JCTC Journal of Chemical Theory and Computation

Rapid Estimation of Solvation Energy for Simulations of Protein−Protein Association

David S. Cerutti,*[,§] Lynn F. Ten Eyck,[‡,§,||] and J. Andrew McCammon[†,§,||]

Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, Department of Pharmacology, and San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0365

Received September 3, 2004

Abstract: We have formulated the Energy by Linear Superposition of Corrections Approximation (ELSCA) for estimating the electrostatic and apolar solvation energy of bringing two proteins into close proximity or into contact as defined by the linearized Poisson−Boltzmann model and a linear function of the solvent-accessible surface area. ELSCA utilizes potentials of mean force between atom types found in the AMBER *ff*99 force field, a uniform distance-dependent dielectric, and a potential that mimics the change in solvent accessible surface area for bringing two solvated spheres into contact. ELSCA was trained by a linear least-squares fit on more than 39 000 putative complexes, each formed from pairs of nonhomologous proteins with a range of shapes, sizes, and charges. The training set was also designed to capture various stages of complex formation. ELSCA was tested against over 8000 non-native complexes of 45 enzyme/inhibitor, antibody/antigen, and other systems that are known to form complexes and gives an overall correlation of 0.962 with PBSA-derived energies for these complexes. The predictions have a slope of 0.89 on the actual values with a bias of 11.1 kcal/mol. When applied to native complexes of these 45 protein systems, ELSCA reproduces PBSA results with a correlation of 0.787, a slope of 1.13, and a bias of 13.0 kcal/mol. We report parameters for ELSCA in the context of the AMBER *ff*99 parameter set. Our model is most useful in macromolecular docking and protein association simulations, where large portions of each molecule may be considered rigid.

Introduction

The protein docking problem, most generally the challenge of finding the structure of a complex of two proteins given the three-dimensional structures of the isolated components, is difficult due to the degree of sampling required and the limitations of scoring functions. In particular, while gas-phase molecular mechanics energies are readily computed by docking programs, the change in solvation energy involved in forming a particular complex is not rigorously treated by most scoring functions presently in use. Although implicit-solvent methods such as Poisson−Boltzmann/Surface Area (PBSA) and Generalized Born/Surface Area (GBSA) are widely used for calculating solvation energies during complex formation,[3,24,37] they are applicable for screening at most a few thousand putative complexes after lower-resolution searches. If the solvation energy involved in forming a complex between two proteins were formulated as a sum of pairwise additive interactions between atoms of the ligand and receptor, this quantity could be readily computed along with the gas-phase molecular mechanics energies that most real-space docking programs already employ, abrogating the need for refinement of the solutions produced and extending implicit-solvent methodology for protein docking and efficient simulations of macromolecular encounters.

* Corresponding author phone: (858)534-2798; e-mail: dcerutti@mccammon.ucsd.edu.
† Howard Hughes Medical Institute.
‡ San Diego Supercomputer Center.
§ Department of Chemistry and Biochemistry.
|| Department of Pharmacology.

10.1021/ct049946f CCC: $30.25
Published on Web 11/24/2004

In implicit solvent methodology, the solvation free energy is typically approximated as the sum of electrostatic and nonpolar contributions.[38] Numerous studies have demonstrated the importance of electrostatics in biomolecular recognition.[2,6,7,21,43,49] In solvated systems, electrostatic forces may act over long ranges to steer protein partners and accelerate the rate of complex formation,[20,22] affect the structure of highly charged nucleic acids,[1] and order the ions around those structures.[11] Favorable electrostatic interactions exist between polar or charged regions of biomolecules and polar water molecules. During the final stages of the association of biomolecules, stripping away nearby water molecules makes an unfavorable contribution to the binding free energy and may even outweigh salt bridges and favorable dipole interactions formed by the association.[40] To whatever degree hydrophobic and apolar forces drive the final stages of complex formation, electrostatic effects may still help to specify interactions. In contrast, nonpolar interactions comprise the work of cavity formation and the dispersion forces acting between solute and solvent.[38] Breaking solute−solvent dispersion interactions during complex formation is energetically unfavorable, but this is compensated by burial of solvent-accessible surface area which reduces the work required to form a solvent cavity for the proteins.

Accurate estimates of the electrostatic interaction energy of two solvated biomolecules at long range can be calculated accurately and cheaply so as to predict parameters such as relative association kinetics through Brownian dynamics studies.[20,23] However, at close range, accurate electrostatic energies are much more expensive to calculate due to desolvation effects and the disparity between the polarity of water and the apolar residues in the cores of proteins. This is a major obstacle to accurate prediction of absolute association rates of biomolecules and the structures of their complexes. The Poisson−Boltzmann approximation, a merger of Poisson's equation for macroscopic electrostatics and a mean-field treatment of the spectator ions, has been used extensively for analyzing the electrostatics of macromolecules, particularly the association of proteins,[18,20,40] but this computation typically requires several minutes on a modern workstation. Approximations to Poisson-derived hydration energies such as Generalized Born parametrizations[17,28] are faster, but are still expensive for lengthy macromolecular simulations and far too expensive for general searches in protein docking applications.

The nonpolar solvation energy of proteins is commonly formulated as a linear function of the solvent-accessible surface area, based on observations of the solvation energy of linear alkanes.[42] While this rule does not generalize perfectly to small molecules of other shapes,[52] nor to proteins,[29] most Poisson−Boltzmann electrostatic models are benchmarked alongside such models (see ref 41), for example) and so require this term for computation of total solvation energy. While calculating the solvent-accessible surface area of a molecule with N atoms can be written as an O(Nlog$N$) problem,[39] the number of computations per atom is still large, and the amount of coding needed to rigorously implement this calculation in a docking or simulation program is considerable. Square-well potentials[33] and other rough measures of shape complementarity[12] are therefore used.

For computational convenience, scoring functions presently used in protein docking applications neglect the most physically meaningful aspects of solvation: the fact that buried atoms in the ligand oppose electrostatic fields much less effectively than polar water molecules, the solute−solvent dispersion interactions, and the change in the work of cavity formation upon binding. At long ranges, these effects are negligible, but protein docking requires millions to billions of accurate estimates for closely interacting proteins, and simulations of macromolecular encounters require accurate energy estimates for all degrees of separation.

