrm{ij}}(\lambda=U_{\mathrm{vdw}}^{\mathrm{ij}}+\lambda U_{\mathrm{el}}^{\mathrm{ij}} \quad (24)$$

To compute ΔG_{att} for each component and the complex, 11 1 ns simulations were run at λ values ranging from 0 to 1 in intervals of 0.1. To compute ΔG_{el} for the components of the complex, 21 1 ns simulations were run at λ values ranging from 0 to 1 in intervals of 0.05. To compute ΔG_{el} for the complex 1 ns simulations were run at the following λ values: 0, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.275, 0.3, 0.325, 0.35, 0.375, 0.4, 0.425, 0.45, 0.475, 0.5, 0.525, 0.55, 0.575, 0.6, 0.625, 0.65, 0.675, 0.7, 0.725, 0.75, 0.775, 0.8, 0.825, 0.85, 0.8625, 0.875, 0.8875, 0.9, 0.9125, 0.925, 0.9375, 0.95, 0.9625, 0.975, 0.9875, and 1.

For the other 8 complexes, $\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$ and $\Delta G_{\mathrm{el}}^{\mathrm{lrt}}$ were computed using eqs 1 and 2 for both the components and the complexes by running 1 ns simulations where the interaction potential between the solute and solvent was $U_{\mathrm{rep}}^{\mathrm{ij}}, U_{\mathrm{att}}^{\mathrm{ij}}$, and U^{ij}. Estimates of $\Delta G_{\mathrm{att}}^{\mathrm{ssp},\,0}$, $\Delta G_{\mathrm{att}}^{\mathrm{ssp},\,1}$, and $\Delta G_{\mathrm{el}}^{\mathrm{ssp}}$ were extracted from these same simulations.

Initial structures for all of these simulations were obtained by immersing the structure in a water box that was 20 Å longer in each dimension than the molecule and adding either Na^+ or Cl^- ions to neutralize the system. The structures were then minimized for 500 steps with the solute–solvent potential set as U^{ij}. Copies of these minimized structure then underwent equilibration at each value of λ. The temperature of each of these systems was increased from 0 to 300 K in units of 25 K with 1000 steps of simulation time at each temperature.

3.4. Free Energy Derivatives

Computing ΔG_{rep} and ΔG_{vdw} for systems as large as these protein–protein complexes would require a great deal of computational time due to the drying of the molecular interior, so we instead adopted a procedure used in our previous publications,[24,34] computing the derivatives of ΔG_{rep} and ΔG_{vdw} with respect to the coordinates (x_i) of the centers of the fixed protein atoms

$$\partial \Delta G_{\mathrm{rep}}/\partial x_i = \langle \partial U_{\mathrm{rep}}/\partial x_i \rangle_{\mathrm{rep}} \quad (25)$$

where this average was taken in the ensemble defined by U_{rep}, and

$$\partial \Delta G_{\mathrm{vdw}}/\partial x_i = \langle \partial U_{\mathrm{att}}(1)/\partial x_i \rangle_1 \quad (26)$$

where these derivatives were computed from the same simulations used to compute the LRT and SSP estimates of ΔG_{att}. These derivatives were computed for each Cartesian coordinate of each protein atom with the rest of the coordinates held fixed.

If ΔG_{rep} and ΔG_{vdw} were linear functions of either A or V, as expected from eqs 6 and 7, then the slopes of plots of these quantities versus A would be γ_{rep} and γ_{vdw}, respectively. Similarly, if $\Delta\Delta G_{\mathrm{rep}}^{\mathrm{desol}}$ and $\Delta\Delta G_{\mathrm{vdw}}^{\mathrm{desol}}$ were linear in ΔA, then plots of

$$\partial \Delta\Delta G_{\mathrm{rep}}^{\mathrm{desol}}/\partial x_i = \partial \Delta G_{\mathrm{rep}}^{\mathrm{c}}/\partial x_i - \partial \Delta G_{\mathrm{rep}}^{\mathrm{b}}/\partial x_i - \partial \Delta G_{\mathrm{rep}}^{\mathrm{a}}/\partial x_i \quad (27)$$

$$\partial \Delta\Delta G_{\mathrm{vdw}}^{\mathrm{desol}}/\partial x_i = \partial \Delta G_{\mathrm{vdw}}^{\mathrm{c}}/\partial x_i - \partial \Delta G_{\mathrm{vdw}}^{\mathrm{b}}/\partial x_i - \partial \Delta G_{\mathrm{vdw}}^{\mathrm{a}}/\partial x_i \quad (28)$$

versus ΔA would have slopes of γ_{rep} and γ_{vdw}, respectively. Estimates of γ_{rep} and γ_{vdw} obtained from such plots are shown in Table 1.

