Efficient Evaluation of the Effective Dielectric Function of a Macromolecule in Aqueous Solution

URS HABERTHÜR, NICOLAS MAJEUX, PHILIPP WERNER, AMEDEO CAFLISCH
Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 17 February 2003; Accepted 17 April 2003

Abstract: We propose an analytical approach to calculate the effective dielectric function of proteins in aqueous solution. The screening effect if quantified by a measure of enclosure which is based on the distribution of solute atomic volumes around a pair of charges in a macromolecule. For protein conformations that vary significantly in size and shape, a comparison with finite difference Poisson calculations shows that pair interaction energies, their sums and solvation energies are well reproduced. The approach rivals the speed of simple distance dependent dielectric functions and the accuracy of the generalized Born model.

 J Comput Chem 24: 1936–1949, 2003

Key words: implicit solvent; continuum dielectric; analytical electrostatic interaction (AEI); generalized Born approach; solvation energy; effective dielectric constant

Introduction

Incorporating solvent effects in molecular dynamics (MD) and Monte Carlo simulations is of key importance to quantitatively understand chemical and physical properties of biomolecular processes. Accurate electrostatic energies of proteins in an aqueous environment are one indispensable component to discriminate between native and non-native conformations. An exact evaluation of electrostatic energies considers the interactions among all possible solute-solute, solute-solvent, and solvent-solvent pairs of charges. However, this is computationally expensive for macromolecules. Continuum dielectric approximations offer a more tractable approach.[1–5] The essential concept in continuum models is to represent the solvent by a high dielectric medium, which eliminates the solvent degrees of freedom, and to describe the macromolecule as a region with a low dielectric constant and a spatial charge distribution. The Poisson equation provides an exact description of such a system. The increase in computation speed for a finite difference solution of the Poisson equation,[6–9] with respect to an explicit treatment of the solvent, is remarkable, but still not enough for effective utilization in computer simulations of macromolecules. The generalized Born (GB) model was introduced to facilitate an efficient evaluation of continuum electrostatic energies.[10] It provides accurate energetics and the most efficient implementations are between five to ten times slower than *in vacuo* simulations.[11–13] The essential element of the GB approach is the calculation of an effective Born radius for each atom in the system, which is a measure of how deeply the atom is buried inside the protein. This information is combined in a heuristic way to obtain a correction to the Coulomb law for each atom pair.[10] For the integration of energy density, necessary to obtain the effective Born radii, both numerical[10,11,13] and analytical[13–16] implementations exist. The former are more accurate but slower than the latter.[13] Moreover, analytical derivatives that are required for MD simulations are not given by numerical implementations.

For efficiency reasons empirical dielectric screening functions are the most common choice in MD simulations with implicit solvent. One kind of solvation model is based on the use of a dielectric function that depends linearly on the distance r between two charges [$\varepsilon(r) = \alpha r$][17,18] or has a sigmoidal shape.[19–22] Although very fast, these options suffer from their inability to discriminate between buried and solvent exposed regions of a macromolecule and are therefore rather inaccurate. Recently, a distance- and exposure-dependent dielectric function was proposed.[23]

The aim of this article is to give an analytical approximation of the effective dielectric screening function that rivals the speed of $\varepsilon(r) = \alpha r$ and the accuracy of the GB model. A measure of

Correspondence to: A. Caflisch

Contract/grant sponsor: Swiss National Competence Center in Structural Biology (NCCR)

Contract/grant sponsor: Swiss National Science Foundation; contract/grant number: 31-64968.01 (A.C.)

enclosure that focuses directly on atom pairs and their neighborhoods is introduced. It provides an approximate description of where the atom pair is located with respect to the bulk of the macromolecule and the solvent. A fit to effective dielectric constants derived from finite difference Poisson (fdP) energies for a set of several protein structures supplies analytical functions with continuous derivatives. The question of transferability and predictive power of the model presented here, henceforth called the analytical electrostatic interaction (AEI) model, is addressed by dividing the set of protein structures into several training and test sets. Various comparisons with electrostatic energies calculated by fdP, the GB approach,[13] and the sigmoidal distance dependent dielectric (SGM) model[20] are given. Finally, the physical relevance of the measure of enclosure is analyzed by comparing AEI with fdP solvation energies.

Methods

AEI Model

Theory

Consider a macromolecule in a fixed configuration immersed in a polar solvent with zero ionic strength. The Poisson equation

$$\nabla[\varepsilon(\vec{x})\nabla\phi(\vec{x})] = -4\pi\rho(\vec{x}) \tag{1}$$

defines the electrostatic potential ϕ given the dielectric function ε and the charge density ρ. In the continuum approximation used in all following calculations $\varepsilon(\vec{x}) = \varepsilon_m$ for the region of the macromolecule and $\varepsilon(\vec{x}) = \varepsilon_s$ for the region of the solvent. The effective dielectric constant $\varepsilon_{ij}^{\mathrm{fdP,eff}}$ for each pair of atoms i and j is defined such that if substituted into the Coulomb law the same electrostatic interaction energy results as when solving the Poisson equation:

$$q_i q_j \phi_i^{\mathrm{fdP}}(\vec{x}_j) =: \frac{q_i q_j}{r_{ij}\varepsilon_{ij}^{\mathrm{fdP,eff}}} \Leftrightarrow \varepsilon_{ij}^{\mathrm{fdP,eff}} := \frac{1}{r_{ij}\phi_i^{\mathrm{fdP}}(\vec{x}_j)} \tag{2}$$

where ϕ_i^{fdP} is the electrostatic potential of a unit charge at the position of atom i; q_i and q_j denote the charges of atoms i and j, respectively; $\vec{x}_j$ represents the position of atom j, and r_{ij} the distance between atoms i and j. Note that $q_i q_j \phi_i^{\mathrm{fdP}}(\vec{x}_j)$ is the electrostatic interaction energy of the (i, j) pair in the presence of solvent.

For an accurate approximation of $\varepsilon_{ij}^{\mathrm{fdP,eff}}$ it is necessary to discriminate between buried and solvent exposed atoms in the macromolecule. In the GB approach[10] the effective dielectric constant $\varepsilon_{ij}^{\mathrm{GB,eff}}$ is defined by

$$\frac{1}{\varepsilon_{ij}^{\mathrm{GB,eff}}} = \frac{1}{\varepsilon_m} - \left(\frac{1}{\varepsilon_m} - \frac{1}{\varepsilon_s}\right)\left(1 + \frac{R_i R_j}{r_{ij}^2} e^{-(1/4)(r_{ij}^2/R_i R_j)}\right)^{-1/2} \tag{3}$$

$$= \frac{1}{\varepsilon_m} - \left(\frac{1}{\varepsilon_m} - \frac{1}{\varepsilon_s}\right)\left(1 + \left(\frac{u_{ij}^{\mathrm{GB}}}{r_{ij}}\right)^2 e^{-(1/4)(r_{ij}/u_{ij}^{\mathrm{GB}})^2}\right)^{-1/2} = f\left(\frac{u_{ij}^{\mathrm{GB}}}{r_{ij}}\right) \tag{4}$$

where R_i and R_j denote the effective Born radii of atoms i and j, respectively, $u_{ij}^{\mathrm{GB}} = \sqrt{R_i R_j}$, and the function f is defined by $f(x) = 1/\varepsilon_m - (1/\varepsilon_m - 1/\varepsilon_s)/\sqrt{1 + x^2 e^{-1/4x^2}}$. The effective Born radius of an atom in the system is a measure of its enclosure. Consequently, the quantity u_{ij}^{GB} could be interpreted as a measure of enclosure of the (i, j) atom pair: the larger u_{ij}^{GB} is the more buried the (i, j) pair is. Because the calculation of the effective Born radii is the time consuming part in the GB approach and because this article is mainly concerned with interaction energies, we seek an alternative way to quantify the degree of enclosure of an atom pair in the macromolecule. We introduce a new and computationally efficient measure of enclosure u_{ij}^{AEI} and approximate $\varepsilon_{ij}^{\mathrm{fdP,eff}}$ in the same spirit as in the GB model by a function of $u_{ij}^{\mathrm{AEI}}/r_{ij}$, that is, $1/\varepsilon_{ij}^{\mathrm{fdP,eff}} \cong 1/\varepsilon_{ij}^{\mathrm{AEI,eff}} = g(u_{ij}^{\mathrm{AEI}}/r_{ij})$.

