# Evaluation of Models of Electrostatic Interactions in Proteins

**Alexandre V. Morozov,[†] Tanja Kortemme,[‡] and David Baker*[,‡]**

*Department of Physics, University of Washington, Box 351560, Seattle, Washington 98195-1560, and Department of Biochemistry, University of Washington, Box 357350, Seattle, Washington 98195-7350*

*Received: August 13, 2002; In Final Form: December 4, 2002*

The conformations of proteins and protein−protein complexes observed in nature must be low in free energy relative to alternative (not observed) conformations, and it is plausible (but not absolutely necessary) that the electrostatic free energies of experimentally observed conformations are also low relative to other conformations. Starting from this assumption, we evaluate alternative models of electrostatic interactions in proteins by comparing the electrostatic free energies of native, nativelike, and non-native structures. We observe that the total electrostatic free energy computed using the Poisson−Boltzmann (PB) equation or the generalized Born (GB) model exhibits free energy gaps that are comparable to, or smaller than, the free energy gaps resulting from Coulomb interactions alone. Detailed characterization of the contributions of different atom types to the total electrostatic free energy showed that, although for most atoms unfavorable solvation energies associated with atom burial are more than compensated by attractive Coulomb interactions, Coulomb interactions do not become more favorable with burial for certain backbone atom types, suggesting inaccuracies in the treatment of backbone electrostatics. Sizable free energy gaps are obtained using simple distance-dependent dielectric models, suggesting their usefulness in approximating the attenuation of long range Coulomb interactions by induced polarization effects. Hydrogen bonding interactions appear to be better modeled with an explicitly orientation-dependent hydrogen bonding potential than with any of the purely electrostatic models of hydrogen bonds, as there are larger free energy gaps with the former. Finally, a combined electrostatics−hydrogen bonding potential is developed that appears to better capture the free energy differences between native, nativelike, and non-native proteins and protein−protein complexes than electrostatic or hydrogen bonding models alone.

## 1. Introduction

Electrostatic effects play an important role in defining structural and functional aspects of biological macromolecules.[1−5] Therefore, there is a need to develop accurate models of electrostatic interactions, which capture the essential physics of the system while being analytically and computationally tractable. Computing electrostatic energies is a well-posed problem within the microscopic electrodynamics framework,[6] provided that charge distributions of all molecules in the system are available. These could in principle be obtained from the density matrix or from the ground-state wave function in the zero-temperature limit;[7] however, this calculation is beyond current ab initio computational approaches for biological macromolecules. Even with fixed atomic charges (i.e., neglecting induced dipoles), it is difficult to compute the electrostatic free energy of biological systems because both solute and solvent degrees of freedom have to be sampled explicitly.

Most current approaches to computing electrostatic free energies are based on the application of macroscopic electrodynamics to biological systems,[6,8] which reduces the number of degrees of freedom by treating the solvent as a continuous medium and by ignoring solute conformational changes. Because protein conformational changes and atomic polarizabilities are ignored, the interior of the protein is often treated as a dielectric with a dielectric constant greater that 1. However, biological macromolecules are too small to be characterized by the methods developed for bulk homogeneous matter. In particular, the notion of the dielectric constant becomes ambiguous;[9] it should be considered a parameter and not a constant with the same physical meaning as in bulk matter. Theoretical computations of dielectric constants inside proteins[10,11] reveal heterogeneous polar environments, which are not well reproduced by any single parameter. Moreover, any explicit solvent effects, such as water molecule penetration into protein interior,[12] are usually disregarded in continuum electrostatics.

Nonetheless, continuum approaches to the study of charged and polar molecules in aqueous solutions appear to be the best current methods for computing electrostatic free energies in proteins.[2,13] Continuum dielectric models describe both the free energy cost of desolvating polar atoms buried in the protein interior and the screening of Coulomb interactions arising from solvent polarization. The problem reduces to a numerical solution of the Poisson−Boltzmann (PB) equation,[6,8] with the system divided into solute (with low dielectric constant) and solvent (with high dielectric constant).[14−16] Solving the PB equation in this way has provided useful insights into the role of electrostatic interactions in proteins,[13] including deriving the Zimm−Bragg parameters for the helix−coil transition,[17] finding the degree of electrostatic optimization and charge complementarity in the barnase−barstar complex,[18,19] and computing electrostatic contributions to the stability of designed homeodomain variants.[20] Implicit solvation models based on the PB equation were also utilized as a part of the free energy function used in native structure discrimination on the EMBL set of

* To whom correspondence should be addressed.
[†] Department of Physics, University of Washington.
[‡] Department of Biochemistry, University of Washington.

10.1021/jp0267555 CCC: $25.00 
Published on Web 02/07/2003

deliberately misfolded proteins,[21,22] CASP3 and Park and Levitt protein models,[23] and ROSETTA protein models.[24]

Analytical approximations to the PB equation such as the generalized Born (GB) model are also widely used.[25−33] Within the GB approach, effective atomic Born radii are computed for each charged atom. For a simple spherical solute with a point charge located at its center, the Born radius is equal to the radius of the solute sphere (e.g., the van der Waals radius of a metal ion in water). For more complex solute shapes, the Born radius is a measure of average distance from the point charge to the solute−solvent dielectric boundary; it depends on the positions and volumes of all other solute atoms. The GB model is less demanding computationally than a numerical solution to the PB equation. Recently, the GB appproach has been used to calculate ligand−receptor binding energies.[34,35] In particular, Zhang et al.[35] found a fair agreement between protein−ligand solvation energies computed using implicit solvent models (both PB and GB) and explicit solvent simulations. GB models were also employed in nucleic acid molecular dynamics simulations, where they were found to reproduce results obtained via PB and explicit solvent approaches,[36,37] and in calculating electrostatic and solvation energies of large sets of misfolded protein conformations, including the Park and Levitt, CASP3, ROSETTA, and Skolnick data sets.[38,39]

Charge−charge interactions screened by solvent and solute polarization can also be modeled in a more heuristic way by introducing an effective distance-dependent dielectric into a simple Coulomb model of electrostatic interactions,[1,40,41] which progressively dampens long-range electrostatic forces. Such electrostatic energies are pairwise additive and offer a significant speedup over GB calculations. Solvation self-energies of individual charges are not considered in this approximation.

Hydrogen bonding (H-bonding) interactions form an especially important class of electrostatic phenomena in biological macromolecules:[42] they play a crucial role in the formation of protein secondary and tertiary structure. Physically, the interaction energy can be divided into classical (electrostatic and polarization) and quantum (exchange repulsion, charge-transfer, etc.) components. There is evidence to suggest that hydrogen bonding interactions are dominated by the electrostatic component, especially at distances >4−5 Å.[43] However, because of the observed directionality of hydrogen bond interactions,[44] it is unclear whether a simple model based, for example, on dipole−dipole interactions of hydrogen bonding groups should suffice to describe hydrogen bonds (H bonds) adequately.

It is a nontrivial problem to set up a rigorous computational test of alternative models of electrostatic interactions. A comprehensive test of electrostatic models is provided by considering a set of compact misfolded protein conformations (decoys) and assuming that the native structure has the lowest total free energy[45] and that, on average, some correlation exists between closeness to the native state on the free energy landscape and the free energy of near-native conformations for sufficiently relaxed structures. Although there are clear counterexamples to the latter assumption (for example small perturbations of the native structure can cause atoms to overlap, leading to very large energy increases), this property of folding free energy landscapes is consistent with many experimental protein folding data and is a central postulate of modern theories of protein folding (for example, the principle of minimal frustration[46]). The decoys used in electrostatic energy computations have to be numerous enough for adequate sampling and should comprise a variety of protein topologies and sizes. If the assumptions described above are correct, one would expect to find a gap in the total free energy while approaching a native state, so that nativelike conformations possess properties not shared by non-native decoys. One can then analyze separate free energy components and determine their contributions to the total free energy gap.

Recent studies[21,23,24,38,39,47−50] have examined the extent to which electrostatics calculations attribute low energies to native structures in sets of alternative conformations (decoys) for small proteins. Recognition of the native structure in sets of alternative conformations for protein−protein and protein−peptide complexes can also provide a useful test,[5,41] particularly since electrostatic effects have been shown experimentally to play an important role.[51] In both the monomeric protein and the protein−protein complex tests, it is also of interest to examine the extent to which conformations close to the correct structure have lower energies than quite non-native conformations (i.e., to what extent are there electrostatic "funnels" around native proteins and protein−protein complexes).

In this paper, we evaluate models of electrostatic interactions in biological macromolecules by testing them on a comprehensive set of decoy conformations for 41 single-domain proteins and 31 protein−protein complexes. Using this set, we compare different electrostatic models with one another by their ability to discriminate native from non-native conformations and nativelike conformations from more distant ones and draw general conclusions about underlying physics of solvation and charge−charge interactions in biological macromolecules. We also compare these models with an effective hydrogen bonding model, which by itself is capable of very good decoy discrimination.[52] We examine the extent to which unfavorable electrostatic desolvation energies for polar atoms are compensated by favorable Coulomb interactions with other polar atoms for the most commonly occurring atom types in proteins. Finally, we combine continuum electrostatics, hydrogen bonding, and van der Waals interactions into a simple physics-based potential exhibiting sizable free energy gaps.

## 2. Methods and Theory

**2.1. Continuum Dielectric Electrostatic Models.** *2.1.1. Poisson−Boltzmann Equation.* Once a molecule is represented as a solute cavity with charged atoms inside, surrounded by solvent, the problem of finding electrostatic energies is reduced to solving the Poisson−Boltzmann equation:[6,13]

$$\nabla(\epsilon(\vec{\mathbf{r}})\nabla\phi(\vec{\mathbf{r}})) - \epsilon(\vec{\mathbf{r}})\kappa^2(\vec{\mathbf{r}})\sinh\phi(\vec{\mathbf{r}}) = -(4\pi/kT)\rho(\vec{\mathbf{r}}) \quad (1)$$

where $\epsilon(\vec{\mathbf{r}})$ is the dielectric constant, $\phi(\vec{\mathbf{r}})$ is the dimensionless electrostatic potential (in units of $kT/e$, where $k$ is the Boltzmann constant, $T$ is the absolute temperature, and $e$ is the magnitude of the electron charge), $\rho(\vec{\mathbf{r}})$ is the free charge density (in units of $e$), and $\kappa^2(\vec{\mathbf{r}}) = (8\pi I)/(\epsilon(\vec{\mathbf{r}})kT)$ ($I = e^2c$ is the ionic strength of the bulk solution and $c$ is the ion concentration). Equation 1 is applicable to salt solutions of the same valence; it reduces to the Poisson equation when we neglect mobile ions in solvent. We used the *DelPhi II* macromolecular electrostatics modeling package to solve the Poisson−Boltzmann equation numerically, via a finite-difference method (see refs 2, 13, and 16 and references therein). We chose AMBER (PARM94) force field parameters[53] (partial charges and atomic radii) in the PB calculation, to be consistent with the parametrization of the GB model we used in this work.

Having found $\phi(\vec{\mathbf{r}})$, we can compute the total electrostatic energy of atomic charges inside the cavity using

$$E^{el} = \sum_i \frac{q_i \phi(\vec{r}_i)}{2} \quad (2)$$

where $\phi(\vec{r}_i)$ is the potential at the location of charge $q_i$, and the sum runs over all solute atoms. Note that direct charge−charge interactions (resulting in Coulomb's law) are included in (2).

*2.1.2. Generalized Born Model.* The generalized Born (GB) model of continuum electrostatics[25] is capable of reproducing the results obtained through the solution of the Poisson−Boltzmann (PB) equation with high accuracy and at a smaller computational cost. This is essential if structural analysis is to involve extensive data sets. Also, different terms in the GB model can be assigned transparent physical interpretations and analyzed separately. We adopt in our calculations the pairwise solute descreening approach to computing atomic Born radii[54,55] and use the AMBER (PARM94) force field parametrization of the GB model.[30,31] Alternative GB model parametrizations consistent with CHARMM all hydrogen and polar hydrogen force fields[29] and with the OPLS force field[27] have also been described in the literature.

The basic GB formula for electrostatic energy is given as follows:

$$E^{el} = \frac{1}{2}\sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon_i r_{ij}} - \frac{\tau}{2}\sum_i \sum_j \frac{q_i q_j}{f^m_{GB}} \quad (3)$$

Here, $\tau = 1/\epsilon_i - 1/\epsilon_s$, $\epsilon_{i(s)}$ is the solute (solvent) dielectric constant, and the modified GB function is given by[30]

$$f^m_{GB} = f_{GB}\frac{\epsilon_s \gamma - \gamma}{\epsilon_s \gamma - 1} \quad (4)$$

where the GB function is

$$f_{GB} = \sqrt{(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/2b_i b_j])}$$

Here, $r_{ij}$ are interatomic distances, $b_i$ are atomic Born radii, and all sums above run over solute atoms. An empirical parameter $\gamma(r_{ij}, b_i, b_j)$ was introduced[30] to improve the correlation between finite-difference PB and GB energies on a test set of small molecules. This parametrization of the GB model is based on the AMBER force field partial charges and van der Waals radii.[53] We computed atomic Born radii using the pairwise solute descreening approach developed in refs 54,55. The first term on the right-hand side of eq 3 gives the Coulomb energy; the $i = j$ contribution to the second term on the right-hand side of eq 3 yields atomic solvation self-energies, whereas the $i \neq j$ contribution describes interatomic screening of solute atoms by solvent polarization. The screened Coulomb energy is given by the sum of the Coulomb and the screening term, the total solvation energy is given by the sum of the self-energy and the screening term, and the GB electrostatic energy is given by the sum of the screened Coulomb energy and the self-energy.

In all GB calculations carried out in the rest of the paper, we reset interatomic distances of atom pairs that are too close to each other:

$$r \Rightarrow d_i + d_j \text{ if } r < d_i + d_j$$

where $d_{i(j)}$ is the van der Waals radius of atom $i(j)$. This helps alleviate unphysical situations in which atomic overlaps occur in our data sets.

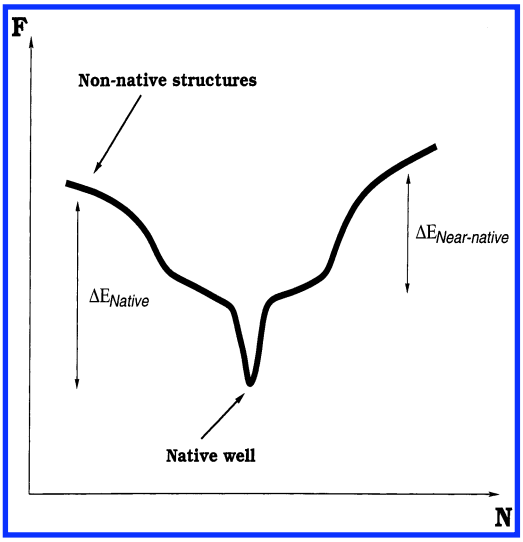


**Figure 1.** Schematic picture of a 1D free energy (F) folding landscape (*N* is an arbitrary reaction coordinate). The native structure resides in the native well, with low RMSD decoys occupying low energy states in the nativelike well (the folding funnel). More distant non-native conformations have higher free energies. $\Delta E_{Native}$ is the native free energy gap, $\Delta E_{Near-native}$ is the nativelike free energy gap.