4. RESULTS

Table 1 gives estimates of γ_{rep} and γ_{vdw} obtained by fitting least-squares lines to plots of $\partial \Delta G_{\mathrm{rep}}/\partial x_i$ and $\partial \Delta G_{\mathrm{vdw}}/\partial x_i$ versus $\partial A/\partial x_i$ and $\partial \Delta\Delta G_{\mathrm{rep}}^{\mathrm{desol}}/\partial x_i$ and $\partial \Delta\Delta G_{\mathrm{vdw}}^{\mathrm{desol}}/\partial x_i$ versus $\partial \Delta A/\partial x_i$. The Supporting Information contains these plots and analyses of the precision of the underlying quantities. If eqs 6–9 were valid, then the squares of the Pearson's correlation coefficients (R^2) would be close to 1, and the estimates of γ_{rep} and γ_{vdw} would be consistent from molecule to molecule as well as between estimates obtained from solvation free energies and those obtained from binding free energies. Instead, plots of $\partial \Delta G_{\mathrm{rep}}/\partial x_i$ and $\partial \Delta\Delta G_{\mathrm{rep}}/\partial x_i$ versus $\partial A/\partial x_i$ and $\partial \Delta A/\partial x_i$ revealed very weak correlations between these quantities. Figure 1 shows a typical example of a plot of $\partial \Delta G_{\mathrm{rep}}/\partial x_i$ versus $\partial A/\partial x_i$ for one protein molecule. These findings are in agreement with our previous findings that the geometry and chemical environment of an atom's surroundings affects how changes in that atom's position change ΔG_{rep}.[24,34]

The estimates of γ_{rep} obtained from plots of $\partial \Delta G_{\mathrm{rep}}/\partial x_i$ versus $\partial A/\partial x_i$ not only differed between molecules but decreased with increasing A (Table 1 and Figure 2). These estimates of γ_{rep} ranged from 0.012 to 0.032 kcal mol^{-1} Å^{-2}, and these values were generally smaller than those we found previously for decaalanine, decaglycine, and alkanes, which generally

had smaller A than the proteins examined here.[24,34] This observation appears to contradict the notion that a critical A exists for which molecules with larger A have ΔG_{rep} that are linear in A and below which ΔG_{rep} is linear in V. This finding may indicate that the vicinity of the biomolecule becomes progressively "dryer" as the size of the cavity increases. Eq 7 is not consistent with these findings.

Table 2 shows estimates of ΔG_{att} and ΔG_{el} for one of the complexes (1BRS) and its components and $\Delta\Delta G_{att}^{desol}$, $\Delta\Delta G_{el}^{desol}$, $\Delta\Delta G_{att}$, and $\Delta\Delta G_{el}$ for this complex given by FEP, TI, LRT, and SSP. These data show that the LRT gives good estimates of ΔG_{att}, $\Delta\Delta G_{att}^{desol}$, and $\Delta\Delta G_{att}$ for this complex, confirming the validity of eq 1 for this complex. However, SSP gave estimates of ΔG_{att}, $\Delta\Delta G_{att}^{desol}$, and $\Delta\Delta G_{att}$ that differed significantly from those given by FEP, indicating that eqs 3 and 4 were not valid for this system.

The data in Table 2 show that the LRT yielded estimates of ΔG_{el} that were more than 100 kcal/mol different from those obtained with FEP, indicating that eq 2 is poor for this system. FEP estimates of $\Delta\Delta G_{el}^{desol}$ and $\Delta\Delta G_{el}$ also differed significantly from those given by LRT.

Table 3 shows the R^2, slopes, and y-intercepts of least-squares lines comparing various computed quantities for the 9 protein–protein complexes. All of these plots and analyses of the precisions of the resulting quantities are in the Supporting Information. These data show that the estimates of ΔG_{att} obtained from the LRT were highly correlated with A, but its estimates of $\Delta\Delta G_{att}^{desol}$ and $\Delta\Delta G_{att}$ were not strongly correlated with ΔA, demonstrating that two theories can yield similar results for ΔG and very different results for $\Delta\Delta G$.

Additionally, the data in Table 3 show that both eqs 3 and 4 produced estimates of ΔG_{att} that were highly correlated with those obtained from the LRT ($R^2 > 0.97$), but eq 3 systematically underestimated ΔG_{att} whereas eq 4 overestimated it. Table 3 also shows that the estimates of $\Delta\Delta G_{att}^{desol}$ obtained from eq 3 were not highly correlated with those obtained from LRT. Conversely, the predictions of $\Delta\Delta G_{att}^{desol}$ and $\Delta\Delta G_{att}$ obtained from eq 4 were reasonably well-correlated with those obtained from LRT, but these estimates were significantly larger than those obtained from LRT, indicating that combining these energy terms with the other components of $\Delta\Delta G$ could lead to unexpected results.

The data in Table 3 also show that the computations of ΔG_{el}, $\Delta\Delta G_{el}^{desol}$, and $\Delta\Delta G_{el}$ given by SSP are very highly correlated ($R^2 > 0.999$) with those obtained from LRT. Therefore, if the LRT is sufficient to predict these quantities, then the SSP would be sufficient as well. Unfortunately, the results in Table 2 indicate that the LRT may not be sufficient for computing these quantities.

Table 3 also shows that the PBE yielded predictions (ΔG_{el}^{APBS} and $\Delta\Delta G_{el}^{APBS,\ desol}$) of ΔG_{el} and $\Delta\Delta G_{el}^{desol}$ that were highly correlated with those obtained from the LRT, but its estimates ($\Delta\Delta G_{el}^{APBS}$ of $\Delta\Delta G_{el}$ were not correlated with those obtained from LRT (R^2 = 0.43). Once again, this finding shows that strong correlations between the predictions of ΔG

given by two theories do not guarantee that the theories will produce highly correlated estimates of $\Delta\Delta G$.

5. CONCLUSIONS

The calculations on protein–protein complexes presented in this paper allow us to draw some conclusions about the assumptions underlying CSMs. As in our previous work, eqs 6 and 7 were not consistent with the results. Neither ΔG_{rep} nor ΔG_{vdw} was proportional to A.