The present approach to calculate a measure of enclosure u_{ij}^{AEI} focuses on a finite region Ω_{ij} of space. This region is chosen around atoms i and j, and is large enough to neglect effects on $\varepsilon_{ij}^{\mathrm{fdP,eff}}$ due to conformational changes outside Ω_{ij}. The exact shape of this region is not important for the following arguments. One could, for instance, imagine a cylinder with an axis along the line joining atoms i and j or a sphere or an ellipsoid with its center somewhere between the two atoms. If only atoms i and j of the macromolecule were present within Ω_{ij}, solving the Poisson equation would result in $\varepsilon_{ij}^{\mathrm{fdP,eff}} \cong \varepsilon_s$ and u_{ij}^{AEI} is required to be small. As more and more atoms are gradually added, $\varepsilon_{ij}^{\mathrm{fdP,eff}}$ decreases and u_{ij}^{AEI} increases in a complex way depending on where the additional atoms are placed. When all the solvent has finally been flushed out from Ω_{ij}, solving the Poisson equation would result in $\varepsilon_{ij}^{\mathrm{fdP,eff}} \cong \varepsilon_m$ and u_{ij}^{AEI} reaches its maximum value. Intuitively, one expects that atoms located near or between charges i and j increase u_{ij}^{AEI} more than atoms located far from the (i, j) pair because the closer an atom is placed to atoms i and j, the more it influences the electric field at their positions.[24,25] Furthermore, adding a large atom is expected to increase u_{ij}^{AEI} more than adding a small one because more solvent is displaced.

The above arguments suggest quantification of the degree of enclosure of the (i, j) atom pair by a function that depends on the sum of the van der Waals volumes within Ω_{ij}, which are weighted according to their positions with respect to atoms i and j. In the GB approach the measure of enclosure u_{ij}^{GB} is the square root of the product of the effective Born radii of atoms i and j. In the AEI model u_{ij}^{AEI} is the square of a sum of weighted van der Waals volumes located around the atom pair. While there are many methods of calculating a weighted sum, the necessity for low computational costs eliminates most of them. For instance, it is not feasible to construct a cylinder around each atom pair and calculate a weighted sum within such a cylinder. We will only use quantities already available in the course of an MD simulation.

Two spheres of radius r_{sphere} with centers at the positions of atoms i and j define Ω_{ij}. Let A denote the set of all atoms with their centers within the sphere around atom i. Note that atom i belongs to A. Let B denote the corresponding set of atoms for the sphere around atom j. Let v_k be the van der Waals volume of any atom k and N the total number of atoms of the macromolecule. Then u_{ij}^{AEI} is defined by the square of a sum of weighted van der Waals volumes of the atoms in A and B (see Fig. 1):

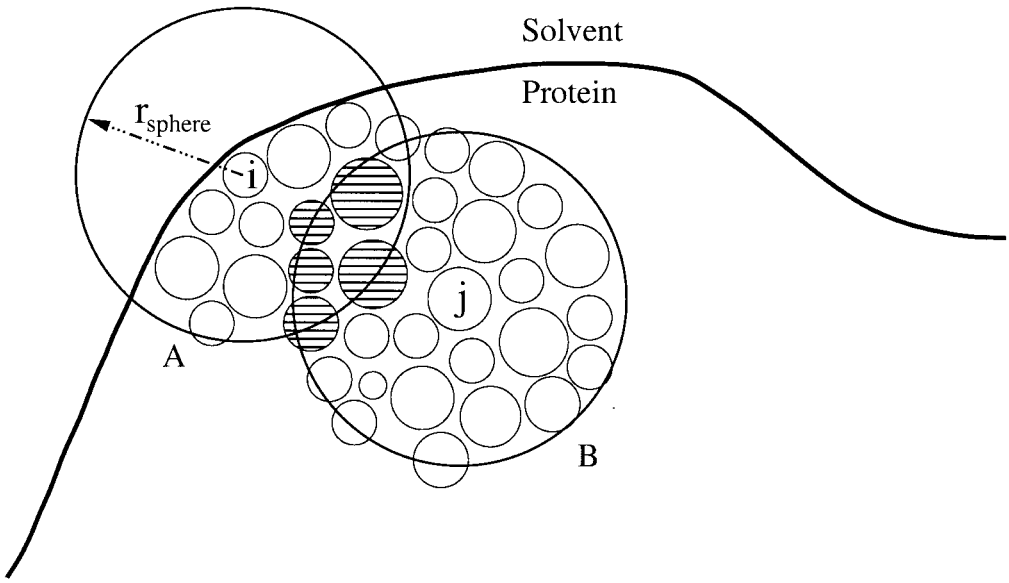

Figure 1. Schematic illustration for the calculation of the measure of enclosure u_{ij}^{AEI} in the case $r_{\mathrm{sphere}} < r_{ij} < 2r_{\mathrm{sphere}}$. The small circles describe protein atoms. The two large circles represent the spheres that define the neighborhood of atoms i and j, which are taken into account to evaluate u_{ij}^{AEI}. Atoms within the large spheres around atoms i and j constitute sets A and B, respectively. The shaded circles represent the atoms in the intersection of A and B. They are weighted with respect to the positions of both atoms i and j. The atoms described by the small empty circles are only weighted with respect to either atom i or atom j.

$$u_{ij}^{\mathrm{AEI}} := \left(\sum_{k=1}^{N} v_k \Theta_{ik} + \sum_{k=1}^{N} v_k \Theta_{jk} \right)^2 \tag{5}$$

$$= \left(\sum_{k \in A \setminus B} v_k \Theta_{ik} + \sum_{k \in B \setminus A} v_k \Theta_{jk} + \sum_{k \in A \cap B} v_k (\Theta_{ik} + \Theta_{jk}) \right)^2 \tag{6}$$

where the weighting function Θ_{ik} is defined by

$$\Theta_{ik} := \begin{cases} \left(1 - \left(\frac{r_{ik}}{r_{\mathrm{sphere}}}\right)^2\right)^2 & r_{ik} < r_{\mathrm{sphere}} \\ 0 & r_{ik} \geq r_{\mathrm{sphere}} \end{cases} \tag{7}$$

The first two sums in eq. (6) include all atoms in A and B that are not in the intersection of A and B. The volumes of these atoms are weighted with respect to the position of either atom i or atom j. The third sum in eq. (6) includes all atoms in the intersection of A and B. The volumes of these atoms are weighted with respect to the positions of both atoms i and j. The weighting function Θ_{ik} assures that the further away an atom k is placed from atom i, the lower its weight and the less it contributes to the sum. Furthermore, the existence of continuous derivatives (required for MD simulations) is guaranteed. Note that Θ_{ik} is the shifting function that is commonly used in MD simulations to have zero Coulomb interaction energy at the cutoff.[26] Only if $r_{ij} < 2r_{\mathrm{sphere}}$ do the spheres around atoms i and j overlap. If $r_{ij} \geq 2r_{\mathrm{sphere}}$, each atom is only weighted with respect to the position of either atom i or atom j, but in this case the distance r_{ij} is large, thus the interaction energy is small and a more crude approximation justified. Following the arguments mentioned previously

$$\frac{1}{\varepsilon_{ij}^{\mathrm{fdP,eff}}} \cong g\left(\frac{u_{ij}^{\mathrm{AEI}}}{r_{ij}}\right) \tag{8}$$

and the function g has to be chosen so as to approximate $1/\varepsilon_{ij}^{\mathrm{fdP,eff}}$ as accurately as possible. The calculation of the measure of enclosure u_{ij}^{AEI} [see eqs. (5) and (7)] can be performed very efficiently. Building a list of atoms within a sphere of radius r_{sphere} around each atom in the system is intrinsic to MD simulations. The same is true of the shifting function and its derivative. Because the derivation given in this section is heuristic, the approach is ultimately only justified if the results are satisfactory.