In the present work we introduce the Energy by Linear Superposition of Corrections Approximation (ELSCA), a correction scheme based on a distance-dependent dielectric (DDD) that is uniform (e.g. its form is the same regardless of the local environment), a scalable function describing buried surface area between two interacting spheres, and a set of potentials of mean force (PMFs) between distinct types of atoms. We compute solvation energies using PBSA parameters for modeling proteins with the AMBER *ff*99 parameter set[14,49] for more than 39 000 putative complexes and close associations of 21 distinct proteins with differing charges, shapes, and sizes. For convenience of implementation, ELSCA is fit to reproduce the change in solvation energy plus the gas-phase Coulombic association energy of two proteins. To demonstrate ELSCA's transferability, we apply it to an independent set of 45 native and over 8000 non-native protein complexes. All functions used in this model can be conveniently superimposed on grid-based potentials or lookup tables, implying no additional computational cost during docking studies and simulations. The purpose of ELSCA is most like that of a set of potentials of mean force developed by Jiang and co-workers[26] for the total free energy change of bringing two proteins together in a particular conformation. However, ELSCA should also be considered in the context of other pairwise potentials for calculating the solvation energy of individual species such as EEF1[27] and atomic contact energies.[54]

Methods

Approximation of the Total Electrostatic and Nonpolar Solvation Energy. To break Poisson−Boltzmann electrostatics and change in solvent-accessible surface area into the pairwise interactions of ELSCA, we make use of four approximations. The first and most basic approximation, a screened Coulombic interaction, stems directly from Debye−Hückel theory[34] and is the dominant contribution for interactions at distances greater than 60 Å. The second approximation, a screened Coulombic interaction attenuated by a uniform distance-dependent dielectric (DDD), affects intermediate and short-range interactions. The third, a set of pair potentials (PMFs) between sixteen atom types, is effective at distances less than 15 Å. The fourth, a function that roughly reproduces the change in solvent accessible surface area as two spheres of given radii approach one another, is

effective at distances on the order of the diameter of the solvent probe. The first approximation has no adjustable parameters, but the second and third have a total of 413 (see below) and the fourth is scaled by a single parameter. All parameters are fit simultaneously by solving a linear least-squares problem. The total electrostatic and solvation energy is the sum of energies from all four approximations.

At large distances, the electrostatic interaction of two charges in a neutral solvent is given by Coulomb's law scaled by the appropriate solvent dielectric constant. In dilute ionic solutions, however, the long-range electrostatic potential is screened exponentially as described by Debye−Hückel theory. We account for this explicitly using [1], where i and j run over the atoms in the ligand and receptor, k runs over the two ionic species in the continuum solvent, q represents an atomic partial charge, β is the inverse product of Boltzmann's constant and the absolute temperature, and N is the number density of an ionic species in bulk solvent.

$$E_{\text{basic}} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_{\text{water}} r_{ij}} \exp\left[- \sqrt{\sum_k \frac{\beta q_k^2 N_k}{\epsilon_0\epsilon_{\text{water}}}} r_{ij} \right] \quad (1)$$

Distance-dependent dielectrics are often computationally convenient, whether realistic or not. In the docking problem, a uniform DDD is particularly convenient as the energy remains a sum of pairwise interactions. We formulate the second part of our method using a set of Gaussians to attenuate the screened Coulombic potential as shown in [2].

$$E_{\text{DDD}} = E_{\text{basic}} \left[1 + \sum_{i,j} \left(\sum_{\alpha=1}^{5} \exp\left(- \frac{0.02}{\alpha^2} r_{ij}^2 \right) S_{\text{DDD}}^{(\alpha)} \right) \right] \quad (2)$$

The value of 0.02 in the exponential argument was chosen along with five Gaussian terms to provide a basis for finely tuning the dielectric constant at distances between 0 and 60 Å, where the interacting ligand and receptor are no longer adequately described as collections of Debye−Hückel spheres.

The electrostatic energy of closely associated proteins is not adequately described by pairwise charge−charge interactions between the solutes due to desolvation effects and the fact that protein interiors dampen electrostatic fields much less effectively than bulk water. To account for these effects while still using pairwise interactions, we parametrize a set of $M(M+1)/2$ PMFs for M atom types. Each PMF is in turn a linear combination of k_{PMF} basis functions $g(r)$ [3], giving $k_{\text{PMF}}M(M+1)/2$ scaling parameters that must be solved for. The correction to the electrostatic energy E_{corr} is given by [3 and 4]:

$$f_{np}(r) = \sum_{\alpha=1}^{k_{\text{PMF}}} g_\alpha(r) S_{\text{np}}^{(\alpha)} \quad (3)$$

$$E_{\text{corr}} = \sum_{i=1}^{L} \sum_{j=1}^{R} \sum_{n=1}^{M-1} \sum_{p=n+1}^{M} \delta_{T_i n} \delta_{T_j p} f_{np}(r_{ij}) \quad (4)$$

Above, n and p are involved in summations over all atom types, T_i and T_j represent the types of the ith and jth atoms, respectively, $S_{np}^{(\alpha)}$ represents the αth scaling parameter of the PMF between atoms of types n and p to be solved for, and δ is the Kronecker delta. The total electrostatic estimate is thus a linear function of the scaling parameters $S_{np}^{(\alpha)}$ and $S_{\text{DDD}}^{(\alpha)}$. With a number of putative complexes generating at least $(k_{\text{PMF}}M(M+1)/2) + 6$ sets of distinct coefficients, the scaling parameters may be obtained by solving a least squares problem $\mathbf{A}\boldsymbol{s} = \boldsymbol{b}$. Each row of the matrix $\mathbf{A}$ is a set of coefficients for each $S_{\text{DDD}}^{(\alpha)}$ and $S_{np}^{(\alpha)}$ obtained from one putative complex. $\boldsymbol{s}$ represents the vector of scaling parameters for the DDD and PMFs, and $\boldsymbol{b}$ represents the vector of total electrostatic and nonpolar solvation energies. Operationally, each row of matrix $\mathbf{A}$ is filled by first initializing it to zero and then looping over atoms j of one molecule nested within a loop over atoms i of the other. In the inner loop, the value of each relevant basis function given the distance r_{ij} is added to the appropriate column of the row of matrix $\mathbf{A}$.