Eqs 8 and 9 were inconsistent with our calculations. Neither $\Delta\Delta G_{\text{rep}}^{\text{desol}}$ nor $\Delta\Delta G_{\text{vdw}}^{\text{desol}}$ was proportional to ΔA. Apparently, the idea that either ΔG_{vdw} or ΔG_{rep} is linear in A and that the resulting ΔG_{vdw} can be combined with the ΔG_{el} obtained from CSMs to obtain good estimates of ΔG is not valid for protein–protein complexes.

Additionally, although these results show that none of ΔG_{rep}, ΔG_{vdw}, $\Delta\Delta G_{\text{rep}}^{\text{desol}}$, or $\Delta\Delta G_{\text{vdw}}^{\text{desol}}$ are truly linear in A or ΔA, if we extract estimates of apparent γ_{rep} from the derivative plots described above, we find our estimates of γ_{rep} decrease with A. This finding contradicts the hypothesis that there is a well-defined A above which ΔG_{rep} is linear in A and below which ΔG_{rep} is linear in V for realistic biomolecular solutes.

We found that LRT was sufficient for estimating $\Delta\Delta G_{\text{att}}^{\text{desol}}$, and $\Delta\Delta G_{\text{att}}$. However, although the SSP using either eq 3 or 4 yielded estimates of ΔG_{att} that were correlated with those given by the LRT, and the predictions of $\Delta\Delta G_{\text{att}}^{\text{desol}}$ and $\Delta\Delta G_{\text{att}}$ given by eq 4 were also correlated to those given by LRT, the magnitudes of these quantities were significantly different from those given by LRT. Whether such estimates would be sufficient for in turn estimating $\Delta\Delta G$ is therefore much less clear. Additionally, whereas ΔG_{att} was correlated with A, neither $\Delta\Delta G_{\text{att}}^{\text{desol}}$ nor $\Delta\Delta G_{\text{att}}$ was correlated with ΔA, highlighting the observation that two theories could produce correlated estimates of solvation free energies while producing uncorrelated estimates of binding free energies.

For the one protein in the data set where ΔG_{el} was obtained with FEP, this estimate differed by more than 100 kcal/mol from those given by SSP and LRT, and the corresponding estimate of $\Delta\Delta G_{\text{el}}$ obtained from FEP differed by more than 10 kcal/mol from those obtained from the LRT and SSP, implying that the LRT and SSP approximations may be problematic in some situations. However, for all of the complexes in the data set, the SSP gave estimates of ΔG_{el}, $\Delta\Delta G_{\text{el}}^{\text{desol}}$, and $\Delta\Delta G_{\text{el}}$ that were highly correlated with those given by the LRT. Therefore, if the LRT is a reasonable approximation for a system, the SSP is probably reasonable as well. These calculations will have to be repeated for other systems to determine how general are these conclusions.

In summary, many of the assumptions underlying common CSMs are brought into question by this work. The predictions yielded by often-used hydrophobic models disagreed with the results from FEP. LRT did not yield estimates of electrostatic energies that were consistent with those obtained with FEP. Furthermore, although the PBE yielded estimates of ΔG_{el} that were highly correlated with those obtained from FEP, its estimates of $\Delta\Delta G_{\text{el}}$ were not in good agreement with those obtained by FEP. In combination with our observation that ΔG_{att}

was correlated with A but that $\Delta\Delta G_{att}$ was not, we can see that we cannot conclude that a theory will give good estimates of $\Delta\Delta G$ simply because its estimates of ΔG have some agreement with experimental data. Attempting to improve the estimates of $\Delta\Delta G$ given by CSMs by comparing their predictions of ΔG to experimental measurements may therefore not be productive.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Robert A. Welch Foundation (H-0037), the National Science Foundation (CHE-1152876), and the National Institutes of Health (GM-037657) are thanked for partial support of this work. This research was performed in part using Xsede resources provided by the National Science Foundation.

REFERENCES

1. Grochowski P, Trylska J. Continuum molecular electrostatics, salt effects, and counterion binding-A review of the Poisson-Boltzmann theory and its modifications. Biopolymers. 2008; 89:93–113. [PubMed: 17969016]
2. Bashford D, Case DA. Generalized Born models of macromolecular solvation effects. Annu. Rev. Phys. Chem. 2000; 51:129–152. [PubMed: 11031278]
3. Lounnas V, Pettitt BM, Phillips GN Jr. A global model of the protein-solvent interface. Biophys. J. 1994; 66:601–614. [PubMed: 8011893]
4. Makarov V, Pettitt BM, Feig M. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment. Acc. Chem. Res. 2002; 35:376–384. [PubMed: 12069622]
5. Lin B, Pettitt BM. Electrostatic solvation free energy of amino acid side chain analogs: Implications for the validity of electrostatic linear response in water. J. Comput. Chem. 2011; 32:878–85. [PubMed: 20941733]
6. Hirata, F., editor. Understanding chemical reactivity: Molecular theory of solvation. Norwell, MA, USA: Kluwer Academic Publishers; 2003.
7. Truchon J-F, Pettitt BM, Labute P. A cavity corrected 3D-RISM functional for accurate solvation free energies. J. Chem. Theory Comput. 2014; 10:934–941. [PubMed: 24634616]
8. Babu CS, Tembe BL. The role of solvent models in stabilizing nonclassical ions. Proc. - Indian Acad. Sci., Chem. Sci. 1987; 98:235–240.
9. Baldwin R. Energetics of Protein Folding. J. Mol. Biol. 2007; 371:283–301. [PubMed: 17582437]
10. Mohan V, Davis ME, McCammon JA, Pettitt BM. Continuum model calculations of solvation free energies: Accurate evaluation of electrostatic contributions. J. Phys. Chem. 1992; 96:6428–6431.
11. Simonson T, Brünger AT. Solvation free energies estimated from macroscopic continuum theory: An accuracy assessment. J. Phys. Chem. 1994; 98:4683–4694.
12. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. J. Med. Chem. 2008; 51:769–779. [PubMed: 18215013]
13. Guthrie JP. A blind challenge for computational solvation free energies: Introduction and overview. J. Phys. Chem. B. 2009; 113:4501–4507. [PubMed: 19338360]
14. Mobley DL, Wymer KL, Lim NM, Guthrie JP. Blind prediction of solvation free energies from the SAMPL4 challenge. J. Comput.-Aided Mol. Des. 2014; 28:135–150. [PubMed: 24615156]
15. Wang J, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc. 2001; 123:5221–5230. [PubMed: 11457384]

16. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J. Chem. Inf. Model. 2011; 51:69–82. [PubMed: 21117705]
17. Adler M, Beroza P. Improved ligand binding energies derived from molecular dynamics: Replicate sampling enhances the search of conformational space. J. Chem. Inf. Model. 2013; 53:2065–2072. [PubMed: 23845109]
18. Harris RC, Mackoy T, Fenley MO. A stochastic solver of the generalized Born model. Mol. Based Math. Biol. 2013; 1:63–74.
19. Li M, Petukh M, Alexov E, Panchenko AR. Predicting the impact of missense mutations on protein-protein binding affinity. J. Chem. Theory Comput. 2014; 10:1770–1780. [PubMed: 24803870]
20. Muddana HS, Fenley AT, Mobley DL, Gilson MK. The SAMPL4 host-guest blind prediction challenge: An overview. J. Comput.-Aided Mol. Des. 2014; 28:305–317. [PubMed: 24599514]
21. Zhu Y-L, Beroza P, Artis DR. Including explicit water molecules as part of the protein structure in MM/PBSA calculations. J. Chem. Inf. Model. 2014; 54:462–469. [PubMed: 24432790]
22. Sharp KA, Honig B. Electrostatic interactions in macromolecules: Theory and applications. Annu. Rev. Biophys. Biophys. Chem. 1990; 19:301–332. [PubMed: 2194479]
23. Cramer CJ, Truhlar DG. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. Chem. Rev. 1999; 99:2161–2200. [PubMed: 11849023]
24. Harris RC, Pettitt BM. Effects of geometry and chemistry on hydrophobic solvation. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:14681–14686. [PubMed: 25258413]
25. Weeks JD, Chandler D, Andersen HC. Role of repulsive forces in determining the equilibrium structure of simple liquids. J. Chem. Phys. 1971; 54:5237–5247.
26. Chandler D, Weeks JD, Andersen HC. Van der Waals picture of liquids, solids, and phase transformations. Science. 1983; 220:787–794. [PubMed: 17834156]
27. Ashbaugh HS, Kaler EW, Paulaitis ME. A “universal” surface area correlation for molecular hydrophobic phenomena. J. Am. Chem. Soc. 1999; 121:9243–9244.
28. Gallicchio E, Kubo MM, Levy RM. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. J. Phys. Chem. B. 2000; 104:6271–6285.
29. Gallicchio E, Zhang LY, Levy RM. The SGB/NP hydration free energy model based on the surface generalized Born solvent reaction field and novel nonpolar hydration free energy estimators. J. Comput. Chem. 2002; 23:517–529. [PubMed: 11948578]
30. Zacharias M. Continuum solvent modeling of nonpolar solvation: Improvement by separating surface area dependent cavity and dispersion contributions. J. Phys. Chem. A. 2003; 107:3000–3004.
31. Choudhury N, Pettitt BM. On the mechanism of hydrophobic association of nanoscopic solutes. J. Am. Chem. Soc. 2005; 127:3556–3567. [PubMed: 15755177]
32. Choudhury N, Montgomery Pettitt B. Local density profiles are coupled to solute size and attractive potential for nanoscopic hydrophobic solutes. Mol. Simul. 2005; 31:457–463.
33. Wagoner JA, Baker NA. Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:8331–8336. [PubMed: 16709675]
34. Harris RC, Drake JA, Pettitt BM. Multibody correlations in the hydrophobic solvation of glycine peptides. J. Chem. Phys. 2014; 141:22D525.
35. Frigerio F, Coda A, Pugliese L, Lionetti C, Menegatti E, Amiconi G, Schnebli HP, Ascenzi P, Bolognesi M. Crystal and molecular structure of the bovine α-chymotrypsin-eglin c complex at 2.0 Å resolution. J. Mol. Biol. 1992; 225:107–123. [PubMed: 1583684]
36. Song HK, Suh SW. Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. J. Mol. Biol. 1998; 275:347–363. [PubMed: 9466914]