For interactions between atom pairs less than the persistence length apart in the chemical sequence, bond stretching and bending may partially offset long range forces. Because atoms close in the linear sequence are likely to also be close in the 3D structure, the contribution of such interactions to the electrostatic free energy can be sizable. We tested a few schemes of atom exclusion for our electrostatics calculations, pinpointing the distance along the chemical sequence at which short-range bonded interactions can be neglected. These included accounting only for atoms separated by at least three other atoms along the chemical sequence; excluding all interactions within the same residue and the neighboring mainchain atoms on both sides; excluding all interactions within the same residue and with all atoms in the adjacent residues. We found the first and second scheme to be similarly optimal choices and use the second scheme when computing GB/effective dielectric energies below (in *DelPhi II*, all atom pairs are included by default; we sum over all atoms when directly comparing PB and GB free energies in Figure 2).

*2.1.3. Distance-Dependent Dielectric Models.* We test three different distance-dependent dielectric models: the Warshel exponential model,[1] the Sternberg pseudo-sigmoidal model,[41] and a linear distance-dependent dielectric model.[40] The Warshel model is given by the following expression:

$$\epsilon_i(r) = \begin{cases} 16.55, & r < 3 \text{ Å} \\ 1 + 60(1 - \exp(-0.1r)) & r \geq 3 \text{ Å} \end{cases}$$

Here and below, $r$ denotes interatomic distances. The value of $\epsilon_i$ in the smaller range is chosen to make the dielectric function continuous.

The Sternberg dielectric model is defined by

$$\epsilon_i(r) = \begin{cases} 4, & r \leq 6 \text{ Å} \\ 38r - 224, & 6 \text{ Å} < r < 8 \text{ Å} \\ 80, & r \geq 8 \text{ Å} \end{cases}$$

This function offers a smooth switchover from the short-distance value of 4 to the long-distance dielectric constant equal to that of bulk water.

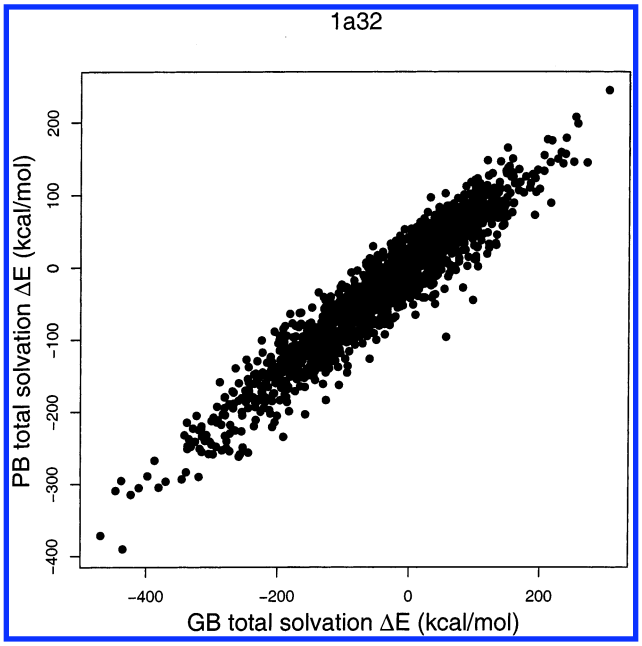


**Figure 2.** Poisson−Boltzmann total solvation energy vs generalized Born total solvation energy for *1a32* decoys (in kcal/mol). All atom−atom pairs are included; the energies are computed relative to the native structure.

Finally, the linear model is given by

$$\epsilon_i(r) = 6r, \forall r$$

Throughout this paper, we use the terms "energy" and "free energy" interchangeably.

**2.2. Hydrogen Bonding Potential.** The energy of a hydrogen bond was taken to be a linear combination of three geometry-dependent energy terms:

$$E^{\mathrm{HB}} = W_\delta E(\delta_{\mathrm{HA}}) + W_\Theta E(\Theta_{\mathrm{H}}) + W_\psi E(\psi_{\mathrm{A}}) \qquad (5)$$

where $E(\delta_{\mathrm{HA}})$ depends on the hydrogen−acceptor distance, $E(\Theta_{\mathrm{H}})$ depends on the angle at the hydrogen atom (donor−H···acceptor) and $E(\psi_{\mathrm{A}})$ depends on the angle at the acceptor atom (H···acceptor−acceptor base). The distance dependence was modeled as a 10−12 potential with an ideal hydrogen−acceptor distance of 1.9 Å. The energy functions $E(\Theta_{\mathrm{H}})$ and $E(\psi_{\mathrm{A}})$ were derived from the logarithm of the probability distributions found in high-resolution crystal structures as described in ref 56. For the dependence on the acceptor angle $\psi_{\mathrm{A}}$, separate statistics were collected for $sp^2$ and $sp^3$ hybridized acceptor atoms to take into account a potentially different electron distribution around the acceptor atom. Because of their divergent geometrical preferences, different statistics were collected for side chain−side chain and mainchain−mainchain hydrogen bonds. The relative weights of the three different energy terms ($W_\delta$, $W_\Theta$, and $W_\psi$) were parametrized as described in refs 56 and 57 to reproduce native sequences of monomeric proteins and were 1.0, 1.03, and 0.2, respectively.

Calculation of hydrogen bond energies in the fashion described above requires explicit placement of polar hydrogen atoms. Polar hydrogens were added in cases where the position of the hydrogen atom was defined by the chemistry of the donor group (backbone amide protons, tryptophan indol, asparagines and glutamine amide groups, and arginine guanido protons). Standard bond lengths and angles were taken from the CHARMM19 force field.[58] Polar hydrogens with variable positions (serine, threonine, and tyrosine hydroxyl groups; the lysine amino group was not rotated) as well as flips of the amide groups of asparagine and glutamine residues and different proton positions of the histidine imidazole groups (assumed to be neutral in all cases) were modeled as rotamers and optimized using a Monte Carlo (MC) simulated annealing procedure with an energy function mainly consistent of a 6−12 Lennard-Jones potential, an effective solvation potential,[49] as well as the hydrogen bonding term described above[56] (only hydrogen bonds with proton positions given by the chemistry of the donor group were considered in the derivation of the energy parameters of the potential).

**2.3. van der Waals Interactions and Cavity Free Energy.** We use a standard 6−12 Lennard-Jones potential[42] with modifications at small and large distances.[57] The van der Waals energy of atoms $i$ and $j$ is given by (in kcal/mol)

$$E_{ij}^{\mathrm{VdW}} = \begin{cases} 10.0(1 - r/(0.89r_{ij})), & r < 0.89r_{ij} \\ -A_{ij}/r^6 + B_{ij}/r^{12} & 0.89r_{ij} \le r < 8.0\ \text{Å} \\ 0.0 & r \ge 8.0\ \text{Å} \end{cases}$$

Here, $r$ is the interatomic distance, $r_{ij} = 0.95(d_i + d_j)$, and $A_{ij}$, $B_{ij}$ are empirical coefficients. The linear ramp-up to 10 kcal/mol at small distances and multiplication of the atomic radii by 0.95 help reduce the height of local van der Waals maxima on the free energy landscape. The cutoff at 8 Å improves computational efficiency of the van der Waals calculations.

The total solvation free energy includes, in addition to the electrostatic contribution, the free energy cost of cavity formation in the solvent and solute−solvent van der Waals interactions. Both of these terms are roughly proportional to the cavity surface area, and can be approximated by $\sum_k \sigma_k SA_k$, where $SA_k$ is the total solvent-accessible surface area of atom type $k$ and $\sigma_k$ is the corresponding empirical solvation parameter.[21,25,28] The set of empirical solvation parameters is the same as in ref 28: $\sigma_{C(sp^3),S} = 10$ cal/(mol Å$^2$), $\sigma_{C(sp^2),C(sp)} = 7$ cal/(mol Å$^2$), $\sigma_{O,N,H} = 0$ cal/(mol Å$^2$). We call this term the "surface area" energy later in the paper.

Combined free energies including hydrogen bonding, van der Waals, and electrostatic interactions were obtained by generating a generalized linear model (GLM) fit via a logistic regression function implemented in the R statistical software package.

**2.4. Description of Data Sets.** If general conclusions about the physical nature of electrostatic interactions in biological macromolecules are to be made, the test set used for model evaluation should be free as much as possible from systematic biases. Protein test sets should be diverse and extensive enough to reproduce a variety of intramolecule, intermolecule, and solute−solvent interactions occurring in nature.

In accordance with this approach, we use two distinct data sets in our analysis (Table 1). The first data set consists of 41 small (less than 90 amino acids) single-domain proteins, for each of which ∼2000 decoys were produced using the ROSETTA method for ab initio protein structure prediction.[59,60] The decoys were generated using a simulated annealing procedure with a protein database derived free energy function using side chains represented as centroids. A subset of low energy decoys was then relaxed, i.e., subjected to a refinement protocol coupling torsion angle move sets and an all atom-based free energy function, dominated by van der Waals interactions.[61] Finally, all side chains were repacked using an MC rotamer-substitution protocol.[52,57] This decoy set is subdivided into two: 25 proteins where high-resolution native structures determined experimentally via X-ray diffraction were available and 23 proteins for which ROSETTA could produce sufficiently

**TABLE 1: 5% RMSD Cutoffs in Å for (from left to right) Single-Domain Decoy Sets Used in Low RMSD *Z* score Calculations and the Protein−Protein Complex Decoy Set[a]**

| monomeric proteins | | | | protein−protein complexes | | |
|---|---|---|---|---|---|---|
| | | 5% RMSD | | | | |
| PDB | SS | −PN | +PN | PDB | ID tag | 5% RMSD |
| 1a32 | α | 1.55 | 1.52 | 1a2y | ab | 2.70 |
| 1am3 | α | 2.09 | 2.06 | 1cz8 | ab | 1.77 |
| 1bw6 | α | 2.68 | 2.71 | 1dqj | ab | 1.83 |
| 1gab | α | 2.22 | 2.24 | 1e6j | ab | 1.24 |
| 1kjs | α | 3.67 | 3.68 | 1egj | ab | 2.43 |
| 1mzm | α | 3.51 | 2.02 | 1eo8 | ab | 2.81 |
| 1nkl | α | 3.57 | 2.67 | 1fdl | ab | 2.65 |
| 1nre | α | 2.72 | 2.31 | 1fj1 | ab | 1.10 |
| 1pou | α | 3.58 | 3.34 | 1g7h | ab | 2.67 |
| 1r69 | α | 1.89 | 1.68 | 1ic4 | ab | 2.13 |
| 1res | α | 1.38 | 1.39 | 1jhl | ab | 2.33 |
| 1uba | α | 3.81 | 3.84 | 1jrh | ab | 1.13 |
| 1uxd | α | 1.34 | 1.36 | 1mlc | ab | 0.83 |
| 2ezh | α | 3.46 | 3.30 | 1nca | ab | 0.97 |
| 2pdd | α | 2.88 | 2.90 | 1nsn | ab | 2.34 |
| 1aa3 | αβ | 3.43 | 3.42 | 1osp | ab | 2.84 |
| 1afi | αβ | 3.23 | 1.96 | 1qfu | ab | 1.30 |
| 1ctf | αβ | 3.60 | 1.28 | 1wej | nab | 2.57 |
| 1pgx | αβ | 2.74 | 1.16 | 1ACB | nab | 2.15 |
| 2fow | αβ | 3.76 | 3.25 | 1AVZ | nab | 1.96 |
| 2ptl | αβ | 2.92 | 2.18 | 1brs | nab | 2.60 |
| 1sro | β | 3.72 | 2.04 | 1CHO | nab | 2.35 |
| 1vif | β | 1.49 | 1.25 | 1MDA | nab | 1.92 |
| mean | | 2.84 | 2.33 | 1PPF | nab | 2.07 |
| | | | | 1SPB | nab | 1.93 |
| | | | | 1UGH | nab | 1.51 |
| | | | | 2PCC | nab | 2.31 |
| | | | | 2PTC | nab | 1.58 |
| | | | | 1CSE | nab | 1.96 |
| | | | | 1FIN | nab | 1.36 |
| | | | | 2BTF | nab | 2.12 |
| | | | | mean | nab | 1.98 |

[a] −PN subcolumn, ab initio single-domain decoy set; +PN subcolumn, ab initio single-domain decoy set enhanced with perturbed-native structures. SS, protein secondary structure assignment (α helix, β strand, or both); ID tag, antibody−antigen complex (ab) or nonantibody complex (nab).

many nativelike decoys [determined by $RMSD_{10\%} \leq 4$ Å, where $RMSD_{10\%}$ is the 10% RMSD cutoff of the resulting decoy distribution (RMSD is the root-mean-square deviation of decoy backbone $C_\alpha$ coordinates from those in the native structure)]. Note that some structures are present in both subsets. The former subset is used in analyzing energy gaps between native structures and decoys; the latter is used between nativelike (low RMSD) and non-native decoys. Additionally, to study the properties of conformations in the native funnel, 300 additional nativelike decoys were created for each structure in the low RMSD subset, starting from the native conformation (perturbed-native decoys), and using the ROSETTA method. Each of these decoys was relaxed and repacked with the same protocol as in the main set. When these extra structures were added to the main low RMSD decoy subset, the average 5% RMSD cutoff (which defines low RMSD decoys, see section 2.5) decreased from 2.84 to 2.33 Å (Table 1). Both the ab initio set and the set enhanced with perturbed native structures are used in the paper.

Our second data set consists of 31 docked protein−protein complexes, with ∼2000 decoys made for each. This set is especially interesting because charged and polar interactions are thought to play an important role in protein−protein association.[5] The set is divided into 18 antibody−antigen complexes and 13 nonantibody (mostly enzyme−inhibitor) complexes, because these two types exhibit consistent differences in terms of the amino acid composition.[62] The decoys are produced by first repacking side chains of the two protein docking partners separately, followed by random-orientation rigid body docking and subsequent minimization using a centroid-based side chain representation, and finally by minimizing the free energy using a side chain repacking all-atom protocol.[52,57,63] Protein backbone conformations stay fixed throughout this procedure. The average 5% RMSD cutoff is 1.98 Å for this decoy set (Table 1).

**2.5. Analysis of Energy Gaps.** For all free energies to be analyzed in the subsequent sections, we use the normalized energy gaps, or *Z* scores as our figures of merit. *Z*-score analysis is a standard way to quantify the signal-to-noise ratio on a data set.[38,64] We use three different *Z*-score measures, defined as follows:

$$Z_{\mathrm{ref}} = \frac{\langle E \rangle - E_{\mathrm{ref}}}{\sigma_E} \tag{6}$$

where $\langle E \rangle = 1/N\sum_{i=1}^{N} E_i$ is an average energy of $N$ decoys

$$\sigma_E{}^2 = \frac{1}{N}\sum_{i=1}^{N}(E_i - \langle E \rangle)^2$$

is the standard deviation of decoy energies, and $E_{\mathrm{ref}}$ is the reference energy which is either $E_{\mathrm{nat}}$ − energy of the native structure obtained through X-ray diffraction or NMR experiments, or $E_{\mathrm{nat\_rep}}$ − energy of the structure with the native backbone but all side chains repacked using the MC rotamer-substitution protocol.[52,57] We will refer to these *Z* scores as the native and native-repacked *Z* scores, respectively. The latter is a more unbiased measure, because all native and decoy side chains have been repacked using the same MC protocol. Finally, the low RMSD (or nativelike) *Z* score is defined as

$$Z_{\mathrm{low\_RMSD}} = \frac{\langle E \rangle_{\mathrm{hi}} - \langle E \rangle_{\mathrm{lo}}}{\sigma_E^{\mathrm{hi}}} \tag{7}$$

where the sums in the averages and the standard deviation run over high RMSD and low RMSD decoys separately. By definition, the low RMSD decoys comprise the lowest 5% of the RMSD distribution. Note that the *Z* scores are invariant with respect to the energy scale. We say that we fail to discriminate a particular structure if $Z < 1$ for its decoy set, where *Z* denotes any of the *Z* scores defined above.