The PMF basis functions $g(r)$ must be well-behaved in that $g(r)$ has finite values at $r = r_0$ and at $r = r_{\text{cut}}$, where r_{cut} is the cutoff distance for applying the PMF (set equal to the cutoff for vdW interactions in our model) and $r_0 < r_{\text{cut}}$. Because the electrostatic interactions between receptor and ligand at long ranges are approximated well by treating the ligand as a set of point charges moving within the electrostatic field of the receptor, it is preferable to have PMFs, and therefore to select basis functions, that diminish as r approaches r_{cut}. Piecewise Gaussians of the form 5 were chosen.

$$g_\alpha(r) = \begin{cases} \exp\left(- \dfrac{\beta_{\text{PMF}}}{(\alpha + \lambda_{\text{PMF}})^2} (r - \sigma_{np})^2 \right), & r \geq \sigma_{np} \\ 1, & r < \sigma_{np} \end{cases} \quad (5)$$

In [5], σ_{np} is the Lennard-Jones parameter for atom types n and p except in cases where n or p is HO (see Table 1), in which case σ_{np} is zero.

For constructing the PMFs, three criteria were considered in determining the number of basis functions k_{PMF} in [4] and

Table 1. Atom Types for ELSCA Parameters[a]

ELSCA name	AMBER *ff*99 names
CT	CT
C	C
CR	CA, CC, CV, CW, CR, CB, C*, CN
N	N, NA, NB, N2, N3
O	O
O2	O2
OH	OH
S	S, SH
H	H, HS
HA	HA
HP	HP
HC	HC
HO	HO
H1	H1
H4	H4
H5	H5

[a] The atom types defined for ELSCA comprise all those found in amino acids in AMBER *ff*99. Distinct Lennard-Jones parameters as well as considerations to broad classes of chemical groups led to these choices.

spread of basis functions as dictated by the β_{PMF} and λ_{PMF} in [5]: the RMSD of estimated energies compared to those derived by PBSA, the similarity in correlations obtained in the training set versus those in the test set, and the range of values taken by the PMFs as r approached σ_{np}. The third criterion was included to ensure that no particular pairwise interaction could drastically affect the estimated score, based on our observation that ELSCA could become particularly sensitive to atoms in close contact if given large values of β_{PMF} in conjunction with a small λ_{PMF}.

Distinct atom types for defining the PMFs were first identified by distinct Lennard-Jones parameters within AMBER *ff*99 (see *Methods*, *Protein Docking*). Distinctions were also made between *sp2* hybridized carbon atoms in aromatic groups, *sp2* hybridized carbon atoms in carbonyl groups, oxygen atoms in amides or carbonyl groups, and oxygen atoms in carboxylic acids. These distinctions help to distinguish between grossly different electronic environments but do not permit ELSCA to parametrize extremely rare atom types, such as carbons in imidazole rings, which would lead to overfitting. The AMBER *ff*99 atom types, and the corresponding names we give them, are found in Table 1. In total, these 16 atom types provided 136 PMFs for fitting.

Finally, at very short distances, the change in solvent-accessible surface area was approximated by a piecewise function that closely reproduces the change in solvent-accessible surface area when two spheres of radii R_i and R_j ($R_i > R_j$) come into contact with each other considering solvent probe radius R_s. This function is $4\pi(Rj + R_s)^2$ for $r < (R_i - R_j)$, a cubic spline connecting the points $(R_i - R_j, 4\pi(R_j + R_s)^2)$ and $(R_i + R_j + 2R_s, 0)$ with zero derivative at its extrema, and zero elsewhere. This function is scaled by a constant fit in tandem with the PMFs and DDD. This is a first-order correction to the SA term, in analogy to the DDD as a first-order correction to the PB term. We set R_s equal to the solvent probe radius used in PBSA calculations (see below). However, for convenience, the atomic radii for this calculation were taken from the Lennard-Jones parameters in *ff*99. While these radii do not perfectly correspond to the AMBER *ff*99 PBSA radii, they are adequate for estimating the change in surface area in our model.

Protein Modeling. For training ELSCA, a set of high-resolution, nonhomologous proteins was chosen from the Protein Data Bank[5] using the PISCES server[48] to have R-factor less than 0.18 and less than 5% sequence similarity. From the resulting list of 441 proteins, 21 were chosen to represent a range of sizes and charges, with the additional criteria that there be no cofactors present in the selected proteins and that their structures be fully resolved. PDB ID numbers and summary data for the systems are given in Table 2. Hydrogens found in AMBER *ff*99 were added using Leap in the AMBER8[55] package, but to give some consideration to the protonation state of histidines, the pdb2gmx utility from the GROMACS software package[4,15,30,45,50] was used to dictate protonation on N^{ϵ} (the default state for pdb2gmx), N^{δ}, or both. Because the initial structure fed to our docking program is taken as a reference state, each structure was energy-minimized in vacuo using the Sander program in AMBER8 through 50 cycles of steepest descent and up to 950 cycles of conjugate gradient minimization or until the potential energy converged to 1.0×10^{-4} kcal/mol. Backbone heavy atoms were restrained to their initial coordinates by 1.0 kcal/mol-Å^2 harmonic potentials during the minimization.

Table 2. Proteins Used in Training

PDB ID[b]	resolution[a]	residues	charge (e)
1jcd	1.30	50	−2
1gvd	1.45	52	+8
1mof	1.70	53	−3
1i2t	1.04	61	−1
3ebx	1.40	62	+2
1ok0	0.93	74	−5
1c5e	1.10	95	−2
1lni	1.00	96	−7
1jo0	1.37	97	+3
1eaj	1.35	127	−3
2lis	1.35	131	+11
1cxq	1.02	162	+10
1nwa	1.50	168	0
1mf7	1.25	194	+3
1qhv	1.51	195	−1
1g61	1.30	225	−14
1dj0	1.50	264	+4
1gxm	1.32	324	+4
1qcx	1.70	359	−7
1ug6	0.99	426	−6
1iat	1.62	556	+3

[a] Crystal diffraction resolution, Å. [b] Of the cases listed, where the PDB file contained multiple chains, the A chain was the one selected by PISCES.