Determination of the Function g

For a training set of structures the function g in eq. (8) was determined by fitting it to the inverse of effective dielectric constants derived from fdP energies [see eq. (2)]. The performance was assessed by comparing interaction energies calculated by the AEI model and by fdP for the conformations in a test set. Of particular interest was whether or not good performance of both folded and unfolded states for peptides and larger proteins could be achieved.

An initial set of 29 proteins (23 single and six multichain proteins) of very different sizes and shapes was used. The struc-

Table 1. Definitions of the Seven Test Cases that Are Used to Perform Cross Correlations.

	Training set		Test set	
Test case	Type of structures	Number	Type of structures	Number
a	All	52	All	52
b	Randomly selected	26	The complementary set	26
c	Test set (b)	26	Training set (b)	26
d	Native	29	Unfolded	23
e	Unfolded	23	Native	29
f	Less than 70 amino acids	27	More than 70 amino acids	25
g	More than 70 amino acids	25	Less than 70 amino acids	27

A test case consists of a training and a test set. For each test case the parameters of the AEI model are derived from the training set and used to calculate interaction energies for the structures in the test set. Apart from test case (a), the training and test sets are disjointed.

tures ranged in size from 11 (1cb3) to 347 (3pte) amino acids. The set included almost spherical geometries with no microcavities, as well as structures with internal cavities. 5hvp, for instance, is the HIV-1 aspartic proteinase in a complex with a peptidic ligand that was removed from the active site to obtain an internal cavity. To further diversify the set of structures with many different kinds of irregular shapes (cavities, open loops, etc.), the single chain proteins were subjected to high temperature unfolding simulations at 450 K for 20 ns with an implicit solvation model.[27] From each trajectory an unfolded conformation was selected and added to the initial set of structures. The average increase in the radius of gyration of the chosen conformations was 32% and their average C_α-RMSD was 12.8 Å. The final set consisted of 52 conformations. All atoms (a total of 47,979 atoms) were assigned unit charges, and all pair interaction energies for every conformation in the set of the 52 conformations (a total of 39,041,961 pairs) were calculated by numerical (finite difference) solution of the Poisson equation.

The set of 52 conformations was divided in seven different ways into a training and a test set in order to perform cross correlations (Table 1). While cases (a), (b), and (c) addressed the convergence of the parameterization in general, cases (d) and (e) investigated how well the parameters extrapolate to different shapes, and cases (f) and (g) investigated how well the parameters extrapolate to different sizes. Note that apart from case (a), the training and test sets are disjointed.

Given a specific training set, $g(u_{ij}^{\mathrm{AEI}}/r_{ij})$ was fitted to $1/\varepsilon_{ij}^{\mathrm{fdP,eff}}$ rather than a function $\tilde{g}(u_{ij}^{\mathrm{AEI}}/r_{ij})$ fitted to $\varepsilon_{ij}^{\mathrm{fdP,eff}}$ in order to obtain accurate values for small effective dielectric constants (only these can result in high energies). Three different cases were distinguished: 1-2 pairs, 1-3 pairs, and all other pairs, that is

$$\frac{1}{\varepsilon_{ij}^{\mathrm{fdP,eff}}} \cong g_k\left(\frac{u_{ij}^{\mathrm{AEI}}}{r_{ij}}\right) \tag{9}$$

where $k = 1$ for 1-2 pairs, $k = 2$ for 1-3 pairs, and $k = 3$ for all remaining pairs. (A 1-2 pair consists of two covalently bonded atoms and a 1-3 pair of two atoms covalently bonded to a common atom.) For each of the three cases the range of the variable $u_{ij}^{\mathrm{AEI}}/r_{ij}$ was divided into 100 bins and the average of all $1/\varepsilon_{ij}^{\mathrm{fdP,eff}}$ values within a given bin was calculated. The functions g_k were obtained by fitting analytical functions of the form of f in eq. (4) to the average curves (see Appendix). Note that the functions g_k have continuous derivatives.

The measure of enclosure u_{ij}^{AEI}, defined by eqs. (5) and (7), and the analytical functions g_k given in the Appendix, are the main results of this article and constitute the AEI model. They were used to calculate electrostatic interaction energies E_{ij} for solute charges immersed in solvent by the formula

$$E_{ij} = 332\,\frac{q_i q_j}{r_{ij}}\,g_k\left(\frac{u_{ij}^{\mathrm{AEI}}}{r_{ij}}\right) \tag{10}$$

where the factor 332 was introduced to obtain values in kcal/mol. Note that only g_3 is relevant for MD simulations because the interaction energies of 1-2 and 1-3 pairs are accounted for in the bonding terms of the force fields. Results are also presented for 1-2 and 1-3 pairs to show that the approach is valid in general. The calculation of solvation energies within the framework of the AEI model is outlined in the section Solvation Energies.

Finite Difference Poisson

The numerical (finite difference) solution of the Poisson equation was calculated with the PBEQ module[28] in CHARMM.[26] A grid spacing of 0.3 Å was used. (Some calculations with a grid spacing of 0.2 Å were also performed. Relative errors of interaction energies for a grid spacing of 0.3 Å with respect to a grid spacing of 0.2 Å are on average only about 0.55%.) The dielectric discontinuity surface was defined by the molecular surface. This is the surface spanned by the surface of a solvent probe sphere of radius 1.4 Å rolled over the van der Waals envelope of the atoms. The molecular volume was treated as a dielectric medium with a low dielectric constant $\varepsilon_m = 1$. Together with the spatial charge distribution of the macromolecule it represents the solute. The remaining space was treated as a dielectric medium with a high dielectric constant $\varepsilon_s = 78.5$ and represents the solvent. Solvation energies were calculated by subtracting the *vacuo* self-energy

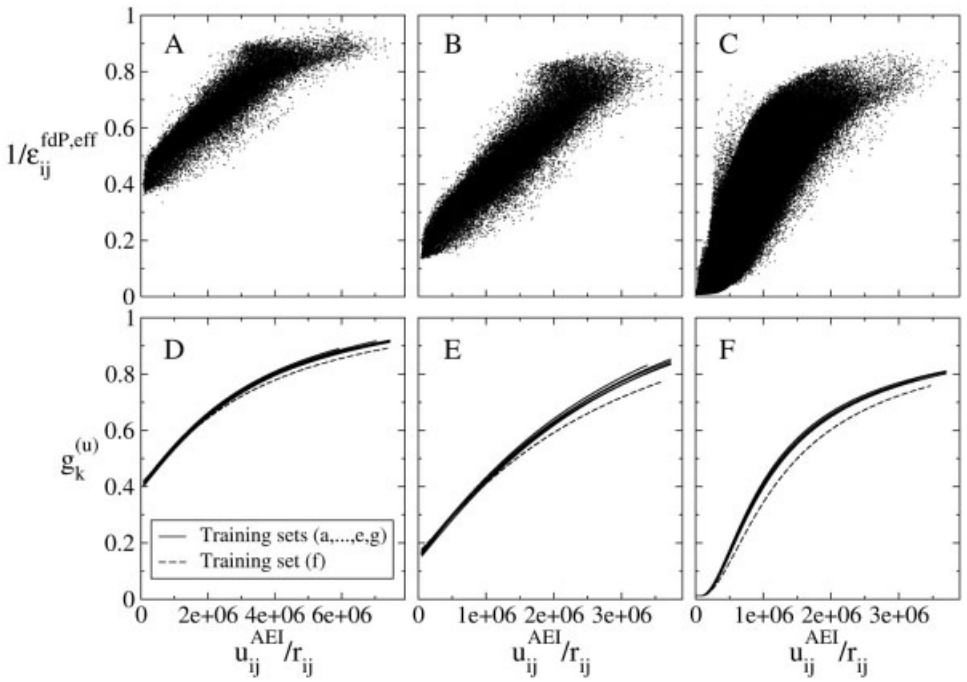

Figure 2. Top: inverse of fdP-derived effective dielectric constants against u_{ij}^{AEI}/r_{ij} for 1-2 pairs (A), 1-3 pairs (B), and all remaining pairs (C) of all 52 conformations. Bottom: analytical functions resulting from the fits (see text) for 1-2 pairs (D), 1-3 pairs (E), and all remaining pairs (F). The solid lines represent the fits for training sets (a) to (e) and (g). The dotted lines represent the fits for training set (f).