Finally, we note that if two individual energies $E_1$ and $E_2$ are known for a decoy set the *Z* score for their linear combination $E = aE_1 + bE_2$ is given by

$$Z_E = \frac{\sigma_{E_1} Z_{E_1} + (b/a)\sigma_{E_2} Z_{E_2}}{\sigma_E} \tag{8}$$

where

$$\sigma_E{}^2 = \sigma_{E_1}{}^2 + (b/a)^2 \sigma_{E_2}{}^2 + 2(b/a)\mathrm{Var}(E_1, E_2)$$

Here, the cross-correlation term is

$$\mathrm{Var}(E_1, E_2) = \langle E_1 E_2 \rangle - \langle E_1 \rangle \langle E_2 \rangle$$

This procedure can be easily extended to a linear combination of three or more scores.

We can use (8) to find the effect of changing the dielectric constant inside the solute cavity. In particular, if we have a set of electrostatic energies computed at some reference value $\epsilon_i^{\mathrm{ref}}$,

we can compute $Z$ scores at a new value $\epsilon_i^{new}$ by simply setting

$$b/a = (\epsilon_s - \epsilon_i^{new})/(\epsilon_s - \epsilon_i^{ref})$$

in (8). Here, $E_1$ is the Coulomb energy, and $E_2$ is the solute−solvent screening term.

## 3. Results and Discussion

In this section, we discuss various electrostatics models and compare their ability to differentiate native and nativelike structures from arbitrary compact decoys. The best model may be capturing the essential physics of solvation and charge−charge interactions better than other, less sensitive approaches. In Figure 1, we show a schematic picture of a 1D free energy landscape with both native and nativelike energy gaps. Native and native-repacked $Z$ scores (energy gaps normalized by standard deviations, see Methods and Theory) assess the depth of the native well, whereas low RMSD $Z$ scores reflect the energy difference between near-native and more distant structures.

**3.1. Poisson−Boltzmann Calculations.** We find the Poisson−Boltzmann (PB) electrostatic energies by solving the PB equation for every structure in our decoy sets. We ignore the dependence of electrostatic energies on the ionic strength by setting the salt concentration to zero in all calculations reported in Table 2a,b; this facilitates comparison with simplified electrostatics models, which are unable to account for the ionic strength explicitly (with the exception of the GB approach extended to low salt concentrations in ref 65). The Debye screening length is ∼1 nm at 0.1 M NaCl, and electrostatic energies are generally dominated by short and medium distance interactions; we did not observe any significant changes in the conclusions described below when the PB calculations were repeated with a salt concentration of 0.1 M (data not shown).

We obtain PB total solvation energies by performing $\epsilon_s = 1$ and $\epsilon_s = 80$ calculations with $\epsilon_i = 1$ ($\epsilon_s$ is the solvent dielectric constant, and $\epsilon_i$ is the dielectric constant within the cavity) for each protein and subtracting the results. PB total solvation energies include both desolvation self-energies and the charge−charge screening induced by solvent polarization. The PB electrostatic energy (cf $Z$ scores in the PB column of Table 2a,b; Table 4) is a sum of the total solvation energy and the Coulomb interactions.

The PB electrostatic energy of native structures is not always lower than that of the misfolded structures; while the Coulomb term favors the native structure, the total solvation energy in many cases actually disfavors the native structure. This solvation energy behavior is expected because native conformations are usually better packed than decoys and therefore incur larger penalties for charged atom burial; indeed, repacking and relaxing of native structures makes them more expanded and eliminates the solvation energy penalty relative to decoys (data not shown; see also refs 23, 38, 39, and 47, where all decoys and native structures were minimized with the same protocol prior to electrostatic calculations). This is also evident from differences between native and native repacked PB $Z$ scores; even though PB solvation energies are still anticorrelated on average, they add up with the Coulomb energies to produce consistently higher $Z$ scores in the native repacked case (but not much higher than Coulomb $Z$ scores alone).

Different sets of atomic radii defining the solute−solvent dielectric boundary have been used in PB calculations,[21,24,29,37] reflecting the uncertainty inherent in the continuum electrostatic models. For example, using PARSE[66] rather than AMBER-(PARM94) radii to define the dielectric boundary would lead to even more favorable decoy solvation energies, because PARSE radii are smaller on average. Placing the dielectric boundary closer to atom sites would affect exposed atoms more significantly than buried ones, lowering their energies because of stronger polarization. This effect would lower decoy solvation energies more than energies of the native structures, because decoys have more atoms exposed to solvent.

**3.2. Generalized Born Calculations.** The GB model was developed as an analytical approximation to the exact solution of the Poisson equation. As such, it is computationally less demanding than the finite-difference PB methods. Moreover, different terms in the GB expression have straightforward physical interpretation. We use $\epsilon_i = 1$ in all GB calculations unless explicitly indicated otherwise.

There is a high degree of correlation between total solvation energies computed using PB and GB approaches,[29,36,37] as shown in Figure 2 for *1a32* decoys (*1a32* is the Protein Data Bank code). Consequently, the GB electrostatic energy, like the PB electrostatic energy, does not exhibit large native and native repacked energy gaps (PB,GB columns of Table 2a,b; Table 4). The best discriminators of native and nativelike structures are Coulomb interactions screened by polarization on the solvent−solute boundary (Screened Coul column of Table 2a,b; Table 4), and constant dielectric Coulomb interactions (Coul column of Table 2a,b; Table 4). To compute GB electrostatic energies, we add solvation self-energies to the screened Coulomb interactions; however, the self-energies usually disfavor native and native repacked structures compared to decoys (Self-Energy column in Table 2a,b; Table 4), and the GB electrostatic energy gaps become considerably smaller.

Total solvation energies are known to be anticorrelated with Coulomb energies,[17,21−23,38] as shown in Figure 3 for *1a32* decoys using the GB model. Therefore, the presence of the gap in the GB electrostatic energy depends on the delicate cancellation of large terms with opposite signs; even a minor error in electrostatic energies might lead to substantial deviations in energy gaps. As Figure 3 shows, solvation penalties of buried atoms are roughly compensated by additional Coulomb interactions they make; atoms exposed to solvent have favorable solvation energies but interact with fewer solute atoms, and vice versa. In the first row of Figure 4a, we show decoy atomic energies, computed relative to native atomic energies: $E_{dec} - E_{nat}$, as a function of the solvent-accessible surface area in the native structure. The energies considered are the self-energies, the screened Coulomb energies, and the total GB electrostatic energies. The self-energy is more negative in decoys by −0.2 kcal/mol per atom, whereas the average screened Coulomb energy is more negative in native structures by 0.2 kcal/mol per atom. The energy gap practically disappears when these two terms are added up to yield the total GB electrostatic energy (the average is −0.003 kcal/mol per atom).

The cancellation between solvation and Coulomb terms is particularly evident for atoms with significant differences in solvation and Coulomb energies in the native structure compared to decoys. In the plots in the second row of Figure 4a, blue triangles designate atoms whose energies are lower in decoys relative to native structures by a certain threshold amount (≥2 kcal/mol for self-energies and total GB energies; ≥5 kcal/mol for screened Coulomb energies), red circles indicate atoms for which decoys have significantly less favorable energies than native structures, and the open green circles are all other atoms for which the energies do not change much. Atoms which are more exposed in the decoy structures (above the diagonal) have

**TABLE 2: Native (Zn) and Native Repacked (Znr) *Z* scores for a Set of 10 α, 9 αβ, and 6 β Single-domain Proteins (Section a) and for a Set of 18 Antibody−Antigen (ab) and 13 Nonantibody (nab) Protein−Protein Complexes (Section b)[a]**

Section a

| | | PB | | GB | | Coul | | self energy | | screened Coul | | surface area | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB | SS | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr |
| 1a32 | α | −0.35 | 0.99 | −0.43 | 0.56 | 1.01 | 0.57 | 1.46 | 1.43 | 1.67 | 1.28 | 0.12 | −0.21 |
| 1ail | α | 0.93 | 2.13 | 0.54 | 2.18 | 2.80 | 1.60 | 0.55 | 0.67 | 2.37 | 2.49 | −0.45 | 0.04 |
| 1am3 | α | −1.43 | −0.35 | −2.32 | −1.03 | −0.68 | −0.98 | 1.07 | 0.13 | 0.77 | 0.51 | 0.10 | −0.64 |
| 1cc5 | α | −2.93 | −0.55 | −4.17 | −1.54 | 1.27 | 1.01 | −1.27 | −0.65 | −2.16 | −1.97 | 1.90 | 1.60 |
| 1cei | α | −0.30 | 1.65 | −1.34 | 0.81 | 2.92 | 0.94 | −0.27 | −0.26 | 2.92 | 1.39 | 2.28 | 2.53 |
| 1hyp | α | −0.34 | 0.50 | −0.87 | −0.29 | −0.10 | −0.49 | 2.06 | 0.68 | 1.76 | 1.41 | 0.94 | 0.61 |
| 1lfb | α | −0.08 | 0.93 | 1.32 | 1.75 | 1.59 | 1.89 | 0.65 | −0.50 | 2.59 | 2.72 | 0.82 | 1.43 |
| 1mzm | α | 1.83 | 1.54 | 0.07 | 0.76 | 1.03 | 1.36 | 2.21 | 1.89 | 1.17 | 1.92 | −0.05 | −0.63 |
| 1r69 | α | 1.04 | −0.12 | 0.27 | −1.22 | 2.80 | 0.87 | 0.12 | −0.77 | 2.32 | 0.27 | 2.26 | 2.61 |
| 1utg | α | −1.44 | 0.76 | −2.45 | −0.39 | 3.05 | 1.90 | 0.75 | 0.16 | 2.72 | 1.54 | −0.83 | −0.60 |
| 1ctf | αβ | −0.14 | −0.09 | −1.19 | 0.92 | 2.32 | 1.44 | 0.55 | 0.08 | 2.46 | 2.52 | 2.83 | 2.02 |
| 1dol | αβ | 1.05 | 1.45 | −0.29 | 0.23 | 0.59 | 0.34 | 2.32 | 1.53 | 1.46 | 1.77 | 2.57 | 2.31 |
| 1orc | αβ | 2.07 | 3.40 | 1.36 | 2.11 | 0.85 | 0.11 | 1.08 | 0.50 | 3.20 | 2.56 | 0.12 | −0.78 |
| 1pgx | αβ | 0.84 | 2.97 | −0.09 | 1.09 | 3.85 | 0.93 | −0.51 | −0.71 | 3.08 | 0.98 | 2.41 | 1.57 |
| 1ptq | αβ | −0.20 | −0.02 | −2.11 | −2.08 | −0.62 | −0.61 | 0.88 | −0.83 | 0.76 | 0.52 | 1.57 | −0.02 |
| 1tif | αβ | 1.77 | 2.83 | 0.82 | 1.79 | 1.57 | 1.10 | 1.24 | 0.23 | 2.73 | 2.66 | 2.76 | 1.68 |
| 1vcc | αβ | 0.75 | 1.28 | −1.42 | −0.28 | 2.38 | 2.17 | −0.28 | −0.26 | 1.63 | 1.52 | 2.89 | 2.38 |
| 2fxb | αβ | −3.48 | −1.34 | −2.89 | 0.25 | 1.62 | 1.01 | −1.84 | −2.47 | −0.09 | 0.45 | 3.98 | 3.69 |
| 5icb | αβ | −3.15 | −0.99 | −2.55 | 0.29 | −1.17 | −2.40 | −0.10 | 0.21 | 1.78 | 1.40 | 1.53 | 0.89 |
| 1bq9 | β | −4.86 | 1.83 | −5.35 | 2.07 | 3.09 | 2.53 | −0.88 | −1.09 | 3.13 | 2.95 | 2.56 | 2.18 |
| 1csp | β | 1.40 | 3.81 | 0.39 | 2.75 | 1.45 | 1.19 | 0.05 | −1.05 | 1.17 | 1.32 | 2.26 | 2.08 |
| 1msi | β | 2.29 | 0.87 | −0.98 | −1.75 | 1.88 | 0.50 | 0.56 | 0.50 | 2.17 | 0.89 | 2.49 | 2.48 |
| 1tuc | β | 0.00 | 2.34 | −1.52 | 0.04 | 1.82 | 1.16 | 1.15 | 0.50 | 1.73 | 1.57 | 2.41 | 1.78 |
| 1vif | β | 1.79 | 2.29 | 1.84 | 2.81 | 3.22 | 3.21 | −0.49 | −1.08 | 2.46 | 2.46 | 1.38 | 1.19 |
| 5pti | β | 1.72 | 1.80 | 1.34 | 1.27 | 1.24 | 1.20 | 0.42 | −0.90 | 2.32 | 1.44 | 2.75 | 2.15 |
| mean | | −0.05 | 1.20 | −0.88 | 0.52 | 1.59 | 0.90 | 0.46 | −0.08 | 1.85 | 1.46 | 1.66 | 1.29 |
| stdev | | 1.91 | 1.36 | 1.79 | 1.39 | 1.31 | 1.17 | 1.04 | 0.97 | 1.18 | 1.05 | 1.25 | 1.25 |