Protein Docking. To parametrize ELSCA as described in *Approximation to the Total Electrostatic Energy*, we required a diverse set of pairs of proteins in various hypothetical complexes that are packed together to varying degrees. We gave consideration to side-chain flexibility in our docking by allowing all side-chains to take on configurations found in a detailed rotamer library,[19] but in order to avoid biasing the training set toward complexes with favorable electrostatics, only steric factors were included in the scoring function. Each of the 21 proteins in the homology-culled set was first docked to every other along 10 000 random approach vectors. This set was pruned of any docked solutions that put the two proteins in very similar orientations. The top-scoring 100 complexes generated in the pruned set (the tightly packed set) were subjected to additional refinement by allowing additional rigid backbone moves with side-chain flexibility as described above. The next 60 high-scoring complexes (the loosely packed set) were refined with flexible side-chains but no additional backbone moves. Every 50th complex from the rest of the pruned set, up to a total of 50 (the perturbed set), was perturbed by pulling the proteins apart by one to four solvent probe diameters and then performing flexible side-chain refinement. Although we intend the software that performed these manipulations to be used in protein docking, this cursory scheme was intended to generate random juxtapositions of proteins with varying degrees of tightness, not correct docked answers.

PBSA Calculations. The thermodynamic cycle used to calculate the total electrostatic energy of complex formation by solving the linearized Poisson−Boltzmann equation [6] has been extensively used in the literature and is clearly illustrated in the work of Elcock and coworkers.[20]

$$\nabla\cdot(\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})) = -4\pi[\rho_{\text{pro}} - \sum_k \beta q_k^2 N_k \phi(\mathbf{r})] \qquad (6)$$

In [6], ϵ represents the dielectric scalar field, ϕ is the electrostatic potential, ρ_{pro} is the macroscopic charge density due to the solute, β is the reciprocal of temperature times Boltzmann's constant, and q_k and N_k are the charge and number density of the kth ion in solution. The electrostatic energy of a solute is given by [7]:

$$E_{\text{elec}} = \frac{1}{2}\int\int_V\int \phi(\mathbf{r})\rho_{\text{pro}}(\mathbf{r})\mathrm{d}\mathbf{r} \qquad (7)$$

The electrostatic free energy of forming a complex between two proteins with low internal dielectric ϵ_{pro} in continuum water with high dielectric ϵ_{water} is equal to the sum of the energies required to bring each isolated protein out of water into a phase in which the dielectric of the continuum solvent is ϵ_{pro}, the energy of bringing the two proteins together as calculated by Coulombic interactions with dielectric ϵ_{pro}, and the energy of returning the complex to the continuum water. This cycle is equivalent to calculating the electrostatic energies E_{elec} of each separate protein and the complex using identical grids, subtracting the self-energy artifacts created by mapping the point charges to the grid points, and taking the difference in the energy of the complex and its components.

We used UHBD[32] to solve the LPBE for each putative complex produced by our docking studies on each protein system. Grids with identical dimensions and alignment were calculated for each separate protein and the complex. To minimize boundary artifacts, a coarse grid was first calculated, large enough that solvent filled roughly 95% of the volume, using Coulombic potentials with Debye−Hückel screening for boundary conditions. A finer, focused grid was used for the thermodynamic calculations, generated using the coarse grid for boundary conditions and large enough to encompass the complex by at least 5 Å on all sides.

In our PB formulation, we conformed to the methods used for PBSA in AMBER *ff*99. For all proteins, ϵ_{pro} was set to 1 and ϵ_{water} was set to 80. Dielectric smoothing[16] was applied in all cases. A monovalent implicit ion concentration of 0.100 M was used throughout the studies with an implicit ion radius of 2.1 Å. A grid spacing of 0.5 Å, or the minimum spacing needed to satisfy the aforementioned size requirements with a grid of 240 points on a side, was used for the focused grids. In all calculations, the low-dielectric region was defined by the solvent-excluded volume with a solvent probe of radius 1.6 Å and solute atom radii optimized for PBSA calculations in the context of AMBER *ff*99 (readers should consult the AMBER 8 source code file pb_init.f for details). To ensure convergence in each result, the UHBD finite-difference solver was permitted to run for up to 200 iterations, twice the default number.

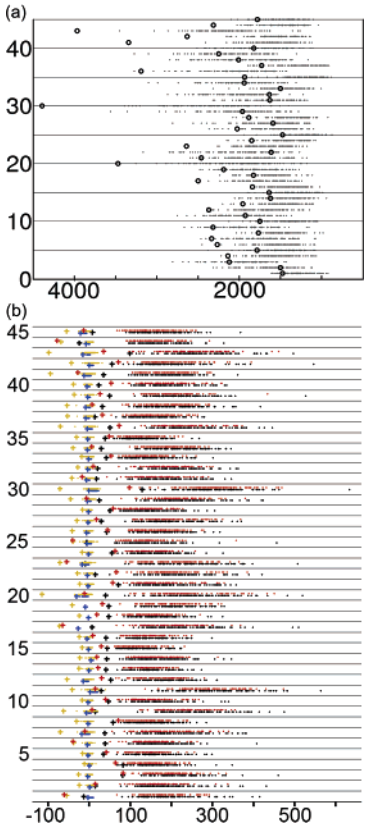

Figure 1. Properties of native vs non-native protein complexes. In (a), the solvent-accessible surface area buried by non-native complexes is shown by gray points, that of native complexes by black donuts. The *x*-axis has units of Å^2. In (b), the electrostatic energy of the complexes according to numerous methods is shown. Plus signs represent native complexes and points represent non-native ones. PB-derived energies (without consideration to the apolar solvation energy given by a change in SASA) are shown in black, analytic Coulombic ($\epsilon = 78$) just below the PB results in blue, linear DDD ($\epsilon = 5r$) just above the PB results in yellow, and ELSCA results (including apolar solvation energy) above those of the linear DDD in red. The *x*-axis has units of kcal/mol. Each pair of proteins is represented on a different level in each chart. In ascending order on both charts, the pairs are the separated components of crystal structures 1A0O, 1BVK, 1KXV, 1STF, 1ACB, 1CGI, 1FSS, 1TAB, 1AHW, 1CHO (10), 1GLA, 1MAH, 1TGS, 2PTC, 1CSE, 1MEL, 1UDI, 2SIC, 1AVW, 1DFJ (20), 1IAI, 1MLC, 1UGH, 2SNI, 1AVZ, 1DQJ, 1IGC, 2TEC, 1BQL, 1EFU (30), 1JHL, 2VIR, 1BRC, 1EO8, 1PPE, 3HHR, 1BRS, 1FBI, 1QFU, 2JEL (40), 4HTC, 1BTH, 1FIN, 1SPB, and 2KAI.