($\varepsilon_m = 1$, $\varepsilon_s = 1$) from the self-energy in solution ($\varepsilon_m = 1$, $\varepsilon_s = 78.5$). Note that using $\varepsilon_m = 1$ (instead of $\varepsilon_m = 2$ or $\varepsilon_m = 4$, for instance) is the most stringent test for the accuracy of the AEI model. For single-point energy calculations (e.g., for ranking in ligand binding), $\varepsilon_m > 1$ would be more appropriate because it accounts for thermal fluctuations. As the AEI model is primarily aimed to be used in MD simulations, the more stringent test with $\varepsilon_m = 1$ was chosen for the present validation.

GB

The GB calculations were performed with the analytical implementation of the GBMV module[13] in CHARMM. The dielectric discontinuity surface and the dielectric constants were defined as in the fdP calculations. Note that the analytical GBMV reproduces the Poisson-derived Born radii with an accuracy of about 2–4% and with a correlation of about 0.95.[13] In previous analytical implementations of the GB model there is cancellation of errors because the Coulomb approximation tends to overestimate the effective Born radii, whereas the analytical approximation of the energy density integration tends to underestimate them.[29]

SGM Function

The SGM model is based on the sigmoidal function[19–21]:

$$\varepsilon^{SGM}(r_{ij}) = A + \frac{B}{1 + \alpha e^{-\beta r_{ij}}} \tag{11}$$

which is similar to the one used recently.[22] The parameters A and B were determined by the conditions $\lim_{r_{ij}\to 0}\varepsilon^{SGM}(r_{ij}) = \varepsilon_m = 1$ and $\lim_{r_{ij}\to\infty}\varepsilon^{SGM}(r_{ij}) = \varepsilon_s = 78.5$. The two remaining parameters α and β were determined by optimizing ε^{SGM} for each of the 52 conformations (separately because of memory requirements) in such a way that fdP interaction energies were reproduced as accurately as possible. Finally, the values for α and β were averaged over the 52 conformations. The resulting parameters were: $\alpha = 60.868$, $\beta = 0.317541$, and $A = -0.273247$, $B = 78.773247$. Note that A and B are such that if interaction energies are calculated by $E_{ij} = 332 q_1 q_2 / r_{ij} \varepsilon^{SGM}(r_{ij})$, the units are kcal/mol.

Parameter Set

All calculations were performed using the van der Waals radii and partial charges of the CHARMM parameter set PARAM19.[26] For the fdP calculations and certain tests (see below) all atoms were assigned unit charges. The CHARMM parameter set PARAM19 treats hydrogens covalently bound to carbons implicitly and polar hydrogens explicitly.

Results and Discussion

For all seven training sets (see Table 1) the functions g_k with $k \in \{1, 2, 3\}$ are determined following the prescription in the section Determination of the Function g. They are denoted by $g_k^{(u)}$ with

Table 2. Cross Correlation Data for Pair Interaction Energies E_{ij} Calculated by the AEI Model and by fdP.

Test case	Correlation	Slope	RMSD
a	0.976	1.033 (0.081)	1.400
b	0.972	0.977 (0.082)	1.383
c	0.980	1.085 (0.091)	1.469
d	0.972	0.981 (0.054)	1.072
e	0.980	1.058 (0.089)	1.576
f	0.978	0.851 (0.149)	1.065
g	0.974	1.087 (0.121)	1.861

For each test case defined in Table 1 the parameters of the AEI model are fitted to the data extracted from the structures in the training set and used to calculate interaction energies for the structures in the test set. A sphere radius $r_{\text{sphere}} = 8.5$ Å is used. Correlation, slope, and RMSD with respect to fdP data are averaged over the conformations in the test set. The unsigned deviations of the slopes from 1, averaged over the conformations in the test set, are shown in parentheses. All atoms are assigned unit charges and 1-2 and 1-3 pairs are excluded. The unit of the RMSD is kcal/mol.

$u \in \{a, b, \ldots, g\}$. Excluding the training set $u = f$, the maximal deviation between any two $g_k^{(u)}$ is 0.0177 for $k = 1$, 0.0287 for $k = 2$, and 0.0229 for $k = 3$. This implies that these curves are very close to each other and basically overlap (see Fig. 2). Only $g_k^{(f)}$ differs significantly from the other curves. The maximal deviation between $g_k^{(f)}$ and all other fits for k values of 1, 2, and 3 is 0.0367, 0.0774, and 0.0750, respectively. However, $g_k^{(f)}$ is expected to be an outlier: in small structures, most of the atoms are exposed to the solvent so that nearly all interactions experience large screening, that is, the average screening is higher than for a training set, which also includes structures with a large hydrophobic core. Therefore, for a given value of $u_{ij}^{\text{AEI}}/r_{ij}$, the average curve is biased towards high effective dielectric constants.

For each test case listed in Table 1 all pair interaction energies for every structure in the test set are calculated for unit charges, using the $g_k^{(u)}$ obtained by the fit based on the corresponding training set. The correlation, slope, and RMSD with respect to fdP data are determined for each structure in every test set. The results are summarized in Table 2. Note that only the results for $k = 3$ are shown, that is, 1-2 and 1-3 pairs are excluded. Taking all pairs into consideration merely improves results and is a less stringent test. None of the correlations is below 0.97 and apart from case (f), all slopes are close to 1 with the overall tendency being to overestimate rather than to underestimate energies. Applying $g_k^{(f)}$ (fit on proteins with less than 70 amino acids) to proteins with more than 70 amino acids [test set (f)] gives a slope of 0.85. The fit $g_k^{(f)}$ underestimates energies for large structures as it overestimates the average screening. The reverse effect, albeit less significant, can be observed for $g_k^{(g)}$. The fit on the training set consisting of proteins with more than 70 amino acids [training set (g)] slightly underestimates screening on average for the small structures. Because the structures in training set (g) include both buried and exposed atoms, the effect is hardly perceivable.

It is clear from the fits shown in Figure 2 and the cross correlation data given in Table 2 that far less than all the 52 structures are sufficient for the parameter optimization to converge [see test cases (a), (b), and (c)]. Furthermore, fitting on only folded, unfolded, or large structures does not introduce any bias [see test cases (d), (e), and (g)]. The model is highly independent of the shape of the proteins in the training set. However, a training set consisting of only small structures is not appropriate as it overestimates screening on average [see test case (f)].

The above analysis was carried out for 17 different values of the sphere radius r_{sphere}, namely $r_{\text{sphere}} = 6.0$ Å up to $r_{\text{sphere}} = 14.0$ Å with a step size of 0.5 Å. A value of $r_{\text{sphere}} = 8.5$ Å was found to perform best, but the model does not depend strongly on the radius. The results with a radius in the range from 7.5 to 9.0 Å differ only slightly. Clearly, a too small or too large sphere radius no longer discriminates whether an atom pair is in the bulk or on the surface, but there seems to be a relatively large range where this information is captured satisfactorily. Moreover, several combinations of different definitions of Θ_{ik} and different exponents for u_{ij}^{AEI} [i.e., $(u_{ij}^{\text{AEI}})^{a/2}$ and a was varied from 1.0 to 2.0 with a step size of 0.25] were investigated. Indeed, there are combinations that perform slightly better than the option presented here, but the marginal gain in accuracy is not considered to be worth the additional complexity in the formulas. In addition, a different weighting function is no longer intrinsic to the calculations in MD simulations. For all the following calculations, the functions g_k derived from the fit on all conformations [training set (a)] with a sphere radius $r_{\text{sphere}} = 8.5$ Å will be used.