Section b

| | | PB | | GB | | Coul | | self energy | | screened Coul | | surface area | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB | ID tag | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr |
| 1a2y | ab | 0.20 | 0.25 | 0.40 | 1.07 | 2.74 | 1.59 | −0.05 | 0.43 | 1.41 | 0.99 | 0.82 | 0.33 |
| 1cz8 | ab | −1.95 | 0.33 | −0.45 | 1.89 | 5.61 | 1.37 | −1.76 | −0.55 | 2.80 | 0.92 | 0.88 | 1.54 |
| 1dqj | ab | −1.76 | −0.79 | 0.36 | 0.63 | 3.71 | 1.51 | −1.24 | −0.72 | 1.02 | 0.61 | 1.08 | 0.72 |
| 1e6j | ab | −3.84 | 1.65 | −3.14 | 1.12 | 7.41 | 1.19 | −1.41 | −0.29 | 2.72 | 0.62 | 2.10 | 1.19 |
| 1egj | ab | −2.05 | 0.13 | −1.98 | 0.56 | 4.23 | 1.58 | −1.10 | −0.14 | 0.40 | 0.56 | 0.72 | 0.59 |
| 1eo8 | ab | −1.44 | 0.93 | −0.22 | 1.10 | 9.83 | 2.22 | −3.77 | −0.28 | 3.25 | 0.88 | 2.93 | 1.99 |
| 1fdl | ab | −0.25 | −0.04 | 0.98 | 0.79 | 2.85 | 0.61 | 0.10 | 0.29 | 1.38 | 0.74 | 1.24 | 1.08 |
| 1fj1 | ab | −9.22 | −0.11 | −8.75 | 0.22 | 3.54 | 0.83 | −3.90 | −0.79 | −1.55 | 0.03 | 1.91 | 1.84 |
| 1g7h | ab | −0.65 | 0.27 | 0.59 | 1.07 | 2.22 | 0.50 | −0.53 | 0.41 | 1.62 | 0.90 | 0.51 | 0.64 |
| 1ic4 | ab | −0.90 | 0.25 | −0.09 | 0.97 | 4.07 | 2.91 | −1.27 | −0.62 | 1.79 | 0.96 | 0.72 | 0.73 |
| 1jhl | ab | −0.44 | 0.77 | 0.22 | 1.30 | 0.16 | 1.68 | 0.62 | 0.56 | 0.45 | 1.34 | 0.84 | −0.14 |
| 1jrh | ab | 0.72 | −0.20 | 0.97 | 0.51 | 2.58 | 1.70 | 0.24 | −0.25 | 1.75 | 0.78 | 0.39 | 0.40 |
| 1mlc | ab | −0.12 | 0.99 | 0.00 | 0.95 | 1.41 | 0.97 | −0.72 | −0.05 | 1.45 | 0.67 | 1.89 | 1.07 |
| 1nca | ab | 1.13 | 0.44 | 1.41 | 0.78 | 6.32 | 2.70 | −1.30 | −0.72 | 1.81 | 0.60 | 1.33 | 1.00 |
| 1nsn | ab | −3.00 | 0.36 | −2.21 | 0.27 | 4.85 | 0.64 | −1.28 | 0.19 | 0.95 | 0.40 | −0.22 | 0.55 |
| 1osp | ab | −4.71 | −0.40 | −4.23 | 0.37 | 2.86 | 0.61 | −0.80 | −0.33 | 0.49 | 0.41 | 2.18 | 0.93 |
| 1qfu | ab | 0.67 | 0.73 | −0.90 | 0.69 | 8.18 | 1.99 | −2.71 | −0.53 | 3.04 | 0.43 | 3.60 | 2.11 |
| 1wej | ab | −0.35 | 1.34 | −0.27 | 1.17 | 1.33 | 1.46 | −0.17 | 0.08 | 0.86 | 1.37 | 0.05 | −0.31 |
| mean | | −1.55 | 0.38 | −0.96 | 0.86 | 4.11 | 1.45 | −1.17 | −0.18 | 1.42 | 0.73 | 1.28 | 0.90 |
| Stdev | | 2.48 | 0.61 | 2.45 | 0.41 | 2.54 | 0.70 | 1.26 | 0.43 | 1.15 | 0.33 | 0.99 | 0.67 |
| 1ACB | nab | −1.65 | 0.78 | −5.02 | 1.01 | 2.05 | 1.20 | −2.40 | −0.09 | −0.69 | 0.12 | 1.93 | 1.48 |
| 1AVZ | nab | −2.56 | −0.02 | −4.28 | 0.21 | 2.29 | 0.51 | −1.05 | 0.24 | 3.43 | 0.63 | 0.28 | 0.24 |
| 1brs | nab | −1.27 | 0.21 | −3.64 | 0.13 | 5.68 | 2.16 | −1.67 | −0.85 | 1.19 | 0.07 | 2.12 | 1.14 |
| 1CHO | nab | −3.89 | 0.25 | −5.03 | 0.93 | 2.59 | 1.73 | −2.05 | −1.26 | 0.83 | −0.24 | 2.28 | 1.36 |
| 1MDA | nab | −13.41 | −0.10 | −9.26 | −0.14 | −2.74 | 0.11 | 0.67 | 0.10 | −0.40 | −0.23 | 3.58 | 0.72 |
| 1PPF | nab | −2.41 | 0.85 | −3.64 | 1.35 | 1.44 | 0.53 | −1.29 | −0.69 | 2.28 | 0.04 | 1.30 | 1.19 |
| 1SPB | nab | −6.32 | −0.57 | −2.08 | −1.32 | 8.42 | 3.58 | −1.00 | −3.65 | 10.13 | −2.26 | 2.75 | 2.60 |
| 1UGH | nab | −5.80 | −0.45 | −6.60 | −0.20 | 3.94 | 0.97 | −2.92 | −1.04 | 2.73 | −0.59 | 2.68 | 2.01 |
| 2PCC | nab | −6.63 | 1.88 | −5.24 | 1.23 | 3.22 | 0.97 | −2.28 | 1.28 | 3.21 | 1.21 | 0.11 | −0.84 |
| 2PTC | nab | −6.12 | −0.15 | −5.81 | −0.48 | 0.30 | 0.55 | −0.98 | −0.78 | 0.23 | −1.04 | 2.10 | 1.35 |
| 1CSE | nab | −2.66 | 0.34 | −1.87 | −0.05 | 5.51 | 2.43 | −1.53 | −1.01 | 1.05 | −0.34 | 1.86 | 1.54 |
| 1FIN | nab | −9.49 | −1.10 | −5.93 | −0.39 | 6.17 | 1.79 | −7.58 | −2.20 | 0.45 | −0.58 | 5.09 | 2.81 |
| 2BTF | nab | −6.65 | −0.19 | −4.20 | −0.91 | 7.11 | 1.53 | −2.82 | −1.04 | 0.42 | −0.75 | 2.06 | 1.35 |
| mean | | −5.30 | 0.13 | −4.81 | 0.11 | 3.54 | 1.39 | −2.07 | −0.85 | 1.91 | −0.31 | 2.17 | 1.30 |
| Stdev | | 3.45 | 0.74 | 1.94 | 0.83 | 3.05 | 0.96 | 1.91 | 1.20 | 2.80 | 0.84 | 1.29 | 0.94 |

[a] SS, protein secondary structure assignment (α helix, β strand, or both). ID tag, antibody−antigen complex (ab) or nonantibody complex (nab). The electrostatic energies are (from left to right) total electrostatic energy computed by solving the Poisson equation (PB); total electrostatic energy computed using the Generalized Born approximation (GB); Coulomb energy of solute charges (Coul); energy of desolvating solute charges (self-energy); Coulomb energy of solute charges screened by solvent polarization (screened Coul, using GB); surface area estimate of cavity free energy and solute−solvent van der Waals interactions (surface area). All atom pairs are included in PB energies; same residue and adjacent mainchain atom pairs are excluded in GB, Coul, and screened Coul energies.

**TABLE 3: Native (Zn) and Native Repacked (Znr) *Z* scores for a Set of 10 α, 9 αβ, and 6 β Proteins (Section a) and for a Set of 18 Antibody−Antigen (ab) and 13 Nonantibody (nab) Protein−Protein Complexes (Section c) and Low RMSD *Z* scores (Zlrm) for a Set of 15 α, 6 αβ, and 2 β Proteins (Section b) and for a Set of 18 Antibody−Antigen (ab) and 13 Nonantibody (nab) Protein−Protein Complexes (Section d)[a]**

Section a

| | | Diel model | | HB scmc | | HB scsc | | HB mcmc | | HB all | | HB Coul | | HB Coul VdW | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB | SS | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr |
| 1a32 | α | 0.72 | 0.44 | 0.08 | −0.78 | 0.92 | 1.36 | 1.84 | 1.84 | 2.14 | 1.69 | 1.86 | 1.68 | 3.93 | 3.58 |
| 1ail | α | 3.26 | 1.45 | −2.74 | −2.44 | 0.12 | −0.52 | 6.33 | 6.33 | 6.17 | 5.26 | 6.45 | 5.27 | 5.23 | 4.75 |
| 1am3 | α | 0.32 | −0.44 | −0.05 | 0.49 | 0.37 | −0.64 | 2.05 | 2.05 | 2.15 | 2.13 | 1.93 | 2.00 | 2.42 | 2.62 |
| 1cc5 | α | 1.75 | 1.05 | −0.87 | 3.02 | −0.69 | −0.14 | −1.29 | −1.29 | −1.63 | 0.47 | −0.34 | 0.58 | 1.29 | 0.67 |
| 1cei | α | 3.85 | 1.39 | 0.20 | 0.06 | 0.50 | −0.12 | 4.40 | 4.40 | 4.69 | 4.18 | 5.75 | 4.34 | 5.49 | 4.80 |
| 1hyp | α | 0.42 | −0.35 | 0.79 | 1.76 | −0.42 | −0.24 | 1.88 | 1.88 | 2.02 | 2.97 | 1.79 | 2.84 | 2.47 | 3.69 |
| 1lfb | α | 0.89 | 1.53 | −0.81 | 0.15 | −0.11 | 1.16 | 2.16 | 2.16 | 1.99 | 2.67 | 1.87 | 2.76 | 3.55 | 3.64 |
| 1mzm | α | 1.71 | 0.97 | −2.07 | −1.84 | 0.09 | 1.38 | 1.61 | 1.61 | 1.38 | 0.93 | 1.94 | 1.04 | 4.47 | 3.06 |
| 1r69 | α | 2.56 | 0.74 | 0.19 | 0.52 | 2.43 | 2.51 | 0.77 | 0.77 | 1.49 | 1.74 | 2.73 | 1.84 | 5.68 | 4.25 |
| 1utg | α | 3.04 | 2.12 | −1.99 | 0.03 | 1.85 | −0.10 | 3.93 | 3.93 | 4.09 | 4.15 | 4.93 | 4.37 | 3.73 | 3.27 |
| 1ctf | αβ | 2.38 | 1.11 | −0.44 | 0.86 | −0.08 | −0.56 | 4.21 | 4.21 | 4.18 | 4.35 | 4.26 | 4.35 | 5.29 | 5.34 |
| 1dol | αβ | 1.18 | 0.71 | −1.81 | −0.12 | −0.72 | −0.58 | 0.89 | 0.89 | 0.26 | 0.60 | 1.15 | 0.72 | 2.91 | 2.29 |
| 1orc | αβ | 0.92 | 0.19 | −3.07 | −1.37 | 0.50 | 1.86 | 2.93 | 2.93 | 2.41 | 2.06 | 2.61 | 2.03 | 1.69 | 1.33 |
| 1pgx | αβ | 4.02 | 1.52 | −1.50 | −0.30 | 0.39 | −0.58 | 4.49 | 4.49 | 4.19 | 3.61 | 5.70 | 3.78 | 4.51 | 2.36 |
| 1ptq | αβ | 0.70 | 0.55 | 2.73 | 4.17 | 4.71 | 4.18 | −1.00 | −1.00 | 0.22 | 2.65 | 0.29 | 2.57 | 2.88 | 3.59 |
| 1tif | αβ | 1.60 | 0.74 | −1.59 | 0.68 | 0.61 | 2.22 | 5.87 | 5.87 | 5.68 | 5.74 | 4.89 | 5.62 | 4.88 | 4.87 |
| 1vcc | αβ | 3.25 | 3.49 | 0.17 | 1.96 | 0.17 | −0.28 | 3.37 | 3.37 | 3.50 | 4.50 | 4.53 | 4.96 | 4.93 | 4.04 |
| 2fxb | αβ | 2.03 | 0.66 | 0.66 | 4.25 | 5.09 | 3.55 | −0.21 | −0.21 | 1.08 | 3.24 | 2.11 | 3.20 | 3.94 | 3.76 |
| 5icb | αβ | −0.99 | −2.47 | 0.49 | 0.77 | 4.38 | 2.93 | 2.62 | 2.62 | 3.56 | 3.46 | 1.41 | 2.60 | 2.73 | 2.90 |
| 1bq9 | β | 3.20 | 2.64 | 0.13 | 0.40 | 1.48 | 4.41 | 4.88 | 4.88 | 5.28 | 5.23 | 5.25 | 5.59 | 4.67 | 4.28 |
| 1csp | β | 1.79 | 1.31 | −1.16 | 0.04 | −0.61 | 0.75 | 4.29 | 4.29 | 4.02 | 4.06 | 4.04 | 4.13 | 3.40 | 3.29 |
| 1msi | β | 5.34 | 4.16 | 0.95 | 1.37 | 0.86 | 0.01 | 1.56 | 1.56 | 1.92 | 2.50 | 3.72 | 2.95 | 4.80 | 4.53 |
| 1tuc | β | 2.51 | 2.05 | −0.20 | 2.53 | 0.16 | −1.18 | 3.05 | 3.05 | 3.25 | 5.00 | 4.07 | 5.20 | 4.39 | 4.54 |
| 1vif | β | 3.31 | 3.26 | −0.45 | −0.69 | −0.24 | 0.76 | 2.87 | 2.87 | 2.68 | 2.44 | 3.48 | 2.76 | 3.32 | 2.77 |
| 5pti | β | 1.42 | 1.82 | −0.33 | 2.64 | 2.05 | 1.22 | 2.90 | 2.90 | 3.27 | 4.74 | 3.13 | 4.81 | 3.24 | 3.34 |
| mean | | 2.05 | 1.23 | −0.51 | 0.73 | 0.95 | 0.93 | 2.66 | 2.66 | 2.80 | 3.21 | 3.18 | 3.28 | 3.83 | 3.50 |
| Stdev | | 1.42 | 1.36 | 1.29 | 1.69 | 1.64 | 1.61 | 1.95 | 1.95 | 1.79 | 1.47 | 1.75 | 1.50 | 1.17 | 1.09 |