Test Cases for ELSCA. To validate ELSCA, we selected 45 protein complexes from a protein:protein docking benchmark[13] of nonredundant test cases. PDB identifiers for these complexes are given in Figures 1 and 2. Because protein complex structures are considerably rarer than those of individual proteins, we accepted members of the benchmark into our test set so long as they did not have deeply buried cofactors or cofactors near the binding site. Residues immediately before or after an unresolved portion of the backbone were treated as C- and N-terminal residues, respectively. Unresolved side-chains were rebuilt using SCWRL3.0.[8,10] Each protein complex was energy-minimized

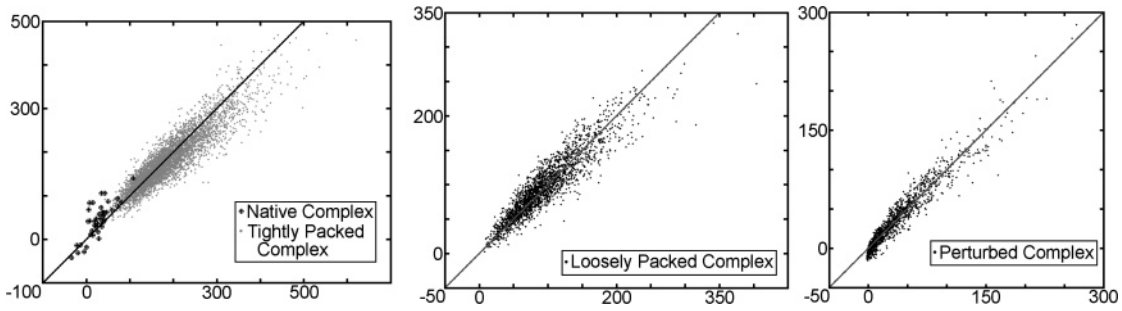

Figure 2. ELSCA predictions for native, non-native, and separated complexes. ELSCA was trained on 39 153 non-native complexes of proteins that are not known to form complexes and then tested on 8161 non-native and 45 native complexes of proteins that are known to associate. The *x*-axis gives the energy, in kcal/mol, of the complex according to the AMBER *ff*99 PBSA formulation, the *y*-axis the ELSCA-predicted value. Predictions for each of the three complex types described in *Methods: Protein Docking* are displayed separately.

according to the protocol used for the training set proteins. The complexes were then separated and docked to their partners in the manner described in *Methods: Protein Docking*. No docking between noninteracting proteins was attempted in the test set. None of the complexes produced by the cursory manipulation scheme came within 10 Å RMSD of the native complex.

Results

Protein Docking and Complex Selection. To parametrize a set of pairwise interactions that emulate PBSA results for general interacting proteins, we created a set of putative protein complexes from a nonhomologous set of peptide chains, all with different orientations and degrees of association. In docking the proteins, we permitted limited flexibility by allowing all side-chains in either molecule to take on discrete rotamer states (see *Methods*, *Protein Docking*), which permitted closer associations and helped to ensure that our training set included some diversity in the conformation of each molecule.

To test whether our methods of producing the training set can generate complexes that appear native by geometric and electrostatic standards, we compared the amount of SASA buried by native complexes in the test set to that buried by tightly packed configurations of the same proteins (Figure 1a). For some complexes our docking method produced non-native structures with comparable buried SASA, although native complexes often bury more surface area than any of the decoys. Figure 1b reveals that our method only rarely creates tightly packed complexes with comparable electrostatics to the native configuration. According to our Poisson−Boltzmann model, the electrostatic energy of the native complexes is optimal for nearly every pair of proteins in the test set, and there is a considerable variation in the energy of non-native complexes.

We are currently unable to test whether PB electrostatics are truly discriminatory of native complexes because we have not yet used PB directly to energy-minimize putative complexes of interacting proteins. Even while we are working to overcome this limitation, the data in Figure 1b suggest that the PB electrostatic model is an outstanding indicator, relative to other simple electrostatic models, of a native complex among other members of the tightly packed set. In contrast to PB, a homogeneous dielectric model ($\epsilon = 78$) typically suggests a "golf-course" electrostatic energy landscape in protein association. The linear DDD ($\epsilon = 5r$) appears to be much better at distinguishing native complexes from non-native, a result that may help to explain the success of linear DDDs in other docking applications.[53,25] However, such a model ultimately has no physical basis, which would hinder attempts to develop more complete potentials that take into account nonpolar solvation effects.

Optimizing Basis Functions. A large set of PMFs was the primary means of accurately reproducing Poisson−Boltzmann electrostatics using pairwise interactions. However, a uniform DDD and a crude surface area function helped make gross corrections to the electrostatic and nonpolar solvation energy estimates so that the PMFs could make finer corrections. A screened Coulombic potential ensured that long-range electrostatic interactions were realistically captured. As stated in *Methods*, the basis set for the DDD was chosen so that the dielectric constant could be tuned at short-range and intermediate distances, where the PMFs taper and the screened Coulombic approximation begins break down. We found $k_{PMF} = 3$, $\beta_{PMF} = 0.6$, and $\lambda_{PMF} = 1$ to yield good results according to the criteria set forth in *Methods*; addition of more basis functions could not significantly improve the quality of the fit nor could changing the broadness of the set of basis functions (data not shown).

A successful model must correctly predict the energy of both native and non-native complexes. As shown in Figure 2, ELSCA is very good at predicting the energies of non-native complexes of proteins in the test-set, yielding an overall correlation coefficient of 0.962 (if ELSCA were applied to its own training set, the correlation is 0.964) and a RMSD of 26.8 kcal/mol between the predicted and calculated energies (Figure 2). The regression line for predicted on calculated values of all non-native complexes has a slope of 0.89 and bias 11.1 kcal/mol. For predicting native complexes, the correlation is somewhat lower (0.787), the RMSD is 29.2 kcal/mol, and the regression line shows a slope of 1.14 with a bias of 13.0 kcal/mol. For comparison, a homogeneous dielectric model yields a poor correlation of 0.229 with our PB results for native complexes, and a linear distance-dependent dielectric model yields a correlation of only 0.083.