37. Brownlow S, Cabral JHM, Cooper R, Flower DR, Yewdall SJ, Polikarpov I, North ACT, Sawyer L. Bovine β-lactoglobulin at 1.8 Å resolution-Still an enigmatic lipocalin. Structure. 1997; 5:481–495. [PubMed: 9115437]
38. Buckle AM, Schreiber G, Fersht AR. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. Biochemistry. 1994; 33:8878–8889. [PubMed: 8043575]
39. Kühlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein-protein interactions: The structural basis for dual recognition in endonuclease colicin-immunity protein complexes. J. Mol. Biol. 2000; 301:1163–1178. [PubMed: 10966813]
40. Würtele M, Wolf E, Pederson KJ, Buchwald G, Ahmadian MR, Barbieri JT, Wittinghofer A. How the *Pseudomonas aeruginosa* ExoS toxin downregulates Rac. Nat. Struct. Biol. 2001; 8:23–26. [PubMed: 11135665]
41. Bode W, Greyling HJ, Huber R, Otlewski J, Wilusz T. The refined 2.0 Å X-ray crystal structure of the complex formed between bovine β-trypsin and CMTI-I, a trypsin inhibitor from squash seeds (*Cucurbita maxima*) Topological similarity of the squash seed inhibitors with the carboxypeptidase A inhibitor from potatoes. FEBS Lett. 1989; 242:285–292. [PubMed: 2914611]
42. Savva R, Pearl LH. Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. Nat. Struct. Biol. 1995; 2:752–757. [PubMed: 7552746]
43. Beveridge DL, DiCapua FM. Free energy via molecular simulation: Applications to chemical and biomolecular systems. Annu. Rev. Biophys. Biophys. Chem. 1989; 18:431–492. [PubMed: 2660832]
44. Straatsma TP, McCammon JA. Computational alchemy. Annu. Rev. Phys. Chem. 1992; 43:407–435.
45. Åqvist J, Hansson A. On the validity of electrostatic linear response in polar solvents. J. Phys. Chem. 1996; 100:9512–9521.
46. Kokubo H, Harris RC, Asthagiri D, Pettitt BM. Solvation free energies of alanine peptides: The effect of flexibility. J. Phys. Chem. B. 2013; 117:16428–16435. [PubMed: 24328358]
47. Young T. An essay on the cohesion of fluids. Philos. Trans. R. Soc. 1805; 95:65–87.
48. Stillinger FH. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. J. Solution Chem. 1973; 2:141–158.
49. Pierotti RA. A scaled particle theory of aqueous and nonaqueous solutions. Chem. Rev. 1976; 76:717–726.
50. Sharp KA, Nicholls A, Fine RF, Honig B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science. 1991; 252:106–109. [PubMed: 2011744]
51. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. 1994; 98:1978–1988.
52. Lum K, Chandler D, Weeks JD. Hydrophobicity at small and large length scales. J. Phys. Chem. B. 1999; 103:4570–4577.
53. Huang DM, Chandler D. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. Proc. Natl. Acad. Sci. U.S.A. 2000; 97:8324–8327. [PubMed: 10890881]
54. Hummer G, Garde S, García AE, Pratt LR. New perspectives on hydrophobic effects. Chem. Phys. 2000; 258:349–370.
55. Rajamani S, Truskett TM, Garde S. Hydrophobic hydration from small to large lengthscales: Understanding and manipulating the crossover. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:9475–9480. [PubMed: 15972804]
56. Hu CY, Kokubo H, Lynch GC, Bolen DW, Pettitt BM. Backbone additivity in the transfer model of protein solvation. Protein Sci. 2010; 19:1011–1022. [PubMed: 20306490]
57. Kokubo H, Hu CY, Pettitt BM. Peptide conformational preferences in osmolyte solutions: Transfer free energies of decaalanine. J. Am. Chem. Soc. 2011; 133:1849–1858. [PubMed: 21250690]
58. Harris RC, Mackoy T, Fenley MO. Problems of robustness in Poisson-Boltzmann binding energies. J. Chem. Theory Comput. 2015; 11:705–712. [PubMed: 26528091]

59. Friedrich R, Fuentes-Prior P, Ong E, Coombs G, Hunter M, Oehler R, Pierson D, Gonzalez R, Huber R, Bode W, Madison EL. Catalytic domain structures of MT-SP1/matriptase, a matrix-degrading transmembrane serine proteinase. J. Biol. Chem. 2002; 277:2160–2168. [PubMed: 11696548]

60. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J. Comput. Chem. 2005; 26:1781–1802. [PubMed: 16222654]

61. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983; 79:926–935.

62. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiókiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 1998; 102:3586–3616. [PubMed: 24889800]

63. Edelsbrunner H, Koehl P. The weighted-volume derivative of a space-filling diagram. Proc. Natl. Acad. Sci. U.S.A. 2003; 100:2203–2208. [PubMed: 12601153]

64. Lee B, Richards FM. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 1971; 55:379–400. [PubMed: 5551392]

65. MacKerell AD Jr, Feig M, Brooks CL III. Improved treatment of the protein backbone in empirical force fields. J. Am. Chem. Soc. 2004; 126:698–699. [PubMed: 14733527]

66. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. J. Chem. Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]

67. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. 2001; 98:10037–10041. [PubMed: 11517324]

68. Höchtl P, Boresch S, Bitomsky W, Steinhauser O. Rationalization of the dielectric properties of common three-site water models in terms of their force field parameters. J. Chem. Phys. 1998; 109:4927–4937.

69. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]

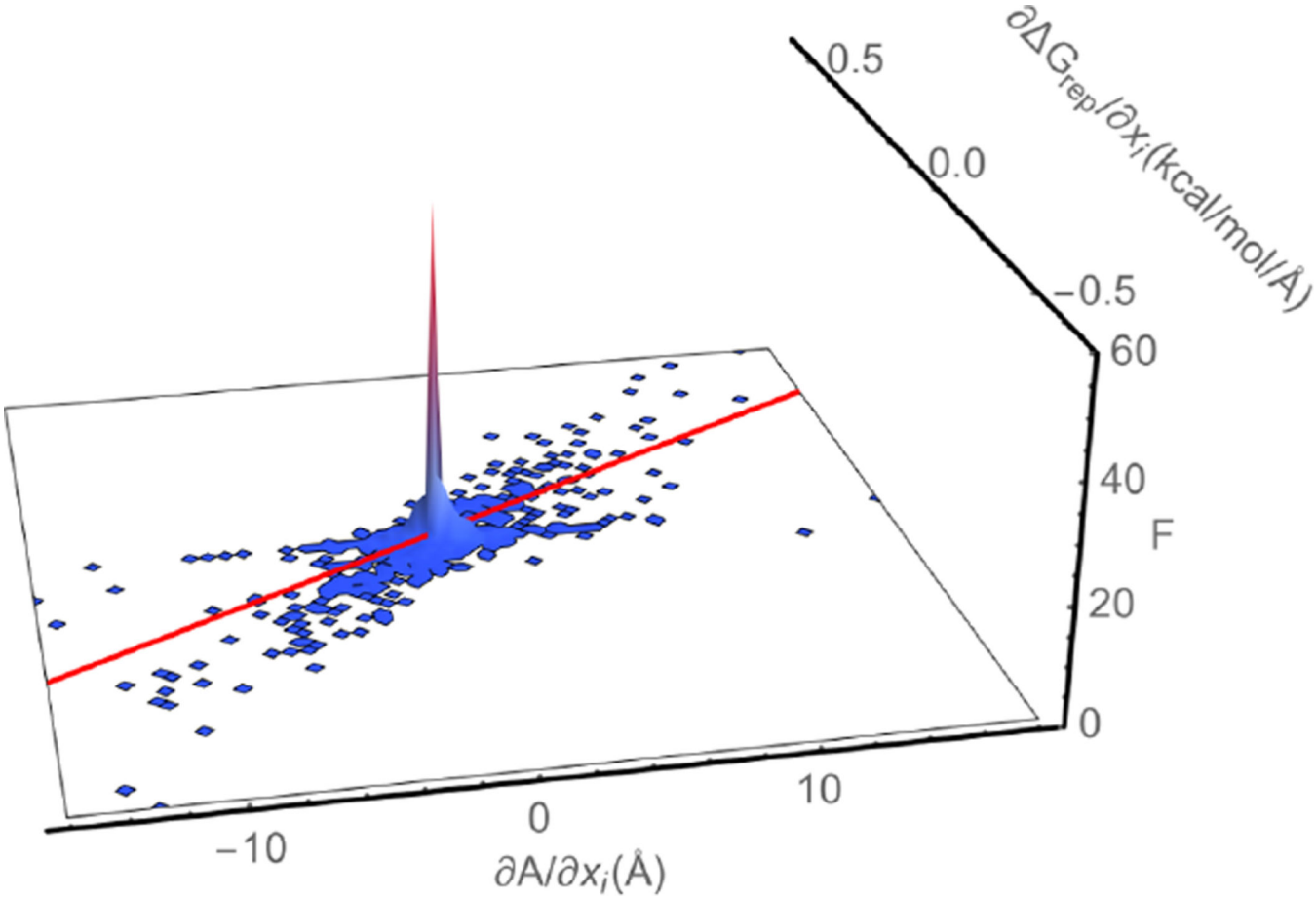


Figure 1.

A plot of the probability density (F) of a given combination of the derivative ($\partial\Delta G_{rep}/\partial x_i$) of the repulsive component (ΔG_{rep}) of the free energy required to insert an uncharged molecule into solution with respect to the atomic coordinates (x_i) and the derivative ($\partial A/\partial x_i$) of the solvent-accessible surface area (A) with respect to x_i for the second chain of the 1UDI complex. The red line is a least-squares line drawn through all of the points. Its slope provides an estimate of γ_{rep} for this protein, and it is equal to 0.022 kcal mol^{-1} Å^{-2}. Those atoms for which both $\partial\Delta G_{rep}/\partial x_i$ and $\partial A/\partial x_i$ were equal to 0 were not included when computing F. If they had been included, the peak in F would have been significantly larger.