Section b

| | | Diel model | | HB scmc | | HB scsc | | HB mcmc | | HB all | | HB Coul | | HB Coul VdW | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Zlrm | | Zlrm | | Zlrm | | Zlrm | | Zlrm | | Zlrm | | Zlrm | |
| PDB | SS | −PN | +PN | −PN | +PN | −PN | +PN | −PN | +PN | −PN | +PN | −PN | +PN | −PN | +PN |
| 1a32 | α | 0.18 | 0.17 | −0.73 | −0.62 | −0.40 | −0.37 | 1.25 | 0.69 | 1.10 | 0.58 | 0.99 | 0.64 | 1.11 | 0.83 |
| 1am3 | α | −0.32 | −0.22 | 0.16 | −0.03 | −0.23 | −0.15 | 0.42 | 0.59 | 0.44 | 0.58 | 0.26 | 0.47 | 0.45 | 0.69 |
| 1bw6 | α | 0.11 | 0.26 | −0.18 | −0.30 | −0.01 | 0.15 | 0.61 | 0.14 | 0.59 | 0.09 | 0.50 | 0.26 | 0.66 | 0.57 |
| 1gab | α | 0.47 | 0.29 | 0.94 | 0.68 | −0.08 | 0.11 | 0.62 | 0.15 | 0.88 | 0.34 | 0.90 | 0.40 | 1.12 | 0.69 |
| 1kjs | α | 0.04 | 0.12 | 0.12 | 0.29 | 0.11 | 0.19 | 0.39 | 0.32 | 0.45 | 0.46 | 0.39 | 0.40 | 0.72 | 0.78 |
| 1mzm | α | −0.10 | 0.64 | −0.16 | −0.94 | −0.01 | 0.80 | 0.56 | 1.71 | 0.57 | 1.81 | 0.47 | 1.71 | 0.42 | 2.10 |
| 1nkl | α | −0.09 | 0.84 | 0.05 | −1.07 | −0.25 | 0.39 | 0.06 | 2.11 | 0.04 | 2.11 | 0.00 | 2.10 | 0.03 | 2.14 |
| 1nre | α | 0.93 | 0.52 | −1.01 | −0.83 | 0.04 | −0.28 | 1.42 | 1.79 | 1.37 | 1.84 | 1.52 | 1.78 | 1.46 | 1.75 |
| 1pou | α | 0.47 | 0.50 | −0.13 | −0.09 | 0.05 | −0.36 | 0.21 | 1.82 | 0.20 | 1.84 | 0.43 | 1.97 | 0.64 | 1.73 |
| 1r69 | α | 0.93 | 0.66 | 0.58 | 0.39 | 0.59 | 0.06 | 0.03 | 1.78 | 0.31 | 1.97 | 0.67 | 2.00 | 1.07 | 2.29 |
| 1res | α | 0.08 | 0.10 | −0.01 | 0.00 | −0.11 | −0.14 | 0.32 | 0.04 | 0.32 | 0.03 | 0.33 | 0.08 | 0.35 | 0.19 |
| 1uba | α | 0.56 | 0.40 | 0.14 | 0.23 | 0.04 | 0.27 | 0.10 | −0.26 | 0.14 | −0.16 | 0.36 | 0.05 | 0.24 | 0.07 |
| 1uxd | α | 0.22 | 0.33 | −0.19 | −0.19 | 0.43 | 0.40 | 1.09 | 0.35 | 1.12 | 0.37 | 1.07 | 0.51 | 1.26 | 0.86 |
| 2ezh | α | 0.06 | 0.00 | −0.35 | 0.01 | −0.23 | −0.47 | 0.71 | 1.54 | 0.64 | 1.57 | 0.59 | 1.38 | 0.40 | 1.42 |
| 2pdd | α | 0.23 | 0.35 | 0.50 | 0.44 | 0.49 | 0.42 | 0.30 | 0.32 | 0.50 | 0.55 | 0.46 | 0.57 | 0.47 | 0.84 |
| 1aa3 | αβ | 0.66 | 0.33 | 0.34 | −0.06 | 0.16 | 0.37 | 0.19 | 0.72 | 0.31 | 0.84 | 0.61 | 0.82 | 0.75 | 0.91 |
| 1afi | αβ | 0.86 | 0.38 | 0.13 | −1.16 | 0.37 | 0.80 | 0.81 | 2.76 | 0.93 | 2.67 | 1.12 | 2.22 | 1.26 | 2.05 |
| 1ctf | αβ | 0.42 | 1.03 | 0.22 | 0.26 | 0.12 | −0.35 | −0.02 | 2.72 | 0.05 | 2.80 | 0.29 | 2.63 | 0.69 | 2.90 |
| 1pgx | αβ | 0.44 | 0.96 | −0.51 | −0.08 | −0.22 | −0.52 | 0.94 | 2.95 | 0.76 | 2.88 | 0.92 | 2.89 | 1.01 | 2.50 |
| 2fow | αβ | 0.00 | −0.11 | 0.69 | 0.09 | −0.32 | −0.07 | −0.05 | 1.36 | 0.10 | 1.49 | 0.07 | 1.08 | −0.12 | 0.87 |
| 2ptl | αβ | 0.34 | 0.48 | −0.57 | −0.22 | −0.14 | −0.37 | 0.69 | 1.89 | 0.57 | 1.88 | 0.66 | 1.74 | 0.95 | 1.55 |
| 1sro | β | 0.81 | 1.83 | −0.35 | 0.68 | 0.19 | 0.41 | 0.82 | 0.76 | 0.83 | 1.18 | 1.19 | 1.97 | 0.84 | 2.06 |
| 1vif | β | 2.67 | 2.33 | 0.35 | −0.16 | 0.22 | 0.18 | 1.53 | 1.62 | 1.61 | 1.58 | 2.32 | 2.24 | 2.20 | 1.89 |
| mean | | 0.43 | 0.53 | 0.00 | −0.12 | 0.03 | 0.06 | 0.56 | 1.21 | 0.60 | 1.27 | 0.70 | 1.30 | 0.78 | 1.38 |
| Stdev | | 0.60 | 0.58 | 0.47 | 0.52 | 0.27 | 0.39 | 0.46 | 0.95 | 0.41 | 0.90 | 0.51 | 0.86 | 0.50 | 0.76 |

Section c

| | | Diel model | | HB scmc | | HB scsc | | HB mcmc | | HB all | | HB Coul | | HB Coul VdW | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB | ID tag | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr |
| 1a2y | ab | 2.91 | 1.12 | 2.95 | 2.62 | 2.93 | 2.57 | 2.65 | 2.65 | 3.41 | 2.96 | 4.59 | 3.17 | 3.67 | 4.72 |
| 1cz8 | ab | 4.93 | 0.93 | 1.98 | 0.33 | 2.01 | 0.63 | 4.77 | 4.77 | 4.56 | 3.88 | 6.31 | 3.96 | 5.75 | 4.97 |
| 1dqj | ab | 3.56 | 1.10 | 0.65 | 1.25 | 0.99 | 1.74 | 3.87 | 3.87 | 2.20 | 2.91 | 3.44 | 2.97 | 4.58 | 4.22 |
| 1e6j | ab | 10.43 | 1.30 | 2.23 | 2.46 | 2.06 | 2.99 | 4.28 | 4.28 | 3.26 | 4.00 | 6.50 | 3.98 | 9.16 | 4.72 |
| 1egj | ab | 2.74 | 1.77 | 1.31 | 0.89 | 1.38 | 1.15 | −0.37 | −0.37 | 1.02 | 0.93 | 2.56 | 1.56 | 4.04 | 2.10 |
| 1eo8 | ab | 10.38 | 2.12 | −0.28 | 0.84 | 1.46 | 3.29 | −0.39 | −0.39 | 1.31 | 3.74 | 7.93 | 4.00 | 13.66 | 4.35 |
| 1fdl | ab | 1.59 | 0.24 | 2.56 | 2.42 | 2.21 | 2.43 | 2.93 | 2.93 | 2.86 | 2.96 | 3.26 | 2.79 | 3.43 | 3.83 |

**TABLE 3: (Continued)**

Section c (Continued)

| | | Diel model | | HB scmc | | HB scsc | | HB mcmc | | HB all | | HB Coul | | HB Coul VdW | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB | ID tag | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr | Zn | Znr |
| 1fj1 | ab | 4.84 | 0.59 | 4.30 | 2.67 | 3.40 | 2.58 | −0.23 | −0.23 | 2.96 | 2.27 | 5.32 | 2.22 | 7.28 | 2.31 |
| 1g7h | ab | 0.78 | −0.23 | 2.71 | 2.90 | 2.54 | 2.51 | 1.92 | 1.92 | 2.82 | 2.52 | 2.60 | 2.13 | 3.19 | 2.85 |
| 1ic4 | ab | 4.61 | 3.96 | 2.52 | 3.17 | 2.55 | 3.12 | 3.85 | 3.85 | 3.54 | 3.84 | 4.81 | 4.52 | 5.38 | 5.60 |
| 1jhl | ab | 1.21 | 1.16 | 0.24 | −0.57 | −0.06 | −0.17 | 2.27 | 2.27 | 0.81 | 0.86 | 1.49 | 1.29 | 4.37 | 2.66 |
| 1jrh | ab | 2.74 | 1.49 | 2.99 | 4.20 | 3.10 | 4.15 | 6.59 | 6.59 | 4.97 | 5.59 | 4.40 | 4.59 | 4.08 | 5.10 |
| 1mlc | ab | 3.40 | 1.24 | 0.49 | 0.49 | 2.32 | 1.53 | 1.91 | 1.91 | 2.88 | 2.23 | 3.55 | 2.32 | 3.71 | 2.97 |
| 1nca | ab | 5.64 | 3.20 | 3.26 | 4.39 | 2.80 | 3.78 | −0.42 | −0.42 | 2.27 | 2.74 | 4.43 | 3.23 | 10.42 | 3.20 |
| 1nsn | ab | 7.01 | 0.61 | −0.30 | −0.46 | −0.19 | −0.66 | −0.30 | −0.30 | −0.24 | −0.78 | 4.49 | −0.34 | 5.78 | 0.44 |
| 1osp | ab | 4.48 | 0.96 | 1.37 | 1.55 | 1.00 | 1.69 | 3.97 | 3.97 | 2.66 | 3.11 | 4.93 | 3.10 | 7.17 | 4.21 |
| 1qfu | ab | 8.29 | 2.16 | −0.14 | 1.22 | 1.41 | 2.83 | −0.46 | −0.46 | 1.23 | 2.84 | 5.62 | 3.00 | 8.60 | 2.84 |
| 1wej | ab | 1.74 | 0.87 | 2.72 | 1.12 | 3.10 | 2.54 | −0.31 | −0.31 | 2.66 | 2.72 | 2.59 | 2.27 | 3.41 | 2.43 |
| mean | | 4.51 | 1.37 | 1.75 | 1.75 | 1.95 | 2.15 | 2.03 | 2.03 | 2.51 | 2.74 | 4.38 | 2.82 | 5.98 | 3.53 |
| Stdev | | 2.91 | 1.01 | 1.38 | 1.44 | 1.05 | 1.29 | 2.24 | 2.24 | 1.27 | 1.36 | 1.58 | 1.20 | 2.81 | 1.29 |
| 1ACB | nab | 3.30 | 1.34 | −0.11 | −1.30 | −0.23 | −1.12 | 11.13 | 11.13 | 7.83 | 6.46 | 6.81 | 6.79 | 5.80 | 7.77 |
| 1AVZ | nab | 2.41 | 0.42 | 0.86 | 1.20 | 0.80 | 1.97 | −0.25 | −0.25 | 0.69 | 1.84 | 2.13 | 1.63 | 4.38 | 2.33 |
| 1brs | nab | 6.93 | 2.42 | 3.31 | 3.32 | 4.13 | 2.96 | −0.46 | −0.46 | 3.06 | 2.33 | 4.44 | 2.73 | 8.88 | 3.04 |
| 1CHO | nab | 2.33 | 0.99 | −0.13 | −0.76 | −0.44 | −0.27 | 9.76 | 9.76 | 6.62 | 6.21 | 6.64 | 6.17 | 11.39 | 6.82 |
| 1MDA | nab | 0.45 | −0.74 | −1.35 | −0.77 | −1.53 | −0.51 | −0.56 | −0.56 | −1.56 | −0.74 | 0.50 | −0.97 | 12.23 | −0.63 |
| 1PPF | nab | 2.85 | 0.74 | −0.90 | −0.73 | −1.03 | −0.76 | 9.06 | 9.06 | 5.56 | 5.26 | 6.86 | 5.44 | 10.80 | 6.19 |
| 1SPB | nab | 9.00 | 3.90 | 6.13 | 5.04 | 5.18 | 4.78 | 9.20 | 9.20 | 9.57 | 9.09 | 11.10 | 9.04 | 14.20 | 8.78 |
| 1UGH | nab | 6.67 | 0.87 | 4.30 | 3.70 | 3.74 | 3.41 | −0.44 | −0.44 | 3.30 | 2.89 | 7.69 | 2.64 | 15.97 | 2.83 |
| 2PCC | nab | 5.29 | 0.72 | −0.56 | 0.14 | −0.82 | 0.07 | −0.48 | −0.48 | −0.88 | −0.12 | 6.41 | 0.24 | 13.38 | 0.45 |
| 2PTC | nab | 0.49 | 0.30 | 3.52 | 1.90 | 2.91 | 1.75 | 5.23 | 5.23 | 5.48 | 4.38 | 3.72 | 4.23 | 8.79 | 4.55 |
| 1CSE | nab | 6.04 | 2.18 | 2.14 | 1.01 | 1.52 | 0.97 | 7.51 | 7.51 | 6.17 | 5.45 | 7.88 | 5.68 | 11.49 | 6.38 |
| 1FIN | nab | 8.14 | 1.86 | 5.45 | 5.38 | 4.96 | 4.85 | −0.33 | −0.33 | 4.74 | 4.49 | 8.87 | 4.17 | 16.17 | 4.14 |
| 2BTF | nab | 6.79 | 1.64 | 1.07 | 2.16 | 1.57 | 2.26 | 2.81 | 2.81 | 2.31 | 2.91 | 5.48 | 3.22 | 19.75 | 3.43 |
| mean | | 4.67 | 1.28 | 1.83 | 1.56 | 1.60 | 1.57 | 4.01 | 4.01 | 4.07 | 3.88 | 6.04 | 3.92 | 11.79 | 4.31 |
| Stdev | | 2.85 | 1.16 | 2.50 | 2.26 | 2.37 | 2.04 | 4.74 | 4.74 | 3.19 | 2.63 | 2.71 | 2.65 | 4.10 | 2.69 |

Section d

| | | Diel model | HB scmc | HB scsc | HB mcmc | HB all | HB Coul | HB Coul VdW |
|---|---|---|---|---|---|---|---|---|
| PDB | ID tag | Zlrm | Zlrm | Zlrm | Zlrm | Zlrm | Zlrm | Zlrm |
| 1a2y | ab | 0.02 | −0.32 | 0.00 | −0.41 | 0.08 | −0.03 | 0.14 |
| 1cz8 | ab | 0.66 | 0.34 | 0.38 | 2.41 | 1.57 | 1.66 | 2.05 |
| 1dqj | ab | 0.69 | 1.74 | 1.83 | 0.36 | 1.81 | 1.89 | 2.51 |
| 1e6j | ab | 1.11 | 1.85 | 2.38 | 1.85 | 2.72 | 2.82 | 2.77 |
| 1egj | ab | 0.68 | 1.54 | 1.86 | −0.37 | 1.68 | 1.98 | 2.26 |
| 1eo8 | ab | 1.17 | −0.05 | 1.02 | −0.06 | 1.45 | 1.72 | 1.76 |
| 1fdl | ab | 0.55 | 0.71 | 0.68 | 0.44 | 0.67 | 0.40 | 0.24 |
| 1fj1 | ab | 0.25 | 2.08 | 2.07 | −0.17 | 1.88 | 1.83 | 2.48 |
| 1g7h | ab | −0.13 | 1.36 | 1.48 | 1.71 | 1.77 | 1.57 | 1.33 |
| 1ic4 | ab | 1.40 | 1.90 | 2.18 | 2.68 | 2.69 | 2.70 | 2.67 |
| 1jhl | ab | 0.58 | −0.04 | −0.04 | 0.04 | −0.02 | 0.10 | 0.12 |
| 1jrh | ab | 1.68 | 1.93 | 1.84 | 1.63 | 1.93 | 2.22 | 2.03 |
| 1mlc | ab | 1.16 | 0.01 | 0.93 | 0.81 | 1.40 | 1.57 | 2.12 |
| 1nca | ab | 1.63 | 2.78 | 2.35 | 0.58 | 1.99 | 2.25 | 2.53 |
| 1nsn | ab | 1.18 | −0.64 | −0.62 | −0.30 | −0.61 | 0.00 | 0.11 |
| 1osp | ab | 1.21 | −0.10 | 0.19 | −0.23 | 0.25 | 0.46 | 0.74 |
| 1qfu | ab | 1.51 | 1.31 | 2.43 | 0.35 | 2.68 | 2.77 | 2.74 |
| 1wej | ab | 0.72 | 0.67 | 0.91 | −0.31 | 0.84 | 0.64 | 0.61 |
| mean | | 0.89 | 0.95 | 1.21 | 0.61 | 1.38 | 1.48 | 1.62 |
| Stdev | | 0.53 | 1.00 | 0.96 | 1.01 | 0.95 | 0.94 | 0.99 |
| 1ACB | nab | 0.70 | −0.02 | −0.19 | 3.88 | 2.61 | 2.26 | 2.03 |
| 1AVZ | nab | 0.81 | 0.42 | 0.56 | −0.25 | 0.42 | 0.76 | 1.04 |
| 1brs | nab | 1.31 | 2.49 | 2.46 | 0.54 | 2.08 | 2.22 | 2.87 |
| 1CHO | nab | 0.65 | 0.16 | 0.42 | 5.00 | 4.23 | 4.13 | 4.28 |
| 1MDA | nab | 0.04 | 0.35 | 0.23 | 0.65 | 0.59 | 0.43 | 1.04 |
| 1PPF | nab | 0.70 | −0.25 | −0.37 | 10.80 | 7.45 | 7.28 | 6.55 |
| 1SPB | nab | 2.35 | 0.94 | 1.01 | 5.39 | 4.57 | 4.73 | 4.40 |
| 1UGH | nab | 0.74 | 1.78 | 1.85 | −0.01 | 1.45 | 1.28 | 1.46 |
| 2PCC | nab | 0.75 | 1.19 | 0.98 | −0.25 | 0.55 | 0.82 | 0.91 |
| 2PTC | nab | 0.47 | 1.43 | 1.37 | 3.13 | 3.05 | 2.82 | 2.82 |
| 1CSE | nab | 1.69 | 0.71 | 0.71 | 4.51 | 3.62 | 4.04 | 4.11 |
| 1FIN | nab | 1.36 | 0.49 | 0.72 | −0.28 | 0.61 | 1.13 | 1.58 |
| 2BTF | nab | 1.41 | 0.06 | 1.08 | 0.69 | 1.55 | 2.00 | 2.24 |
| mean | | 1.00 | 0.75 | 0.83 | 2.60 | 2.52 | 2.61 | 2.72 |
| Stdev | | 0.60 | 0.79 | 0.77 | 3.28 | 1.98 | 1.90 | 1.63 |