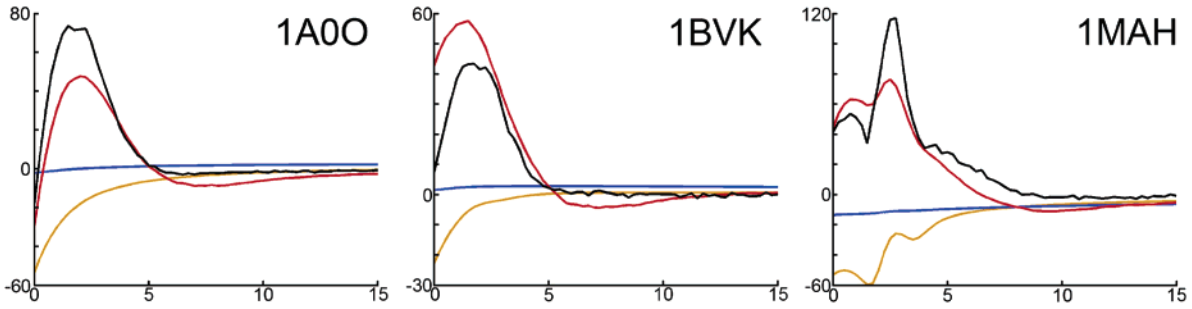

Figure 3. Reproduction of energy landscapes for three complexes. ELSCA was used to reproduce electrostatic plus apolar solvation energy (*y*-axis, units of kcal/mol) along a hypothetical, linear reaction pathway for three different protein systems (*x*-axis denotes distance from the native state, in Å). Similar to Figure 1b, black represents PBSA energy, red ELSCA, blue a homogeneous dielectric ($\epsilon = 78$), and yellow a linear DDD ($\epsilon = 5r$).

Energy along Reaction Pathways. To demonstrate ELSCA's ability to predict the features of the energy landscape, we selected three protein complexes with different characteristics from the test set and computed the total electrostatic and solvation energy using PBSA, ELSCA, $\epsilon = 78$, and $\epsilon = 5r$ as the ligand was pulled directly away from the receptor along a putative reaction pathway. Proteins were held rigid during these tests to eliminate energy contributions from conformational change in either protein. Measurements were taken at 0.25 Å intervals, up to 15.0 Å from the native state. We selected systems 1MAH (acetylcholinesterase and fasciculin-2), 1BVK (lysozyme bound to a lysozyme-specific antibody), and 1A0O (chea in complex with chey). Despite quantitative errors in the energy estimates noted above and in Figure 2, Figure 3 demonstrates that ELSCA creates a smooth potential energy landscape that reproduces the features of the benchmark PBSA model, and thus the critical effects of solvation in protein interactions, in ways that simpler electrostatic models cannot. ELSCA can produce energies that are wrong in the native state, but as shown by the plot of 1BVK approach, ELSCA at least reproduces the qualitative features of the energy landscape. Comparison of the parametrized ELSCA model to the other electrostatic models (Figure 1b) reveals that the native state is again predicted to be the minimum or (in one case only) within 3 kcal/mol of the minimum energy of the tightly packed set. Figure 1b compares ELSCA to PB electrostatic energies; we remind the reader that ELSCA includes a SASA term in its formulation (see *Methods: Approximation to the Total Electrostatic and Nonpolar Solvation Energy*) that cannot be simply dissected from the electrostatic contributions. However, the effect of the SASA term is small (5 kcal/mol per 1000 Å^2 buried SASA) compared to the magnitudes of the PB electrostatic energies.

The parameters for the surface area term, PMFs, and uniform DDD obtained by the least-squares fitting are given in the Supporting Information.

Discussion

Reproduction of PBSA Results. ELSCA's estimates compare to PBSA-derived energies with very high correlations. Furthermore, ELSCA is equally predictive of two very large, nonredundant sets of proteins, indicating that it is applicable to proteins in general. The parameters k_{PMF}, β_{PMF}, and λ_{PMF}, which are most important for correctly determining the energy of closely packed complexes, are somewhat arbitrary; a comparable quality of fit can be obtained with numerous PMF and DDD bases, but significant improvements over our formulation were not found. These results suggest that ELSCA is nearly the best possible rendering of the PBSA free energy model in the form of pairwise interactions between the atoms of each solute; additional improvements require consideration of the local atomic environments.

In this study, we have parametrized ELSCA over a set of non-native complexes that were not optimized in terms of electrostatics because the vast majority of structures that will be sampled in protein docking or association simulations will be non-native. Despite the fact that electrostatic complementarity typifies protein interfaces, the parameters transfer well to native complexes. Indeed, the data points for native complexes in Figure 2 can be seen as an extension of the data for the most tightly packed non-native complexes. The lower correlation is largely due to the fact that native complexes have consistently low energy in the PBSA model, whereas non-native, tightly packed complexes can have a broad range of energies. Further improvements in this regime may be possible by including hundreds to thousands of natural interfaces in the training set.

ELSCA's potentials of mean force and uniform distance-dependent dielectric are meant to be superimposed on van der Waals and electrostatic grids and lookup tables so that ELSCA may be as computationally tractable as any gas-phase energy calculation or grid-based calculation used in simulations and docking studies. Implicit solvent simulations of protein association can be significantly accelerated using ELSCA rather than a model such as GBSA which depends on frequently recalculating atomic polarization radii. Significant enhancements are also possible for methods based on rotamer libraries which require pairwise additive interactions. Once the ELSCA functions are stored in lookup tables (and, ideally, mapped to grids), the implementation is trivial. Because the PMFs used in ELSCA are smoothly varying, grid-based artifacts should be minimal when ELSCA is applied in grid-based calculations.

Alone, none of the PMFs has physical significance and the DDD is realistic only at distances greater than 60 Å, where it reproduces Debye−Hückel screening effects of the ionic solvent. Together, however, all these functions repro-

duce a model that can realistically account for desolvation and solvent ion effects. For docking studies, ELSCA is expected be nearly as useful as the PBSA model because it reproduces the features of the PBSA energy landscape (Figure 3) and obtains few competing minima that PBSA would not have (Figure 1a,b). However, there is ongoing debate about the validity of implicit solvent models. ELSCA has promise as an investigative tool, but like Generalized Born and other theoretically benchmarked approximations it is ultimately limited by the model it emulates.