Comparison Between the AEI, GB, and SGM Models

In the point charge approximation the total electrostatic interaction energy of a macromolecule in aqueous solution is

$$E_{\text{elec}}^{\text{inter}} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} E_{ij} \tag{12}$$

where N is the number of charges in the system and E_{ij} the screened Coulomb interaction energy of the (i, j) pair. In the following sections, energies calculated by the AEI, GB, and SGM models are compared with the appropriate fdP values. In the next section, pair interaction energies E_{ij} are analyzed, while in the following sections the sum of all interaction energies of atom i, $\sum_{j \neq i} E_{ij}$, and the total electrostatic interaction energy of the macromolecule, $E_{\text{elec}}^{\text{inter}}$, are discussed. The sums are useful to investigate cancellation of errors.

Pair Interaction Energies E_{ij}

Pair interaction energies E_{ij} are calculated by the AEI, GB, and SGM models for each of the 52 conformations. The correlation, slope, and RMSD with respect to fdP data are evaluated and the results are shown in Figure 3 and Table 3. Unit charges are assigned to all atoms, and two sets of pairs are distinguished: all pairs and all but 1-2 and 1-3 pairs. Both the AEI and GB models perform distinctly better than the SGM model. The AEI and GB models show similar accuracy, and for most of the conformations the GB model has only a marginal advantage. However, the range in the correlation, slope, and RMSD are larger for the AEI than for

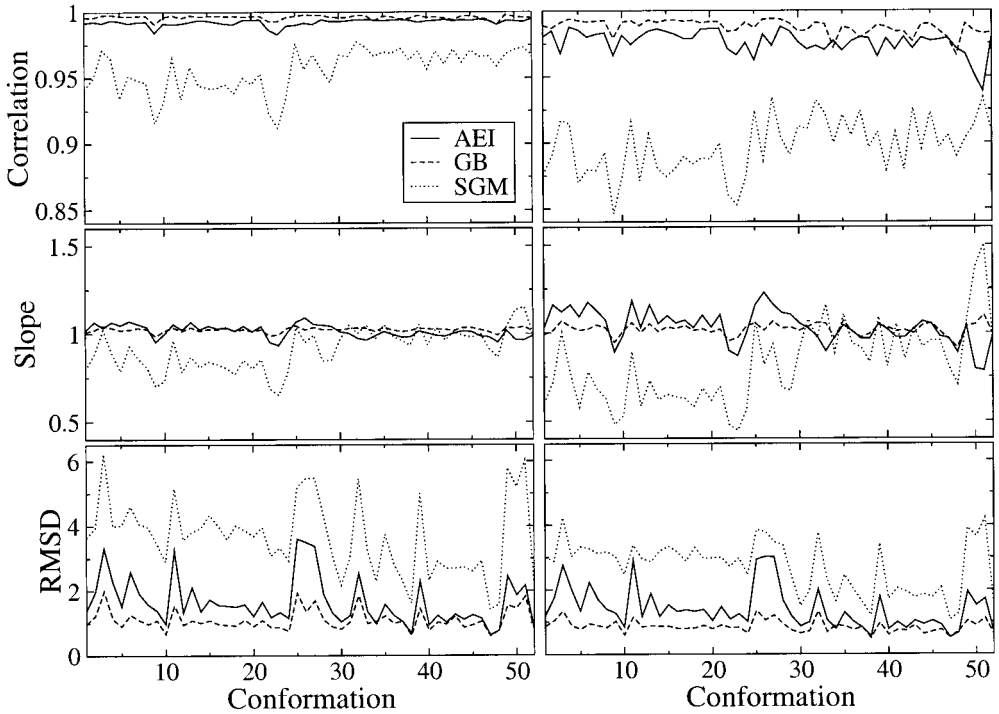

Figure 3. Pair interaction energies E_{ij}, calculated by the AEI [fit on training set (a), $r_{\text{sphere}} = 8.5$ Å], GB, and SGM models, are compared to the corresponding fdP values. Correlation, slope, and RMSD are shown for each of the 52 conformations. Unit charges are assigned to all atoms and energy values are in kcal/mol. The data presented on the left hand column include all pairs and the data on the right hand column all but 1-2 and 1-3 pairs. The conformations are ordered such that the folded ones (conformations 1 to 29) precede the unfolded ones (conformations 30 to 52).

the GB model. In particular, the slopes vary more (see Table 3), but most of the single structure values are close to the mean value and the extremes are gathered in few conformations (see Fig. 3). From the data presented in Figure 3 one can deduce that both the AEI and GB models tend to overestimate interaction energies, that is, underestimate the screening effect.

A closer inspection of the data and structures reveals that conformations where most residues are exposed are less accurately represented in the AEI model compared to its average performance. Figure 4 shows pair interaction energies E_{ij} as calculated by the AEI model against fdP data for a folded protein (2ins) and an unfolded structure [originating from a helix cut out from protein G (1pgb), named hlx1 in this article] that has the poorest correlation in the set of the 52 conformations. Note that the results for most of the structures are similar to 2ins. A strongly extended conformation of 17 residues (that is not in the set of the 52 conformations) gives a

Table 3. Minimal, Maximal, and Average Values of the Data Shown in Figure 3.

	Correlation			Slope			RMSD		
Model	Min	Max	Ave	Min	Max	Ave	Min	Max	Ave
All pairs									
AEI	0.983	0.995	0.992	0.933	1.092	0.032	0.626	3.603	1.649
GB	0.990	0.998	0.996	0.986	1.045	0.025	0.603	1.992	1.102
SGM	0.912	0.977	0.956	0.651	1.152	0.113	1.410	6.230	3.661
All but 1-2 and 1-3 pairs									
AEI	0.939	0.988	0.976	0.784	1.228	0.081	0.513	3.019	1.400
GB	0.966	0.994	0.988	0.915	1.100	0.032	0.492	1.349	0.854
SGM	0.846	0.934	0.897	0.444	1.502	0.240	1.068	4.230	2.712

The average values shown in the slope column are the averages of the unsigned deviations of the slopes from 1. For the AEI model the fit on training set (a) with $r_{\text{sphere}} = 8.5$ Å is used. The unit of the RMSD is kcal/mol.