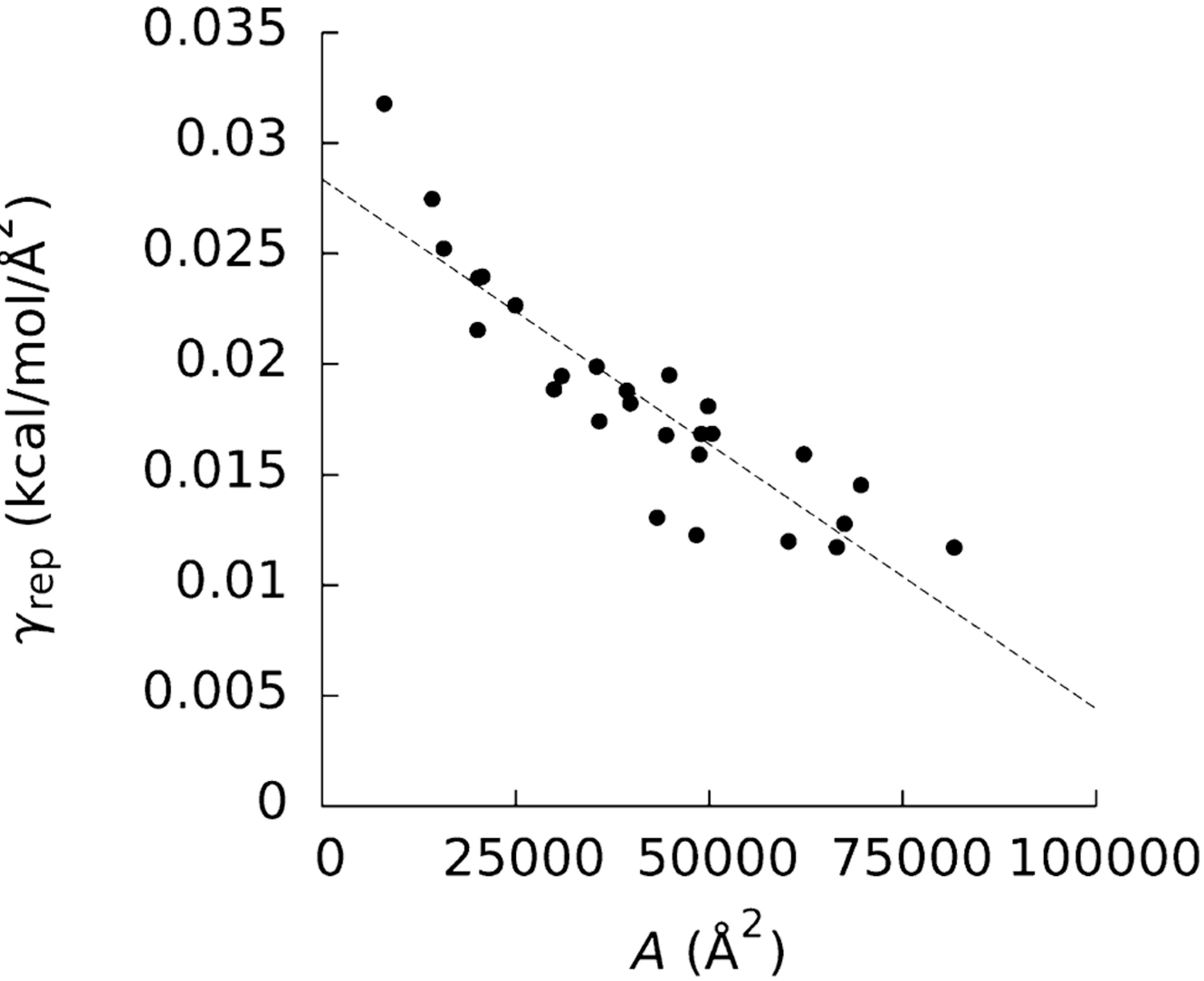


Figure 2.

Slopes (γ_{rep}) of the least-squares fit of the derivatives ($\partial G_{rep}/\partial x_i$) of the repulsive ($\Delta G_{rep}$) component of the cavity-insertion energy (ΔG_{vdw}) with respect to the coordinates (x_i) of the atomic centers versus derivatives ($\partial A/\partial x_i$) of the solvent-accessible surface areas (A) with respect to x_i versus A for all complexes and components in this study.

Table 1

Chains from the Structure Files Used in Each Calculation[a]

	molecule first chain second chain	**1ACB E I**	**1AVX A B**	**1BEB A B**	**1BRS A D**	**1EAW A B**	**1EMV A B**	**1HE1 A C**	**1PPE E I**	**1UDI E I**
protein A	A	48978	44830	35787	24965	48348	20160	29952	44429	50395
	γ_{rep}	0.017	0.020	0.017	0.023	0.012	0.024	0.019	0.017	0.017
	R^2	0.51	0.55	0.46	0.61	0.40	0.64	0.53	0.49	0.47
	γ_{vdw}	0.040	0.042	0.034	0.051	0.027	0.048	0.040	0.041	0.039
	R^2	0.15	0.14	0.12	0.20	0.08	0.21	0.15	0.14	0.13
protein B	A	15699	39342	35458	20705	14221	30953	39824	8048	20087
	γ_{rep}	0.025	0.019	0.020	0.024	0.027	0.019	0.018	0.032	0.022
	R^2	0.64	0.54	0.54	0.64	0.66	0.55	0.50	0.73	0.55
	γ_{vdw}	0.051	0.041	0.043	0.051	0.051	0.038	0.035	0.062	0.042
	R^2	0.19	0.15	0.16	0.19	0.22	0.13	0.10	0.34	0.15
complex	A	62232	81657	69623	43280	60246	48726	66480	49854	67477
	γ_{rep}	0.016	0.012	0.015	0.013	0.012	0.016	0.012	0.018	0.013
	R^2	0.49	0.38	0.40	0.39	0.37	0.49	0.36	0.49	0.38
	γ_{vdw}	0.042	0.029	0.030	0.026	0.025	0.037	0.028	0.045	0.030
	R^2	0.15	0.08	0.09	0.07	0.06	0.13	0.09	0.16	0.08
binding	γ_{rep}	0.017	0.014	0.017	0.017	0.014	0.019	0.015	0.019	0.016
	R^2	0.50	0.40	0.47	0.47	0.40	0.52	0.42	0.50	0.44
	γ_{vdw}	0.041	0.031	0.034	0.037	0.028	0.039	0.033	0.044	0.036
	R^2	0.15	0.09	0.12	0.14	0.09	0.14	0.12	0.16	0.12

[a]Horizontal sections labeled protein A, protein B, and complex give estimates of γ_{rep} and γ_{vdw} obtained by fitting $\partial\Delta G_{rep}/\partial x_i$ and $\partial\Delta G_{vdw}/\partial x_i$ versus $\partial A/\partial x_i$ and the squares (R^2) of the Pearson's correlation coefficients of these plots. The horizontal section labeled binding contains estimates of γ_{rep} and γ_{vdw} obtained by fitting $\partial\Delta\Delta G_{rep}/\partial x_i$ and $\partial\Delta\Delta G_{vdw}/\partial x_i$ versus $\partial\Delta A/\partial x_i$ and the squares (R^2) of the Pearson's correlation coefficients of these plots. The plots are in the Supporting Information. All A are in units of Å^2, and all γ_{vdw} and γ_{rep} are in units of kcal mol^{-1} Å^{-2}.