[a] SS, protein secondary structure assignment (α helix, β strand, or both). ID tag, antibody−antigen complex (ab) or nonantibody complex (nab). The electrostatic energies are (from left to right): Coulomb interactions with the Warshel distance-dependent dielectric (Diel model); side chain-mainchain H bonds (HB scmc); side chain-side chain H-bonds (HB scsc); mainchain-mainchain H bonds (HB mcmc). HB all, combined H-bond energies; HB Coul, combined Warshel Coulomb and H-bond energies; HB Coul VdW, combined Warshel Coulomb, H bond, and van der Waals energies. SS, protein secondary structure assignment (α helix, β strand, or both). −PN subcolumn, *Z* scores for the ab initio decoy set. +PN subcolumn, *Z* scores for the ab initio decoy set enhanced with perturbed-native structures.

**TABLE 4: Average Native (Zn), Native Repacked (Znr), Low RMSD (Zlrm) *Z* scores, and the Number of Successful Discriminations (#SD, defined as *Z* score > 1) for the Energy Functions in the Left Column[a]**

| energy function | SDM | | | | | AB | | | | NAB | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Zlrm | | | | | | | | | | |
| | Zn | Znr | −PN | +PN | #SD | Zn | Znr | Zlrm | #SD | Zn | Znr | Zlrm | #SD |
| PB | −0.05 | 1.20 | 0.25 | | | −1.55 | 0.38 | −0.32 | 0 | −5.30 | 0.13 | −0.23 | 0 |
| PB total solv | −1.04 | −0.68 | −0.44 | | | −2.73 | −0.80 | −0.75 | 0 | −2.90 | −1.20 | −1.06 | 0 |
| GB | −0.88 | 0.52 | 0.33 | 0.38 | 2 | −0.96 | 0.86 | −0.01 | 0 | −4.81 | 0.11 | −0.09 | 0 |
| GB total solv | −0.22 | −0.25 | −0.32 | 0.03 | 0 | −1.34 | −0.25 | −0.74 | 0 | −2.03 | −1.06 | −1.04 | 0 |
| Coul | 1.59 | 0.90 | 0.49 | 0.59 | 4 | 4.11 | 1.45 | 0.84 | 6 | 3.54 | 1.39 | 1.15 | 6 |
| self-energy | 0.46 | −0.08 | −0.14 | 0.32 | 1 | −1.17 | −0.18 | −0.70 | 0 | −2.07 | −0.85 | −0.84 | 0 |
| screened Coul | 1.85 | 1.46 | 0.25 | 0.58 | 5 | 1.42 | 0.73 | −0.21 | 0 | 1.91 | −0.31 | −0.52 | 0 |
| surface area | 1.66 | 1.29 | 0.49 | 0.30 | 2 | 1.28 | 0.90 | 0.99 | 10 | 2.17 | 1.30 | 1.28 | 11 |
| Diel model | 2.05 | 1.23 | 0.43 | 0.53 | 3 | 4.51 | 1.37 | 0.89 | 9 | 4.67 | 1.28 | 1.00 | 5 |
| HB scmc | −0.51 | 0.73 | 0.00 | −0.12 | 0 | 1.75 | 1.75 | 0.95 | 9 | 1.83 | 1.56 | 0.75 | 4 |
| HB scsc | 0.95 | 0.93 | 0.03 | 0.06 | 0 | 1.95 | 2.15 | 1.21 | 10 | 1.60 | 1.57 | 0.83 | 5 |
| HB mcmc | 2.66 | 2.66 | 0.56 | 1.21 | 12 | 2.03 | 2.03 | 0.61 | 5 | 4.01 | 4.01 | 2.60 | 6 |
| HB all | 2.80 | 3.21 | 0.60 | 1.27 | 13 | 2.51 | 2.74 | 1.38 | 12 | 4.07 | 3.88 | 2.52 | 9 |
| HB Coul | 3.18 | 3.28 | 0.70 | 1.30 | 13 | 4.38 | 2.82 | 1.48 | 12 | 6.04 | 3.92 | 2.61 | 9 |
| HB Coul VdW | 3.83 | 3.50 | 0.78 | 1.38 | 12 | 5.98 | 3.53 | 1.62 | 12 | 11.79 | 4.31 | 2.72 | 12 |

[a] SDM, single domain set; AB, antibody−antigen set; NAB, nonantibody set. −PN subcolumn, ab initio single domain set. +PN subcolumn, ab initio single domain set enhanced with perturbed−native structures. #SD refers to Zlrm (+PN) for single domain proteins and to Zlrm for protein−protein complexes. Energy functions are as in Tables 2 and 3; additionally, PB total solv is the total PB solvation energy, and GB total solv is the total GB solvation energy. Bonded atoms and atoms in the same residue were excluded from all interatomic energy functions except PB, PB total solv.

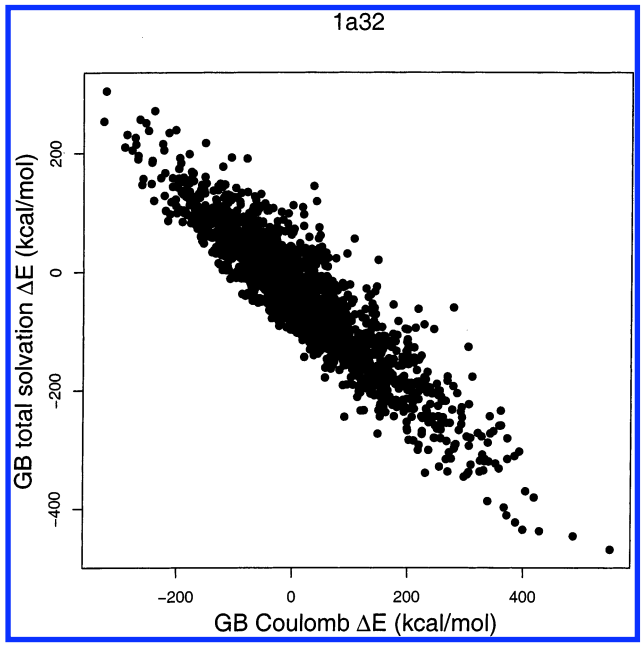


**Figure 3.** Generalized Born total solvation energy vs Coulomb energy for *1a32* decoys (in kcal/mol). All atom−atom pairs are included; the energies are computed relative to the native structure.

lower solvation self-energies but higher screened Coulomb energies with respect to the native structure. When these two terms are added together to yield the GB electrostatic energy, there is no clear separation any more, and the energy gaps disappear (rightmost plot, second row of Figure 4a).

To investigate further the extent to which self-energies are compensated by favorable electrostatic interactions with other protein atoms, we considered the electrostatic energies of different atom types in a set of monomeric native structures. Figure 4b shows, from left to right, the self-energy, the screened Coulomb energy, and the GB electrostatic energy as a function of the number of atoms within 10 Å. Interactions between all protein atom pairs are computed. The atom types shown (from top to bottom of Figure 4b) are backbone carbonyl oxygen, the side chain $N_\zeta$ nitrogen (of lysine), the backbone carbonyl carbon, and the backbone amide hydrogen. For the mainchain carbonyl oxygen, the side chain $N_\zeta$, the $C_\alpha$ carbon (not shown), and most other side chain heavy atoms (not shown), the self-energy increases with the number of neighbors, disfavoring the native structure which is better packed than decoys, whereas the screened Coulomb energy becomes lower for buried atoms. The extent of their compensation is evident in the GB electrostatic energy; there is still energy decrease with burial, but it is less marked than for the screened Coulomb energy alone. Different results are obtained for the backbone carbonyl carbons, backbone amide hydrogens, and most other hydrogen atoms, where the self-energy is again less favorable for buried atoms, but the screened Coulomb energy stays approximately constant throughout the range of burial. This makes the GB electrostatic energies unfavorable for native structures relative to decoys for these atom types.

A likely reason for the observed lack of compensation of self-energies by screened Coulomb interactions is the neglect of polarization effects in current continuum models of electrostatic interactions. Polarization effects are expected to alleviate the unfavorable self-energy term, and thus facilitate compensation of the solvation and Coulomb terms. The divergent behavior of some backbone atoms might reflect particularly strong polarization effects in regular secondary structure elements in proteins, suggesting the need for more accurate description of backbone electrostatics.

The size of the electrostatic energy gaps is also considerably affected by the atom exclusion scheme (see Methods and Theory). Keeping interactions between all atom pairs is necessary for obtaining the classical electrostatics energy of a point-charge system; for this reason, all interactions are included in PB calculations[16] and in the corresponding GB models.[30] On the other hand, bonded interactions are treated differently in most molecular force fields,[53,58] because quantum-mechanical effects are pronounced for bonded atoms, and the simple point-charge model is inaccurate. In the PB column of Table 2a,b, we sum up the energies of all atoms, including covalently bonded ones; however, we exclude same residue and neighboring mainchain atoms in the GB, Coul, and Screened Coul columns of Table 2a,b and in the corresponding rows of Table

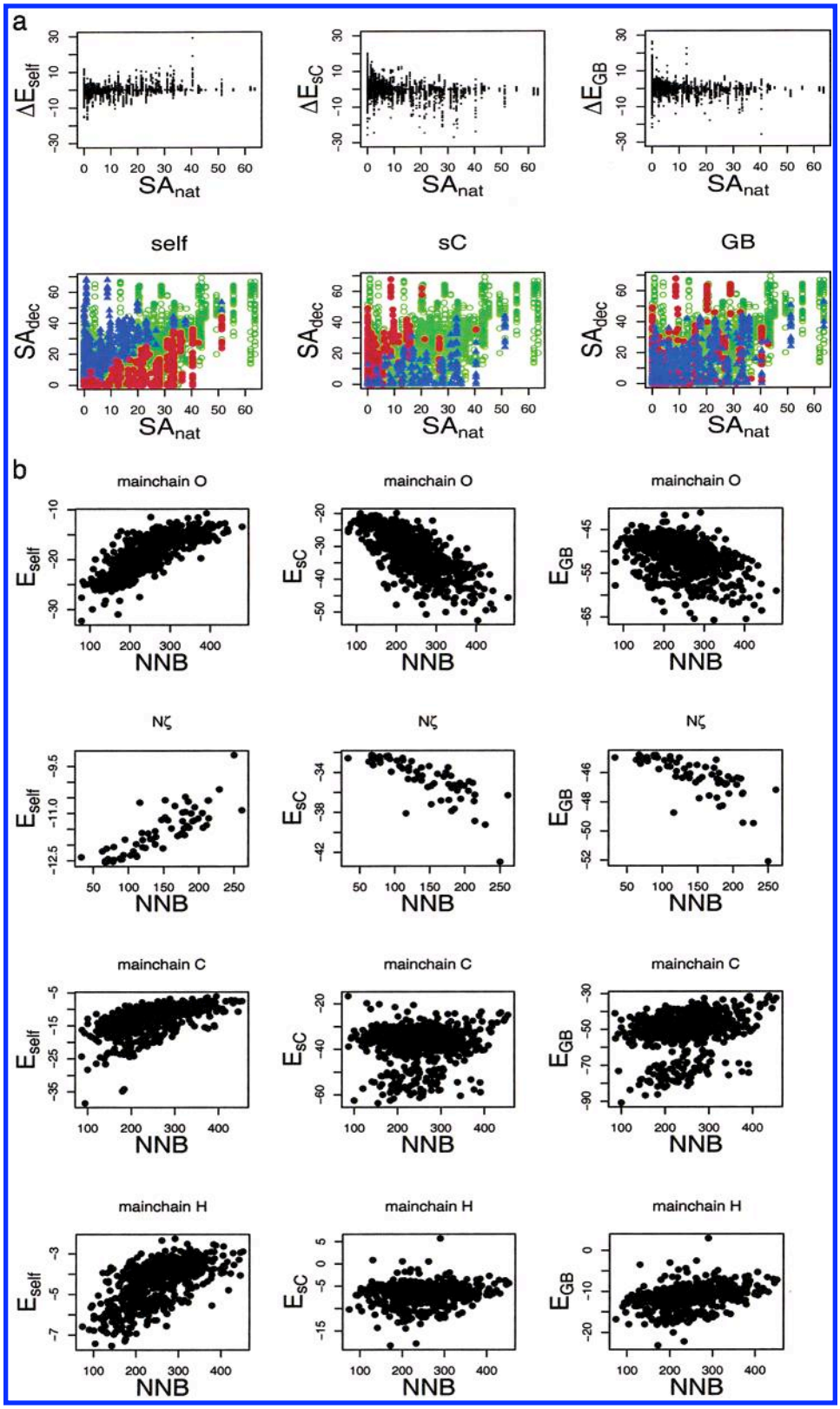


**Figure 4.** Atomic generalized Born energies for *1pgx* relaxed decoys (a) and monomeric native structures (b). (a) First row: solvation self-energy ($E_{self}$, left plot), screened Coulomb interactions ($E_{sC}$; middle plot) and GB electrostatic energy ($E_{GB}$; right plot), computed relative to the native structure ($\Delta E = E_{dec} - E_{nat}$), vs native solvent-accessible surface area ($SA_{nat}$). Second row: decoy solvent-accessible surface area ($SA_{dec}$) vs native solvent-accessible surface area ($SA_{nat}$), with blue triangles indicating atoms for which $\Delta E < -\Delta E_{thr}$, red circles indicating atoms for which $\Delta E > \Delta E_{thr}$, and green open circles indicating atoms for which $-\Delta E_{thr} < \Delta E < \Delta E_{thr}$. The energies considered are the same as in the first row. $E_{thr} = 2$ kcal/mol for $E_{self}$ and $E_{GB}$; 5 kcal/mol for $E_{sC}$. (b) Solvation self-energy ($E_{self}$, left column), screened Coulomb energy ($E_{sC}$; middle column) and GB electrostatic energy ($E_{GB}$; right column) vs the number of atoms within 10 Å (NNB), for mainchain carbonyl O (first row), side chain $N_\zeta$ of K (second row), mainchain carbonyl C (third row), and mainchain amide H (fourth row). The bimodal distribution of the mainchain carbonyl C energies is due to different charges on the C atoms of K,R in the AMBER force field (0.73e vs 0.54−0.60e for the other amino acids). The number of neighbors and the solvent-accessible surface area provide alternative measures of atom burial.