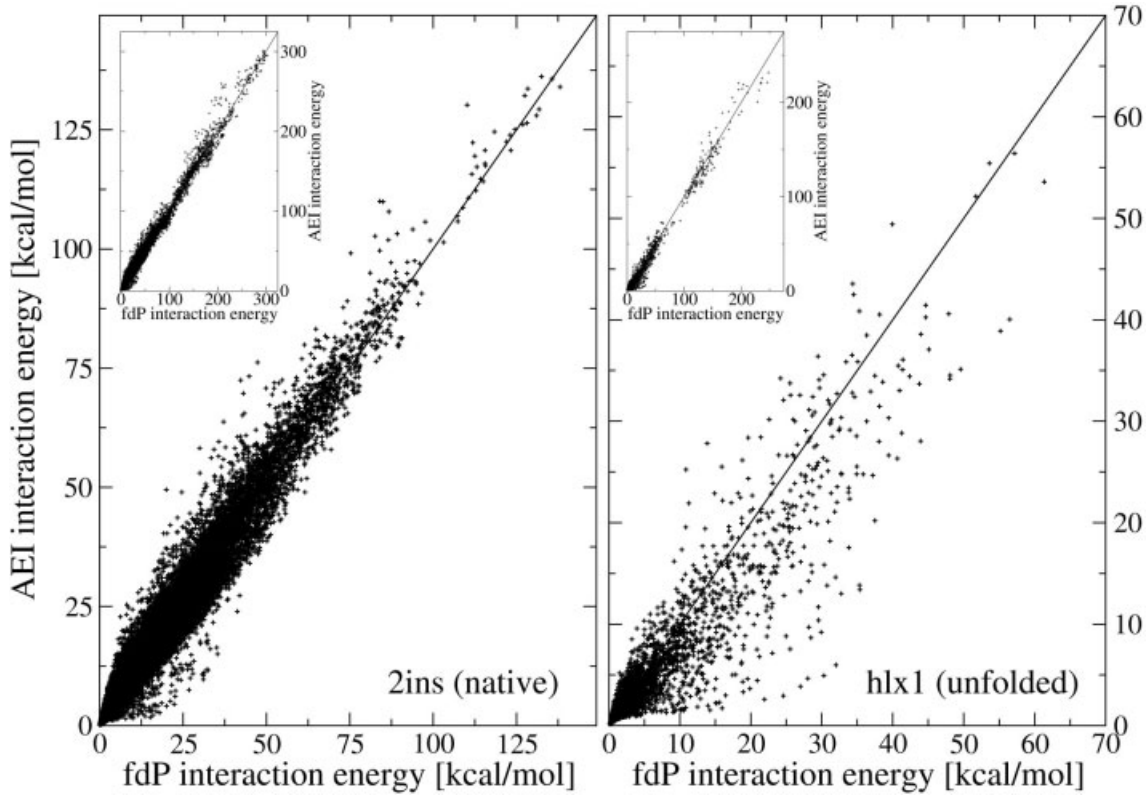

Figure 4. Pair interaction energies E_{ij} for unit charges, calculated by the AEI model [fit on training set (a), $r_{\text{sphere}} = 8.5$ Å], are plotted against fdP values for all but 1-2 and 1-3 pairs for 2ins folded (left) and hlx1 unfolded (right). hlx1 unfolded is the conformation with the poorest correlation in the set of the 52 conformations. The insets show the interaction energies for all pairs. The unit of energy is kcal/mol.

correlation and slope of only 0.91 and 0.66, respectively (excluding 1-2 and 1-3 pairs). The AEI model underestimates interaction energies for very extended structures. However, this is not a serious disadvantage because they are not realistic and are not usually sampled in conventional MD simulations.

The measure of enclosure u_{ij}^{AEI} presented in this work is applicable for any macromolecular system. Yet, for peptides (20 residues or less) results can be improved by a slightly different choice. Using $(u_{ij}^{\text{AEI}})^{1/2}$ instead of u_{ij}^{AEI} improves the accuracy for very extended conformations without any major deterioration of the values of the folded peptides.

To further compare the accuracy of the three models, Figure 5A shows a histogram of the deviations of the interaction energies calculated by the AEI, GB, and SGM models from the fdP values. Again, the GB model is slightly more accurate than the AEI model, whereas the SGM model has by far the largest errors. The deviations are essentially the same if 1-2 and 1-3 pairs are included in the calculations.

Error Cancellation for Atomic Energies

The relevant quantities in MD simulations are sums over pair interaction energies and their derivatives and not single pair values. The model that best reproduces E_{ij} with respect to fdP data is not necessarily the best at reproducing $\Sigma_{j\neq i} E_{ij}$ if the errors in E_{ij} cancel each other poorly. There is a fortuitous cancellation of errors in the GB model because a systematic overestimation (or underestimation) of the effective Born radii has a compensating effect on sums of interactions involving like and opposite charges.[29] In the following, the cancellation of errors in the AEI, GB, and SGM models is compared. For this purpose all atoms of the 52 conformations are assigned partial charges. Let E_{ij}^{fdP} denote the interaction energy calculated by fdP and E_{ij} the energy calculated by the AEI, GB, or SGM model. In a first step the error of each pair interaction energy E_{ij} with respect to the fdP value, that is, $E_{ij} - E_{ij}^{\text{fdP}}$, is determined. Then, for each nonzero partial charge i individually (a total of 37,869 charges), positive errors ($E_{ij} - E_{ij}^{\text{fdP}} > 0$) and negative errors ($E_{ij} - E_{ij}^{\text{fdP}} < 0$) are added up separately. The total error of charge i is the sum of the two and the values are shown in Figure 6. The sums $\Sigma_{j\neq i} E_{ij}$, as calculated by the AEI, GB, and SGM models for each atom i with a nonzero partial charge, are plotted against the appropriate fdP values in Figure 7. The corresponding correlation, slope, and RMSD are shown in Table 4. Note that the vertical deviations from the diagonal line in Figure 7 that are used to calculate the RMSD correspond to the total errors in Figure 6. The correlation and slope, however, cannot be deduced from the data in Figure 6. A histogram of the distri-

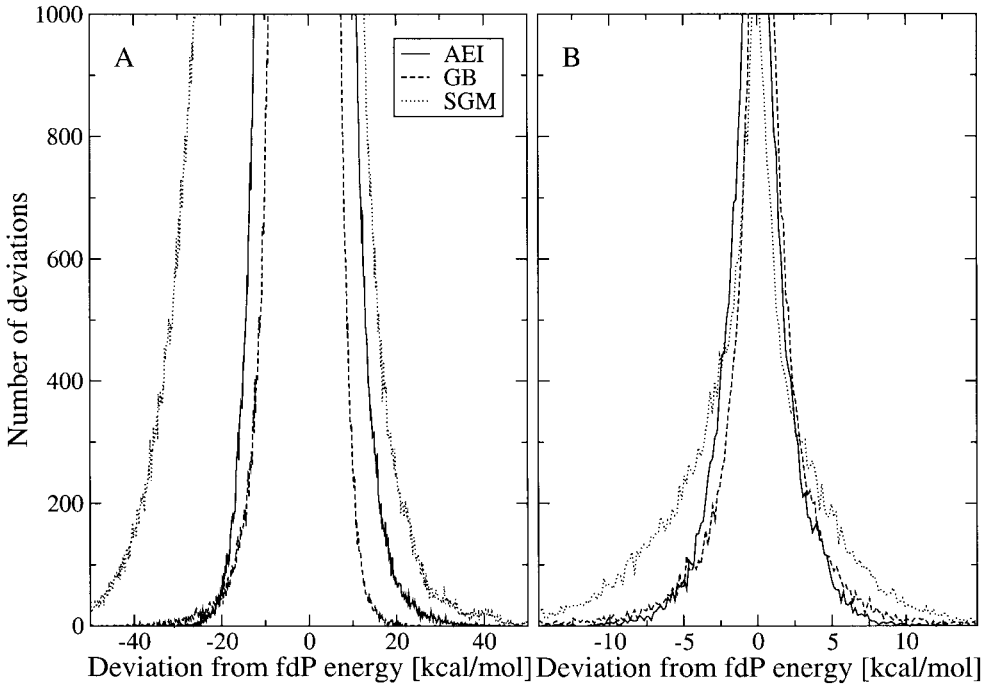

Figure 5. Distribution of deviations from fdP values of pair interaction energies E_{ij} (A) and the sums $\Sigma_{j\neq i} E_{ij}$ (B), calculated by the AEI, GB, and SGM models. A total of 625 and 185 bins of 0.16 kcal/mol each span the energy intervals given on the abscissa axes in (A) and (B), respectively. The number of data points for which the deviation in energy falls within a given bin is shown on the ordinate axis. For (A), all atoms are assigned unit charges, whereas for (B), partial charges are used. In both cases 1-2 and 1-3 pairs are excluded from the calculations.

bution of errors is shown in Figure 5B. Moreover, Table 4 shows the results if the sums $\Sigma_{j\neq i} E_{ij}$ are calculated with a cutoff of 7.5 Å (the default cutoff of CHARMM PARAM19) in the AEI, GB, and SGM models and compared to the corresponding fdP values obtained by adding up interaction energies with no cutoff. Note that 1-2 and 1-3 pairs are excluded from the data presented in Figures 5B, 6, and 7 and Table 4, but the results look similar if all pairs are taken into account.