ELSCA in Context. A number of studies have used potentials of mean force between multiple atom types to approximate nonlinear functions in molecular modeling, many of them stemming from the work of Miyzawa and Jernigan.[35,36] Such models often rely on short-ranged potentials parametrized by statistical distributions found in a homology-culled database of protein structures such as the Protein Data Bank.[5] The set of atomic contact energies (ACE)[54] is a very detailed example of such "knowledge-based" potentials, using step-functions to describe the transfer of various atoms in the selected force field between hydrophobic and aqueous environments. ACE accounts for solvation energy contributions to protein stability but must be augmented by an approximation for solvent screening electrostatic interactions. In contrast to other PMF models that are only able to assess the energy of a bound state, ELSCA is intended specifically for estimating the electrostatic energy of protein:protein association as the partners come within a few solvent layers of one another and as they bind. The basis functions used in ELSCA are continuously differentiable and longer-ranged than those used in most other PMFs, essential features of a model that is intended to predict interactions over a range of intermolecular separations.

By assuming a Hamiltonian that separates into solvation free energy and molecular-mechanics interactions and the existence of pairwise additive potentials of mean force for various chemical groups, Lazaridis and Karplus built the EEF1 model[27] for solvation energy of proteins. These are the same basic assumptions that were used to construct ELSCA. Again like ELSCA, EEF1 uses Gaussians as the basis functions for its PMFs and works in tandem with a uniform distance-dependent dielectric ($\epsilon = r$ in this case) and can be superimposed on the gas-phase molecular mechanics energy for a modest computational cost. EEF1 is intended for use in protein folding and single-protein dynamics problems, where the internal energy of the molecule varies as its conformation changes. Such a model could be a powerful complement to ELSCA, which addresses the change in solvation and electrostatic association energy in complex formation, not the energy of changing each molecule's configuration from a particular reference state.

One other pairwise additive PMF model should be mentioned because of its notable similarity to ELSCA in design and purpose. Jiang and co-workers[26] built a model for the total free energy of protein:protein association using four general heavy atom types, distinguishing between atoms of each type on the ligand and receptor to construct sixteen PMFs. This model achieved a correlation coefficient of 0.75 between its scores and the experimental binding energies of 28 protein complexes and recovered important features such as a potential well between hydrogen-bond donors and acceptors at the correct hydrogen-bonding distance. This is nearly the consistency by which ELSCA, given its current training set, reproduces native complex energies, but there is no way to judge the model's accuracy in reproducing nonisolable protein complexes and intermediates. ELSCA's primary advantages in reproducing its training data are likely the consistency by which the data is generated and the limitless number of data points available, which permits distinctions between many more atom types. Again, this is an advantage only if the model ELSCA emulates is realistic.

Future Directions. We plan to extend ELSCA's PMFs to estimate the solute−solvent dispersion energy as derived by a continuum model based on volume integration while retaining the linear function of surface area for the work of cavity formation.[29,52] ELSCA is implemented in a developmental docking program, GRAPPLE, which is ultimately intended to emulate state-of-the-art binding energy calculation techniques[24,31,37] to dock proteins and determine transition paths in binding by computing the free-energy landscape of the interaction at high resolution.

Acknowledgment. David Cerutti thanks Drs. Stewart Adcock and Steven Bond for advice on numerical methods and Dr. Robert Konecny and the Center for Theoretical Biophysics (NSF grants MCB-0084797, PHY-0216576, and PHY-0225630, with additional support from NIH grant NBCR-RR08605i) for assistance with large-scale computing resources. This research was funded in part by NIH grant 2 T32 GM07240-27, the Achievement Rewards for Collegiate Scholars San Diego chapter, and grants from Howard Hughes Medical Institute, the National Science Foundation, the National Institutes of Health, the National Biomedical Computing Resource, the San Deigo Supercomputer Center, the W.M. Keck Foundation, and Accelrys, Inc.

Supporting Information Available: Coefficients for ELSCA to simulate protein encounters and notes on implementation of ELSCA-based potentials. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Angelini, T. E.; Liang, H.; Wriggers W.; Wong, G. C. Like-Charge Attraction Between Polyelectrolytes Induced by Counterion Charge Density Waves. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*(*15*), 8634−8637.

(2) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. Dissecting Subunit Interfaces in Homodimeric Proteins. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 708−719.

(3) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(4) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROM American Chemical Society: A Message-Passing Molecular Dynamics Implementation. *Comput. Phys. Commun*. **1995**, *91*, 43−45.

(5) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(6) Bogan, A. A.; Thorn, K. S. Anatomy of Hot Spots in Protein−Protein Interfaces. *J. Mol. Biol.* **1998**, *280*, 1−9.

(7) Bourne, Y.; Taylor, P.; Marchot, P. Acetylcholinesterase Inhibition by Fasciculin: Crystal Structure of the Complex. *Cell* **1995**, *83(3)*, 503−12.

(8) Bower, M. J.; Cohen, F. E.; Dunbrack, R. L. Prediction of Protein Side-chain Rotamers from a Backbone-dependent Rotamer Library: A New Homology Modeling Tool. *J. Mol. Biol.* **1997**, *267*, 1268−1282.

(9) Caflisch, A.; Karplus, M. Acid and Thermal Denaturation of Barnase Investigated by Molecular Dynamics Simulations. *J. Mol. Biol.* **1995**, *252*, 672−708.

(10) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. A Graph Theory Algorithm for Protein Side-Chain Prediction. *Protein Sci.* **2003**, *12*, 2001−2014.

(11) Cerutti, D. S.; Wong, C. F.; McCammon, J. A. Brownian Dynamics Simulations of Ion Atmospheres around Polyalanine and B-DNA: Effects of biomolecular dielectric. *Biopolymers* **2003**, *70(3)*, 391−402.

(12) Chen, R.; Weng, Z. A Novel Shape-Complementarity Scoring Function for Protein−Protein Docking. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 397−408.

(13) Chen, R.; Mintseris, J.; Janin, J.; Weng, Z. A Protein−Protein Docking Benchmark. *Proteins: Struct., Funct., Genet.* **2003**, *52(1)*, 88−91.

(14) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr..; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force field for the Simulation of Proteins, Nucleic Acids.; Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(15) Daura, X.; Oliva, B.; Querol, E.; Avilés, F. X.; Tapia, O. On the Sensitivity of MD Trajectories to Changes in Water−Protein Interaction Parameters: The Potato Carboxypeptidase Inhibitor in Water as a Test Case for the GROMOS Force Field. *Proteins: Struct., Funct. Genet.* **1996**, *25(1)*, 89−103.