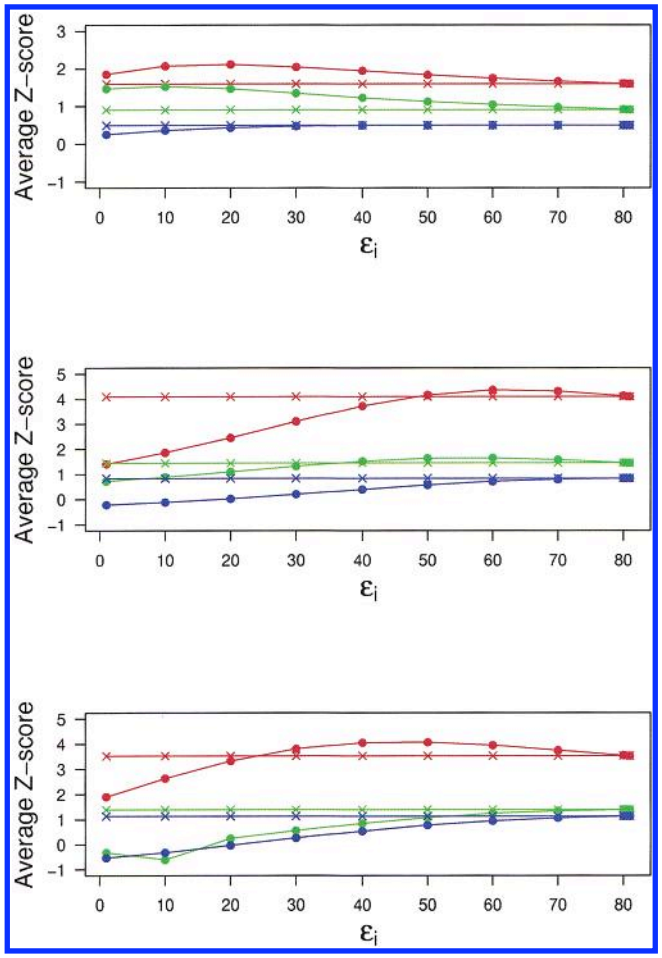


**Figure 5.** *Z* scores of Coulomb interactions screened by solvent polarization (using the GB model) and averaged over decoy sets as a function of $\epsilon_i$, the dielectric constant inside the protein cavity. Red, native *Z* scores; green, native-repacked *Z* scores; blue, low RMSD *Z* scores. Filled circles, screened Coulomb energies; crosses, Coulomb energies with constant dielectric. Upper plot, single-domain proteins; middle plot, antibody−antigen complexes; lower plot, other protein−protein complexes.

4. This contributes to some of the discrepancies between the PB and GB columns of Table 2a,b, because small differences in bond lengths and bond angles between idealized decoys (created using standard force-field bond lengths and angles) and experimentally determined native structures often result in noticeable energy gap variations. This effect is also partially responsible for consistent discrepancies between native and native repacked *Z* scores in the PB and GB columns; when the side chains are repacked, they are also idealized.

We observed that in the GB model the largest energy gaps are provided by the Coulomb interactions screened by solvent polarization, with chemically bonded atoms excluded from the energy sums. The solute cavity dielectric constant is a variable input parameter, and can be adjusted to obtain maximum *Z* scores. The assumption of a single uniform dielectric constant for the protein interior is clearly incorrect;[1,3,4] because an exact value of the protein dielectric constant cannot be defined, the *Z*-score maximization procedure can be viewed as one way of setting its effective average value. The optimum value of $\epsilon_i$ is shown in Figure 5 for average native, native repacked, and low RMSD *Z* scores; it is an an indicator of the degree of screening of Coulomb interactions by solvent polarization. For single-domain proteins, the optimum value of $\epsilon_i$ lies in the 10−20 range for native and native repacked *Z* scores; for low RMSD *Z* scores, the Coulomb term always has the largest energy gap. Optimum values of $\epsilon_i$ are ∼55−65 for antibody−antigen complexes and ∼45−55 for other complexes (for native *Z* scores). Antibody−antigen interfaces are known to be more solvated than the other interface types;[67] this is consistent with our finding that the optimum dielectric constant is closer to water in the former case. On the other hand, the optimum dielectric constant is much lower in single-domain structures, where the protein core is well packed and water penetration is negligible.

Finally, we consider the surface area term designed to capture the cost of making an empty cavity in solvent; by itself, it constitutes a simplified solvation model similar to other effective models discussed in refs 47 and 48. The surface area term exhibits native and native repacked energy gaps (Surface Area column of Table 2a,b; Table 4), showing that native structures are indeed more compact than decoys. Unfortunately, surface areas do not help discriminate distant decoys from nativelike ones; in fact, none of the energies discussed so far produce statistically significant low RMSD *Z* scores for single-domain proteins, even when ab initio decoy sets are complemented with low RMSD perturbed-native structures (see Methods; Table 4).

**3.3. Effective Dielectric Models.** Next, we considered three effective dielectric models widely used for computing electrostatic effects in protein structure prediction and design: the Warshel model,[1] the Sternberg model,[41] and the linear model.[40] These models are pairwise additive and therefore as efficient as a Coulomb calculation with $\epsilon_i(r) = \text{const}$. They describe the same physical interactions as the screened Coulomb model derived using the GB approach; we assume that a model that produces the largest free energy gaps is likely to describe essential physics of charge−charge interactions better than the other approaches.

Figure 6 shows that the Warshel and Sternberg models produce native energy gaps comparable in magnitude to those obtained using the screened Coulomb GB energy with an optimum dielectric constant inside a protein cavity. Indeed, the average native *Z* score for the single domain set is 2.05 for the Sternberg and Warshel models and 2.11 for the screened Coulomb GB model with $\epsilon_i = 20$. The same is true for native repacked structures (data not shown). The linear model does not produce comparable *Z* scores (the average native *Z* score in the single domain set is 1.63). The improved performance of the nonlinear models suggests that the attenuation of electric fields inside proteins, perhaps due to induced polarization and side chain conformational changes, is greater than in a linear dielectric medium.

Next, we investigate how different force fields available for biological macromolecules affect our comparison of effective dielectric models with PB and GB calculations. In Figure 7, we present native repacked *Z* scores for Coulomb calculations with constant dielectric permittivity for three widely used force fields: CHARMM19,[58] AMBER,[53] and OPLS.[68] The AMBER parametrization requires all hydrogen atoms, whereas CHARMM19 and OPLS only consider polar hydrogens explicitly (nonpolar hydrogens are combined with the attached heavy atoms). We observe a high degree of correlation between results employing these different parameter sets, with the average native repacked *Z* score of 0.70 (CHARMM19), 0.90 (AMBER), and 0.95 (OPLS). This correlation is also observed when native *Z* scores are considered or when effective dielectric models with different force-field parameter sets are compared with each other.

Finally, there is a question of which atoms and residues contribute most to the signal observed in the *Z*-score analysis. For example, it is not unreasonable to expect that only side chain

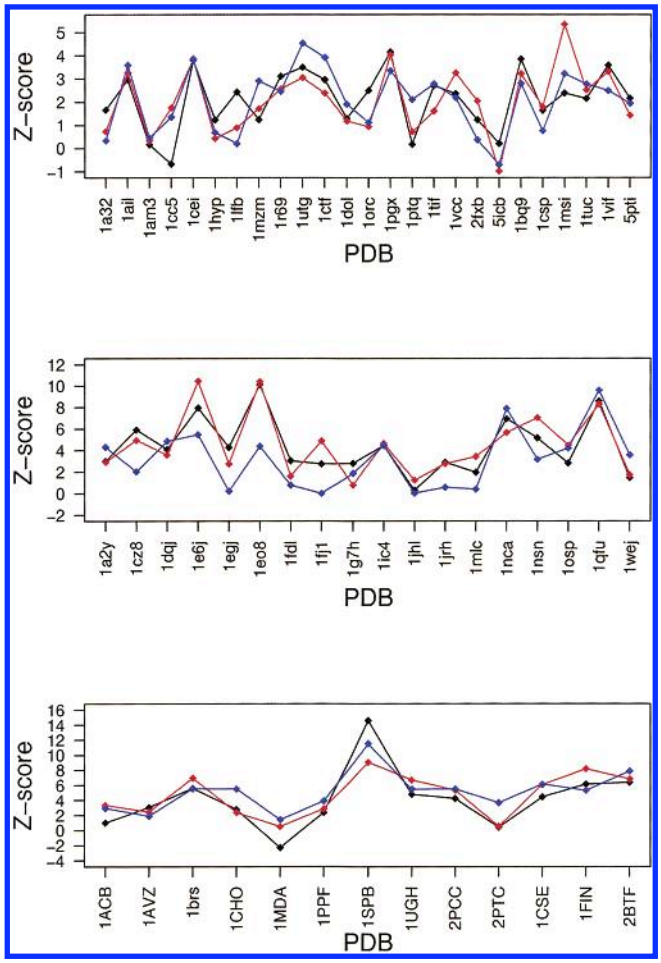


**Figure 6.** Native $Z$ scores of distance-dependent dielectric models versus GB model with optimized $\epsilon_i$ for different protein structures. Black, screened GB Coulomb ($\epsilon_i = 20$ for single-domain proteins, $\epsilon_i = 65$ for antibody−antigens; $\epsilon_i = 45$ for other protein−protein complexes); red, Warshel dielectric model; blue, Sternberg dielectric model. Upper plot, single-domain proteins; middle plot, antibody−antigen complexes; lower plot, other protein−protein complexes. AMBER atom types were used in the GB model; CHARMM19 atom types were used for effective dielectric models.

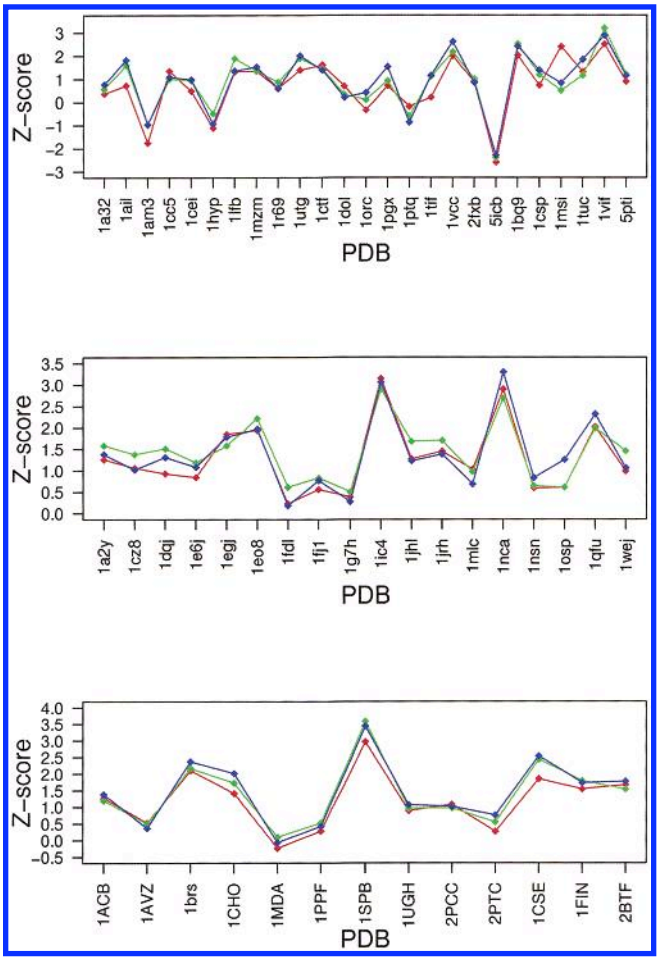


**Figure 7.** Native repacked $Z$ scores of Coulomb interactions with $\epsilon_i$ = const for different protein structures. Color code: red, CHARMM19 atom types; green, AMBER atom types; blue, OPLS atom types. Upper plot, single-domain proteins; middle plot, antibody−antigen complexes; lower plot, other protein−protein complexes.

groups of charged polar residues need be considered, possibly with a distance cutoff set to include only interactions of close pairs of residues of opposite charge.[47] Results shown in Figure 8 suggest, however, that this is not the general case; the interactions of all atoms and residues, perhaps with the exception of mainchain−mainchain interactions, contribute to the energy gap. Better performance in the all-atom case suggests that including partial charges on noncharged residues is preferable to treating them as totally neutral. This observation is also confirmed by excluding atoms participating in hydrogen bonds from electrostatic calculations; the drop in $Z$ scores is particularly striking for protein−protein interfaces.

**3.4. Hydrogen Bonding Potential and Combined Free Energy.** Even though hydrogen bonds are believed to be predominantly electrostatic in origin,[42] their directionality makes them similar to weak covalent bonds. This angular dependence is not captured using the electrostatic models described above. In this subsection, we discuss the results of applying the empirical hydrogen bonding potential we developed in refs 52 and 56 to our decoy sets (using the parameterization described in ref 56). We also investigate the extent to which decoy discrimination is improved by combining other free energy components with the hydrogen bonding potential.

On the single-domain protein set, we observe that mainchain−mainchain hydrogen bonds are the best discriminator of native structures (Table 3a; Table 4). The lack of discrimination by side chain−side chain and side chain−mainchain hydrogen bonds indicates that side chains of most decoys are repacked locally as well as those of the native structures (at least as far as the hydrogen bonding potential is concerned). The similarity between hydrogen bond native and native repacked $Z$ scores suggests that the number of rotamers was sufficient, because the same hydrogen bonding potential was used in the side chain repacking protocol applied to all decoys and native repacked structures. The difference between the two types of $Z$ scores is more pronounced for energies not included into (or down-weighed in) the rotamer repacking protocol, such as the Warshel electrostatics model or any of the PB and GB energies.