All three models do in fact benefit greatly from cancellation of errors. Clearly, the SGM model has the largest total errors. It is interesting to note that for small errors (<5 kcal/mol) the AEI and GB models show essentially the same frequencies, whereas larger errors (>5 kcal/mol) occur less often in the AEI than in the GB model (see Fig. 7 and tails in Fig. 5B). From the data presented in Figures 5B and 7 one can deduce that in the case of an infinite cutoff, errors cancel each other better in the AEI than in the GB model. For a cutoff of 7.5 Å the two models show similar accuracy (Table 4).

Total Electrostatic Interaction Energy of Native and Non-Native Conformations

It is important to test the accuracy of the total electrostatic interaction energy calculated by the AEI model for different conformations of the same macromolecule. For this purpose high temperature unfolding simulations at 450 K for 20 ns using an implicit solvation model[27] of a SH3 domain (1shg, 57 residues) and a three-stranded antiparallel β-sheet (beta3s, 20 residues[30]) were performed. Coordinates were saved every 5 ps and all snapshots were sorted according to increasing radius of gyration (Rg). Then 100 conformations were chosen as follows: every 25th conformation of the 500 snapshots with the lowest Rg (20 conformations), every 25th conformation of the 500 snapshots with the largest Rg (20 conformations), and every 50th conformation of the remaining 3000 snapshots (60 conformations). Furthermore, the native state was added. The conformations ranged from folded to significantly extended. They covered a range in the radius of gyration from 10.2 to 25.4 Å for 1shg and from 6.9 to 12.3 Å for beta3s. Note that from the 101 conformations of 1shg or beta3s, only the native state and one of the unfolded states were in training set (a) that was used to parameterize the AEI model. The comparison was limited to two proteins because of the large computational requirements for the fdP calculations on the set of 100 conformations. For each structure the total electrostatic interaction energy $E_{\text{elec}}^{\text{inter}}$ [see eq. (12)] was calculated. The results for the AEI, GB, and SGM models are compared to the fdP data and are shown in Figure 8 and Table 5. Also shown in Table 5 are the results if $E_{\text{elec}}^{\text{inter}}$ is calculated in the AEI, GB, and SGM models with a cutoff of 7.5 Å and compared to the appropriate fdP data obtained with no cutoff. The plots in Figure 8 indicate that the AEI model is accurate enough not only for compact and unfolded structures but also for conformations

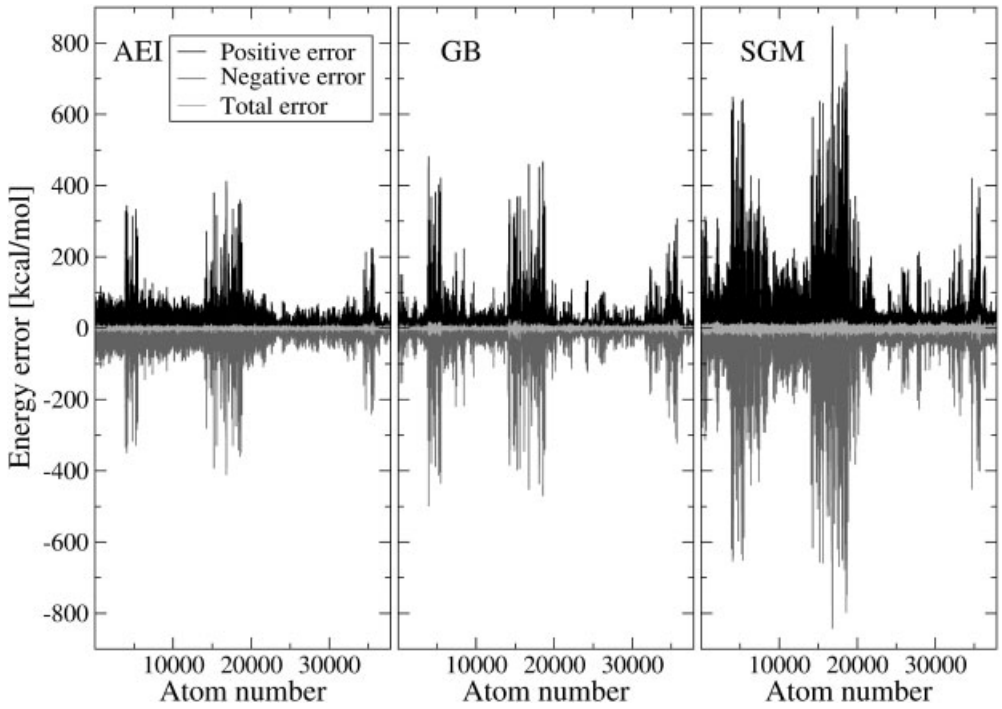

Figure 6. Error cancellation for pair interaction energies as calculated by the AEI [fit on training set (a), $r_{\text{sphere}} = 8.5$ Å], GB, and SGM models. For each atom i with nonzero partial charge, the total error $\Sigma_{j \neq i} (E_{ij} - E_{ij}^{\text{fdP}})$ is given in light gray. The positive and negative contributions to the total error are shown in black and gray, respectively. Partial charges are used and 1-2 and 1-3 pairs are excluded.

with an intermediate value of the Rg. Furthermore, the RMSD values of the total electrostatic interaction energy are smaller in the AEI than in the GB approach. There is a systematic shift towards lower and higher energy values for the AEI and GB model, respectively (see Fig. 8), and the shift is smaller in the AEI than in the GB model. Note that the electrostatic interaction energy alone is not expected to discriminate the native state from other compact conformations.

Efficiency

Finally, we comment on the computational requirements. The AEI model is highly efficient; it is only 10% slower than *vacuo,* irrespective of the size of the molecule. According to ref. 13, the GB approach is slower by a factor of 5 (for large proteins with more than 60 amino acids) to 10 (for small proteins and peptides with up to 60 amino acids) compared to *vacuo.*

Solvation Energies

To further assess the physical relevance of a measure of enclosure based on the sum over weighted volumes of neighbors, this section addresses the evaluation of solvation energies in the framework of the AEI model. A brief description is given here because the focus of this article is on screened interaction energies.

In the Methods section a measure of enclosure for a pair of atoms (i, j) is introduced. In the same spirit one can define a measure of enclosure for a single atom i by

$$w_i^{\text{AEI}} = \sum_{k=1}^{N} v_k \Theta_{ik} \tag{13}$$

where v_k is the van der Waals volume of atom k, N the total number of atoms in the system, and Θ_{ik} is defined in the Methods section with $r_{\text{sphere}} = 8.5$ Å. Note that Θ_{ik} is different from zero only for the atoms in a sphere of radius 8.5 Å centered on atom i. Analytical functions of w_i^{AEI}, that is, $\Delta E_i^{\text{AEI}} = h_p(w_i^{\text{AEI}})$, where ΔE_i^{AEI} denotes the solvation energy of atom i calculated in the AEI model, are fitted to fdP-derived solvation energies for unit charges of the atoms of one protein (1a2p, 1,073 atoms). The index p accounts for the fact that different functions are necessary for different ranges of the van der Waals radii (see Appendix). The AEI solvation energies for all atoms with nonzero partial charge of 10 proteins not used to parameterize the model (1bpi, 1crn, 1hdn, 1pgb, 1pht, 1ycq, 1ycr, 2ci2, 2ptl, beta3s), are shown in Figure 9. For the AEI model, the correlation and slope are 0.987 and 0.952, respectively, and for the GB model 0.986 and 0.632, respectively. These results indicate that a measure of enclosure for a single atom i based on summing over neighbors allows the modeling of not only interaction but also solvation energies.