(16) Davis, M. E.; McCammon, J. A. Dielectric Boundary Smoothing in Finite Difference Solutions of the Poisson Equation: An Approach to Improve Accuracy and Convergence. *J. Comput. Chem.* **1991**, *12(7)*, 909−912.

(17) Dominy, B. N.; Brooks, C. L., III Development of a Generalized Born Model Parameterization for Proteins and Nucleic Acids. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.

(18) Dong, F.; Vijayakumar, M.; Zhou, H. Comparison of Calculation and Experiment Indicates Significant Electrostatic Contributions to the Binding Stability of barnase and Barstar. *Biophys. J.* **2003**, *85*, 49−60.

(19) Dunbrack, R. L., Jr.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins: Application to Side-Chain Prediction. *J. Mol. Biol.* **1993**, *230*, 543−574.

(20) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. Computer Simulation of Protein−Protein Association Kinetics: Acetylcholinesterase-Fasciculin. *J. Mol. Biol.* **1999**, *291*, 149−162.

(21) Frisch, C.; Schreiber, G.; Johnson, C. M.; Fersht, A. R. Thermodynamics of the Interaction of barnase and Barstar: Changes in Free Energy Versus Changes in Enthalpy upon Mutation. *J. Mol. Biol.* **1997**, *267(3)*, 696−706.

(22) Gabdoulline, R. R.; Wade, R. C. Protein−Protein Association: Investigation of Factors Influencing Association Rates by Brownian Dynamics Simulations. *J. Mol. Biol.* **2001**, *306*, 1139−1155.

(23) Gabdoulline, R. R.; Wade, R. C. Effective Charges for Macromolecules in Solvent. *J. Phys. Chem.* **1996**, *100*, 3868−3878.

(24) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein−Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras-Raf and Ras-RalGDS Complexes. *J. Mol. Biol.* **2003**, *330*, 891−913.

(25) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein−Protein Docking with Simultaneous Optimization of Rigid-Body Displacement and Side-Chain Conformations. *J. Mol. Biol.* **2003**, *331(1)*, 281−299.

(26) Jiang, L.; Gao, Y.; Fenglou, M.; Liu, Z.; Lai, L. Potential of Mean Force for Protein−Protein Interaction Studies. *Proteins: Struct., Funct., Genet.* **2002**, *46(2)*, 190−196.

(27) Lazardis, T.; Karplus, M. Effective Energy Functions for Proteins in Solution. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133−152.

(28) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. New Analytic Approximation to the Standard Molecular Volume Definition and Is Application to Generalized Born Calculations. *J. Comput. Chem.* **2003**, *24(11)*, 1348−1356.

(29) Levy, R.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy. *J. Am. Chem. Soc.* **2003**, *125(31)*, 9523−9530.

(30) Lindahl, E.; Hess, B.; ans van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulations and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(31) Luo, H.; Sharp, K. On the Calculation of Absolute Macromolecular Binding Free Energies. *Proc. Natl. Acad. Sci.* **2002**, *99(16)*, 10399−10404.

(32) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **1995**, *91*, 57−95.

(33) Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovyi, V.; Mitchell, J. C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L. F. Protein Docking Using Continuum Electrostatics and Geometric Fit. *Protein Eng.* **2001**, *14(2)*, 105−113.

(34) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, California, 2000.

(35) Miyazawa, S.; Jernigan, R. L. Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term for Simulation and Threading. *J. Mol. Biol.* **1996**, *256*, 632−644.

(36) Miyazawa, S.; Jernigan, R. L. Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation. *Macromolecules* **1985**, *18*, 534−552.

(37) Noskov, S. Y.; Lim, C. Free Energy Decomposition of Protein−Protein Interactions. *Biophys. J.* **2001**, *81*, 737−750.

(38) Roux, B.; Simonson, T. Implicit Solvent Models. *Biophys. Chem.* **1999**, *78*, 1−20.

(39) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Fast and Robust Computation of Molecular Surfaces. *Proceedings of the 11th ACM Symposium on Computational Geometry* C6−C7, 1995.

(40) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein−Protein Interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153−159.

(41) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(42) Sitkoff, D.; Sharp, K. A.; Honig, B. Correlating Solvation Free Energies and Surface Tensions of Hydrocarbon Solutes. *Biophys. Chem.* **1994**, *51(3)*, 397−409.

(43) Tai, K.; Shen, T.; Henchman, R. H.; Bourne, Y.; Marchot, P.; McCammon, J. A. Mechanism of Acetylcholinesterase Inhibition by Fasciculin: A 5-ns Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2002**, *124(21)*, 6153−6161.

(44) van Buuren, A. R.; Marrink, S. J.; Berendsen, H. J. C. A Molecular Dynamics Simulation of the Decane/Water Interface. *J. Phys. Chem.* **1993**, *97(36)*, 9206−9212.

(45) van der Spoel, D.; van Buuren, A. R.; Tieleman, P.; Berendsen, H. J. C. Molecular Dynamics Simulations of Peptides from BPTI: A Closer Look at Amide-Aromatic Interactions. *J. Biomolecular NMR* **1996**, *8*, 229−238.

(46) van Gunsteren, W. F.; Berendsen, H. J. C. *Gromos-87 Manual*; Biomos BV: Groningen, The Netherlands, 1987.

(47) Wang, T.; Wade, R. C. Implicit-Solvent Models for Flexible Protein−Protein Docking by Molecular Dynamics Simulation. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 158−169.

(48) Wang, G.; Dunbrack, R. L. PISCES: a protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589−1591.

(49) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(50) Warshel, A.; Papazyan, A. Electrostatic Effects in Macromolecules: Fundamental Concepts and Practical Modeling. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211−217.

(51) Yu, B.; Blaber, M.; Gronenborn, A. M.; Clore, G. M.; Caspar, D. L. Disordered Water within a Hydrophobic Protein Cavity Visualized by X-ray Crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 103−108.

(52) 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.

(53) Zacharias, M. Protein−Protein Docking with a Reduced Model Accounting for Side-Chain Flexibility. *Protein Sci.* **2003**, *12*, 1271−1282.

(54) Zhang, C.; Vasmatzis, G.; Cornette, J. L.; DeLisi, C. Determination of Atomic Desolvation Energies from the Structures of Crystallized Proteins. *J. Mol. Biol.* **1997**, *267-(3)*, 707−725.

(55) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

CT049946F