Can an improved model be generated by combining the orientation-dependent hydrogen bonding potential with electrostatics and van der Waals interactions? We used logistic regression to create a combined free energy capable of discriminating monomeric native and native repacked structures from decoys. Table 3a shows that a linear combination of the Warshel electrostatics model with hydrogen bonding energies is capable of discriminating 23 out of 25 structures in our X-RAY single-domain subset (for both native and native-repacked $Z$ scores; $Z$ score $< 1$ was considered a failure). On average, the $Z$ scores are higher than those of the combined free energy involving only the three types of hydrogen bonds.

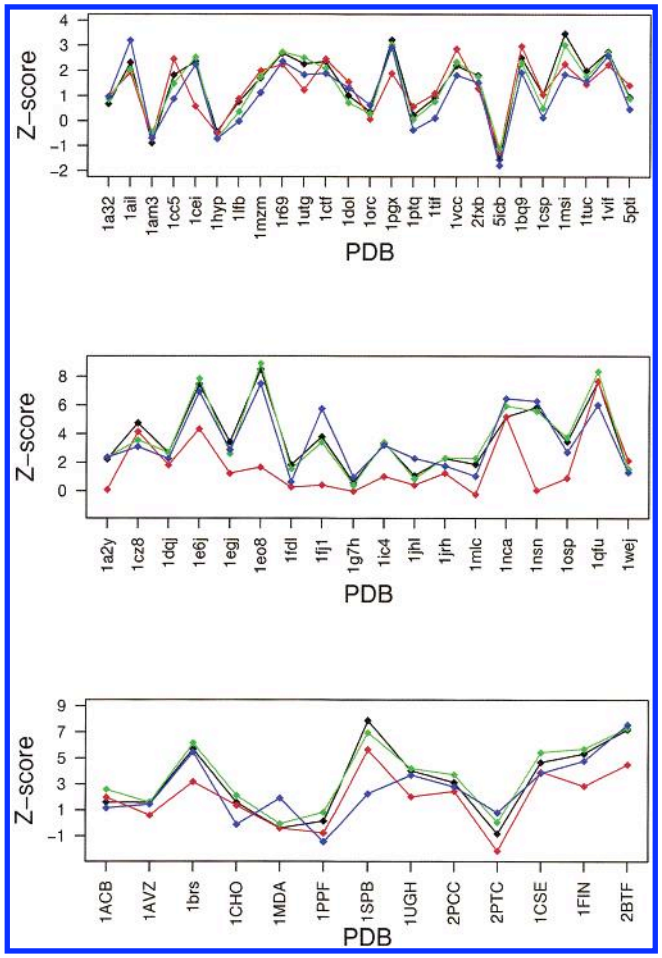


**Figure 8.** Native *Z* scores of Coulomb interactions with $\epsilon_i$ = const. Black, all atoms included; red, interactions involving Hbond-making atom pairs excluded; green, mainchain-mainchain interactions excluded; blue, only charged polar residue interactions (involving amino acids D,E,K,R) included. CHARMM19 force field atom types were used in this calculation. Upper plot, single-domain proteins; middle plot, antibody−antigen complexes; lower plot, other protein−protein complexes.

However, the main contribution is clearly due to the mainchain hydrogen bonds, with the Warshel model playing a secondary role. The Warshel model by itself fails in 7 (native) and 11 (native-repacked) out of 25 cases and is very sensitive in general to changing side chain conformations, which occurs for solvent-exposed residues during repacking of the native structures. When we include van der Waals interactions into the combined free energy, we observe an additional *Z*-score improvement, with no failures except for the native-repacked *1cc5* structure.

Next we looked at the low RMSD *Z* scores using a subset of single-domain proteins for which some ab initio nativelike decoys exist (see RMSD cutoffs in Table 1). We generate the same combined free energies as above and show the results in Tables 3b and 4. We are able to improve discrimination by adding van der Waals interactions and Warshel electrostatics into the free energy function, but the average *Z* scores are not high for most structures. The most likely reason for this is that we do not have enough nativelike structures in the ab initio monomeric decoy data set, so that the native funnel (Figure 1) is not reached by the decoys we designate as low RMSD. To test this hypothesis, we added perturbed-native structures (see Methods and Theory) to the ab initio decoy set and repeated Warshel electrostatics, hydrogen bonding, and logistic regression calculations on this new decoy set. We observed an increase of low RMSD *Z* scores for mainchain−mainchain hydrogen bonds and Warshel electrostatics, whereas side chain−side chain and side chain−mainchain hydrogen bond *Z* scores did not increase, probably because of the high degree of local optimization achieved for these energies by side chain repacking in the decoys. The free energy function including van der Waals, Warshel electrostatics, and hydrogen bonding interactions produces well-formed low RMSD funnels in 12 out of 23 cases (Table 4; versus 8 in the original set); 6 more have *Z* scores between 0.7 and 1.0. The average width of the folding funnel appears to be about 2 Å; if not enough structures are produced in this range, low RMSD *Z*-score discrimination is generally not possible.

We carried out a similar analysis on the protein−protein complex decoy set (Table 3c,d; Table 4), produced by rigid body perturbations of bound protein complexes. Although the protein backbones were taken from the bound protein−protein complex structure, all side chains were repacked to eliminate the information contained in the exact native conformation (see Methods and Theory). The combined free energy including a linear combination of side chain−side chain, side chain−mainchain, and mainchain−mainchain hydrogen bonds can reliably discriminate native and native-repacked structures in 26 out of 31 cases (Table 3c; Table 4). All three hydrogen bonding terms provide a significant contribution to the energy gaps. Because there are also sizable energy gaps between native structures and alternatively docked decoy conformations using the Warshel dielectric model, we expect the decoy discrimination to improve when the Warshel electrostatic energy is combined with the hydrogen bond free energy function, and indeed, in this case, we fail only once for native structures and three times for native-repacked ones. The addition of van der Waals interactions produces a further increase of the *Z* scores; this effect is especially dramatic for complexes other than antibody−antigen. The combined free energy including van der Waals, electrostatics, and hydrogen bonding terms discriminates all the native structures and fails three times when native side chains are repacked. In all of the failures, we observe low *Z* scores for the electrostatics and van der Waals components alone, so their addition to the hydrogen bonding potential does not result in a dramatic improvement.

Finally, we observe that low RMSD decoy discrimination is better with protein−protein complexes than it was in the single-domain case, because of a larger number of nativelike decoys available in the former data set. The hydrogen bonding terms are again a main contributor; we have only observed a gradual improvement upon adding extra terms to the free energy function. We have 10 failures out of 31 with the combined hydrogen bond free energy function (Table 3d; Table 4). This number drops to 7 when all of the extra terms are included, 2 of these being borderline cases with *Z* scores between 0.7 and 1.0.

## 4. Conclusions

In this paper, we evaluated continuum electrostatic models in proteins by considering electrostatic free energy gaps between native, nativelike, and non-native protein conformations, using both monomeric proteins and protein−protein complex data sets. Free energy gaps are necessary for discrimination of native structures and nativelike decoys from arbitrary compact conformations. Electrostatic free energies were computed using numerical finite-difference solutions to the PB equation; an analytical approximation to it provided by the GB model consistent with the AMBER force field; and pairwise-additive

models with effective distance-dependent dielectric constants. We also used an empirical hydrogen bonding potential developed in refs 52 and 56, by itself and in combination with van der Waals and electrostatic energies.

The total electrostatic energies obtained using either the PB or the GB approach do not produce large native or nativelike free energy gaps, because desolvation self-energies of charged atom burial typically disfavor native structures. In many cases, the desolvation self-energies of individual atoms appear to be sufficiently compensated by favorable screened Coulomb interactions in the protein interior; however, this is not true for backbone carbonyl carbons and amide hydrogens where the screened Coulomb interactions do not become more favorable with burial. This behavior suggests that effects ignored in the continuum models, such as induced polarization, protein dynamics, and the molecular nature of water, are not approximated correctly and might be sizable, particularly for backbone atoms.

The largest electrostatic free energy gaps using continuum electrostatic models are produced by the Coulomb interactions screened by solvent polarization, with unfavorable self-energies excluded. The dielectric constant in the protein interior ($\epsilon_i$) can be viewed as a variable parameter; it cannot be determined ab initio or even defined within the macroscopic approach.[1,3,4] The high values of $\epsilon_i$ we obtain in the process of maximizing average $Z$ scores on our decoy sets indicate that downweighted solvation contributions are preferable for decoy discrimination; in fact, the Coulomb term alone has performed nearly as well. Our results do not distinguish between the possibilities that (1) electrostatic solvation energies so strongly disfavor the native structure that the total electrostatic energy is, in reality, lower for the alternative (decoy) conformations or (2) the total electrostatic energy indeed favors the native structure, but this is not recaptured by current continuum electrostatics models because of imperfect balancing of the two large and opposing contributions.

Furthermore, simple distance-dependent dielectric models produce energy gaps similar to the screened Coulomb term of the more detailed GB approach. This suggests that the main physical effect captured by the appropriate GB components and distance-dependent models consists of gradual shielding of Coulomb interactions with increasing interatomic distances. This phenomenon is reproduced to some extent by simple analytical expressions of distance-dependent dielectrics.[47]

On a single domain protein decoy set, we were able to obtain significant energy gaps for native and repacked native structures when a distance-dependent dielectric model is combined with the hydrogen bonding and van der Waals interactions. The main contribution is provided by hydrogen bonding, with the other two terms assuming a secondary role. The orientation-dependent effective hydrogen bonding potential appears to be a better model of hydrogen bonds than the purely Coulomb description as it produces larger energy gaps (see also ref 52).

On a protein−protein complex decoy set, the free energy function with the same components discriminates the native structure in all cases, with three failures when side chains are repacked. There is also strong score-RMSD correlation in this case, which is detected by the hydrogen bonding potential alone and can be somewhat improved by combining van der Waals and electrostatics interactions with the hydrogen bonding potential. The combined free energy function is capable of very good low RMSD decoy discrimination (24 of 31 structures) and fails only when most of the decoys are too distant to be in a native funnel (as occurs in the ab initio single domain set; cf Figure 1, Table 1).

The tests carried out in this paper suggest areas for improvement of models of electrostatic interactions in proteins. In particular, improved descriptions should explicitly incorporate the orientation dependence of the hydrogen bond and better treat the delicate balance (Figure 4b) between the free energy cost of desolvating backbone polar atoms and the gain of favorable hydrogen bonding and electrostatic interactions, perhaps by explicitly modeling induced polarization effects along the backbone and in the protein interior.

**Acknowledgment.** We express gratitude to Jerry Tsai, Jeff Gray, and Stewart Moughon for their help with creating original single-domain and protein−protein complex decoy data sets, to Kira Misura for relaxing a subset of single-domain decoys and native structures, and to Chris Saunders for useful suggestions on the manuscript. We also thank Keith E. Laidig for his effective administration of the computational resources instrumental in completing the numerical part of these calculations. A.M. and D.B. were supported by the Howard Hughes Medical Institute. T.K. was supported by fellowships from the European Molecular Biology Organization and the Human Frontier Science Program Organization.

## References and Notes

(1) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283−422.
(2) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.
(3) Warshel, A.; Åqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267−298.
(4) Nakamura, H. *Q. Rev. Biophys.* **1996**, *29*, 1−90.
(5) Sheinerman, F. B.; Norel, R.; Honig, B. *Cur. Opin. Str. Biol.* **2000**, *10*, 153−159.
(6) Jackson, J. D. *Classical electrodynamics*; John Wiley & Sons: New York, 1975.
(7) Huang, K. *Statistical mechanics*; John Wiley & Sons: New York, 1987.
(8) Landau, L. D.; Lifshitz, E. M. *Electrodynamics of continuous media*; Pergamon: Oxford, 1984.
(9) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400−417.
(10) King, G.; Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366−4377.
(11) Nakamura, H.; Sakamoto, T.; Wada, A. *Prot. Eng.* **1988**, *2*, 177−183.
(12) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; Garcia-Moreno, B. *Biophys. J.* **2000**, *79*, 1610−1620.
(13) Honig, B.; Nichols, A. *Science* **1995**, *268*, 1144−1194.
(14) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.
(15) Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1987**, *9*, 327−335.
(16) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507−6514.
(17) Yang, A.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 351−365.
(18) Lee, L. P.; Tidor, B. *Nat. Struct. Biol.* **2001**, *8*, 73−76.
(19) Lee, L. P.; Tidor, B. *Prot. Sci.* **2001**, *10*, 362−377.
(20) Marshall, S. A.; Morgan, C. S.; Mayo, S. L. *J. Mol. Biol.* **2002**, *316*, 189−199.
(21) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 399−413.
(22) Vorobjev, Y. N.; Hermans, J. *Biophys. Chem.* **1999**, *78*, 195−205.
(23) Vorobjev, Y. N.; Hermans, J. *Protein Sci.* **2001**, *10*, 2498−2506, addendum in *Protein Sci.* **2002**, *11*, 994.
(24) Lee, M. R.; Tsai, J.; Baker, D.; Kollman, P. *J. Mol. Biol.* **2001**, *313*, 417−430.
(25) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(26) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(27) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.
(28) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(29) Dominy, B. N.; Brooks, C. L., III. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.
(30) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571−9576.

(31) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465−1471.
(32) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(33) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(34) Zou, X.; Sun, Y.; Kuntz, I. D. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.
(35) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591−607.
(36) Srinivasan, J.; Cheatham, T. E., III.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(37) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.
(38) Wallqvist, A.; Gallicchio, E.; Felts, A. K.; Levy, R. M. *Adv. Chem. Phys.* **2002**, *120*, 459−486.
(39) Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 404−422.
(40) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(41) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. *J. Mol. Biol.* **1997**, *272*, 106−120.
(42) Israelachvili, J. *Intermolecular and surface forces*; Academic Press: London, 1997.
(43) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6490−6498.
(44) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(45) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(46) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci.* **1987**, *84*, 7524−7528.
(47) Petrey, D.; Honig, B. *Prot. Sci.* **2000**, *9*, 2181−2191.
(48) Gatchell, D. W.; Dennis, S.; Vajda, S. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 518−534.
(49) Lazaridis, T.; Karplus, M. *J. Mol. Biol.* **1998**, *288*, 477−487.
(50) Hassan, S. A.; Mehler, E. L. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 45−61.
(51) Norel, R.; Sheinerman, F.; Petrey, D.; Honig, B. *Protein Sci.* **2001**, *10*, 2147−2161.
(52) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol.* In press.
(53) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(54) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(55) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(56) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci.* **2002**, *99*, 14116−14121.
(57) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci.* **2000**, *97*, 10383−10388.
(58) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(59) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, *268*, 209−225.
(60) Simons, K. T.; Ruczinski, I.; Kooperberg, C.; Fox, B. A.; Bystroff, C.; Baker, D. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 82−95.
(61) Tsai, J.; Bonneau, R.; Morozov, A. V.; Kuhlman, B.; Rohl, C.; Baker, D. *Proteins: Struct., Funct., Genet.* In press.
(62) Conte, L. L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177−2198.
(63) Gray, J. J.; Moughon, S.; Kortemme, T.; Schueler-Furman, O.; Misura, K. M. S.; Morozov, A. V.; Baker, D. *Proteins: Struct., Funct., Genet.* In press.
(64) Hao, M. H.; Scheraga, H. A. *Curr. Opin. Struct. Biol.* **1999**, *9*, 184−188.
(65) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426−434.
(66) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(67) Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946−950.
(68) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.