Table 2

Attractive (ΔG_{att}) and Electrostatic Components (ΔG_{el}) of the Solvation Energy of the 1BRS Complex and Its Two Components and the Attractive ($\Delta\Delta G_{att}^{desol}$ and $\Delta\Delta G_{att}$) and Electrostatic ($\Delta\Delta G_{el}^{desol}$ and $\Delta\Delta G_{el}$) Components of the Desolvation and Binding Free Energies[a]

	ΔG_{att}^{a}	ΔG_{att}^{b}	ΔG_{att}^{c}	$\Delta\Delta G_{att}^{desol}$	$\Delta\Delta G_{att}$
forward FEP	−315	−259	−458	116	2
backward FEP	−312	−259	−458	113	−1
TI	−311	−260	−457	114	−1
LRT	−325	−267	−480	112	−2
SSP λ = 0	−165	−143	−218	89	−25
SSP λ = 1	−485	−392	−742	135	21
	ΔG_{el}^{a}	ΔG_{el}^{b}	ΔG_{el}^{c}	$\Delta\Delta G_{el}^{desol}$	$\Delta\Delta G_{el}$
forward FEP	−1385	−1556	−2449	492	−17
backward FEP	−1387	−1558	−2443	501	−9
TI	−1388	−1562	−2448	501	−9
LRT	−1517	−1694	−2691	520	10
SSP λ = 1	−1528	−1670	−2678	520	10

[a]These energies were obtained by forward and backward free energy perturbation (FEP), thermodynamic integration (TI), linear response theory (LRT), and single-step perturbation (SSP) from either the beginning (λ = 0) or ends (λ = 1) of the integration. All energies are in units of kcal/mol.

Table 3

Slopes (*m*), *y*-Intercepts (*b*), and the Squares of Pearson's Correlation Factors (R^2) of Least-Squares Lines of the Quantities Listed in the *y*-Column versus Those in the *x*-Column[a]

x	y	R^2	m	b
A	$\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$	0.992	−0.050	−19
ΔA	$\Delta\Delta G_{\mathrm{att}}^{\mathrm{lrt,\ desol}}$	0.64	−0.077	−33
ΔA	$\Delta\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$	0.25	−0.040	−77
$\Delta G_{\mathrm{att}}^{\mathrm{SSP,\ 0}}$	$\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$	0.97	3.4	203
$\Delta\Delta G_{\mathrm{att}}^{\mathrm{SSP,\ 0,\ desol}}$	$\Delta\Delta G_{\mathrm{att}}^{\mathrm{lrt,\ desol}}$	0.64	1.60	−37
$\Delta\Delta G_{\mathrm{att}}^{\mathrm{SSP,\ 0}}$	$\Delta\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$	0.38	0.89	19
$\Delta G_{\mathrm{att}}^{\mathrm{SSP,\ 1}}$	$\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$	0.9991	0.58	−38
$\Delta\Delta G_{\mathrm{att}}^{\mathrm{SSP,\ 1,\ desol}}$	$\Delta\Delta G_{\mathrm{att}}^{\mathrm{lrt,\ desol}}$	0.97	0.60	35
$\Delta\Delta G_{\mathrm{att}}^{\mathrm{SSP,\ 1}}$	$\Delta\Delta G_{\mathrm{att}}^{\mathrm{lrt}}$	0.89	0.57	−10.5
$\Delta G_{\mathrm{el}}^{\mathrm{SSP}}$	$\Delta G_{\mathrm{el}}^{\mathrm{lrt}}$	0.9997	1.01	16
$\Delta\Delta G_{\mathrm{el}}^{\mathrm{SSP,\ desol}}$	$\Delta\Delta G_{\mathrm{el}}^{\mathrm{lrt,\ desol}}$	0.99997	1.0007	2.0
$\Delta\Delta G_{\mathrm{el}}^{\mathrm{SSP}}$	$\Delta\Delta G_{\mathrm{el}}^{\mathrm{lrt}}$	0.9999	1	1.6
$\Delta G_{\mathrm{el}}^{\mathrm{APBS}}$	$\Delta G_{\mathrm{el}}^{\mathrm{lrt}}$	0.96	1.03	−138
$\Delta\Delta G_{\mathrm{el}}^{\mathrm{APBS,\ desol}}$	$\Delta\Delta G_{\mathrm{el}}^{\mathrm{lrt,\ desol}}$	0.92	0.40	125
$\Delta\Delta G_{\mathrm{el}}^{\mathrm{APBS}}$	$\Delta G_{\mathrm{el}}^{\mathrm{lrt}}$	0.42	−7.57	510

[a]The corresponding plots are in Supporting Information. All energies are in kcal/mol. All areas and changes in area are in units of Å^2. All *m* and *b* are in the corresponding units.