Conclusion

Both the AEI and GB models utilize a measure of enclosure for pairs of charges to calculate the screened electrostatic interac-

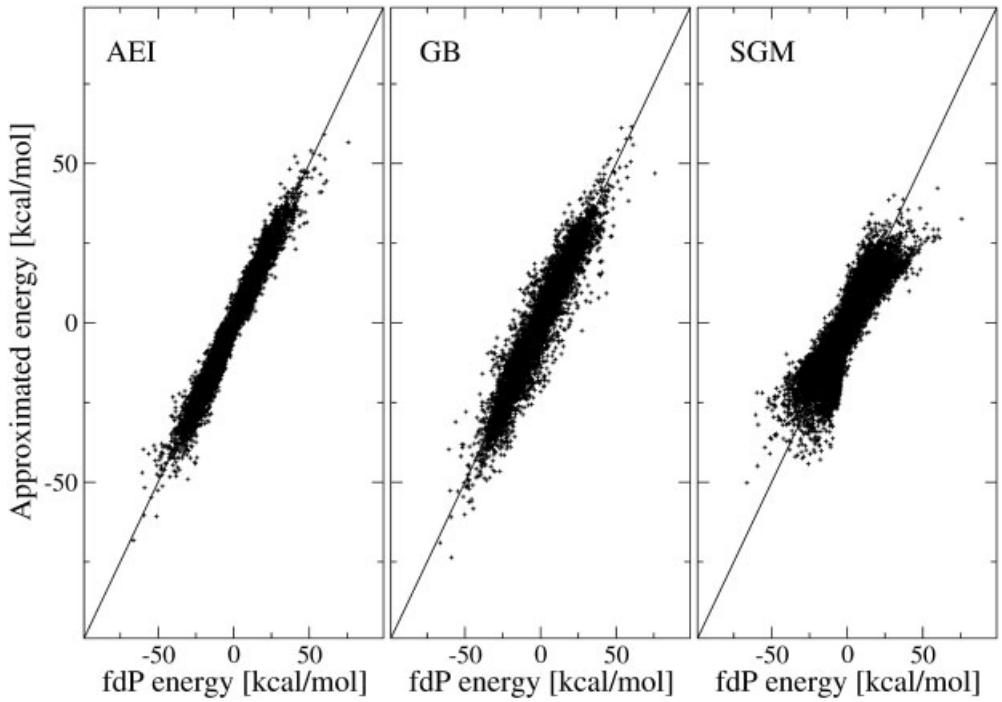

Figure 7. Each data point represents the sum of all pair interaction energies of an atom i with nonzero partial charge, $\Sigma_{j \neq i} E_{ij}$. The AEI [fit on training set (a), $r_{\text{sphere}} = 8.5$ Å], GB, and SGM values are plotted against the fdP data. Partial charges are used and 1-2 and 1-3 pairs are excluded. See text for details.

tion energy. For each charge in the system an effective volume and an effective Born radius is evaluated in the AEI and the GB approach, respectively. These quantities are combined in a heuristic way to obtain a measure of enclosure for a pair of charges. The essential element of the AEI model is to define such a measure with information easily available from a reasonably large neighborhood of a given pair. The degree of enclosure of two charges quantifies the distribution of solute atomic volumes surrounding the pair. The appealing feature of the AEI model is the efficiency with which the measure of enclosure can be calculated. It is simply the square of the sum of weighted atomic volumes. Hence, the present implementation of the AEI model uses only quantities whose calculation is intrinsic to MD simulations so that the computational overhead is negligible with respect to *vacuo*. In the GB approach the measure of enclosure of a pair of charges is the square root of the product of their effective Born radii, whose calculation requires integration of the electrostatic energy density over the solute volume. The sum of all pair interaction energies of an atom i (relevant for MD simulations) and the total electrostatic interaction energy of a macromolecule are reproduced more accurately in the AEI than in the GB approach. Only single pair interaction energies are slightly better approximated in the GB model. The validity of the AEI model is further assessed by

Table 4. Comparison of the Sum of All Interaction Energies of Atom i, $\Sigma_{j \neq i} E_{ij}$, Calculated by the AEI, GB, and SGM Models and fdP.

	No cutoff			Cutoff of 7.5 Å		
Model	Correlation	Slope	RMSD	Correlation	Slope	RMSD
AEI	0.981	1.016	2.152	0.957	0.994	3.258
GB	0.966	1.003	2.895	0.955	0.989	3.312
SGM	0.890	0.794	4.958	0.878	0.782	5.197

For the AEI model the fit on training set (a) with $r_{\text{sphere}} = 8.5$ Å is used. Partial charges are assigned to all atoms and 1-2 and 1-3 pairs are excluded. No cutoff is applied for the calculation of the data on the left hand part of the table. For the data on the right hand part, the sums $\Sigma_{j \neq i} E_{ij}$ are calculated with a cutoff of 7.5 Å in the AEI, GB, and SGM models and compared to the corresponding fdP values obtained with no cutoff. The unit of the RMSD is kcal/mol.

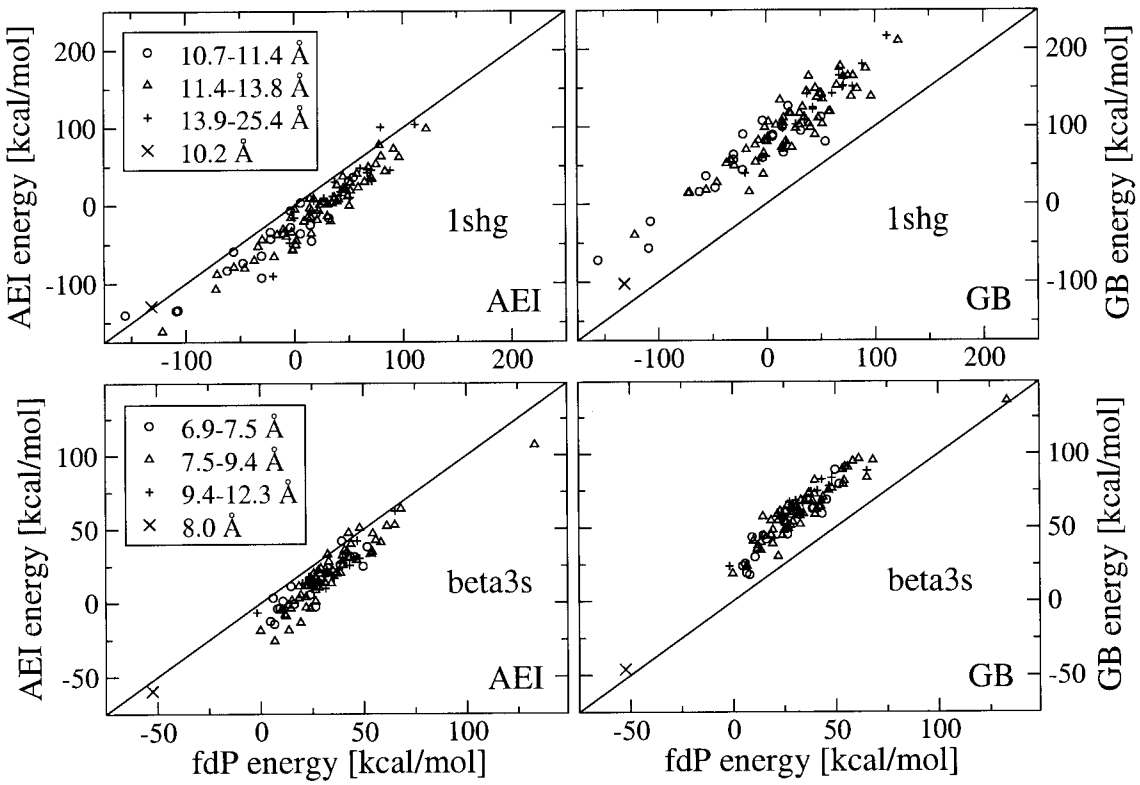

Figure 8. Each data point represents the total electrostatic interaction energy $E_{\text{elec}}^{\text{inter}}$ for a given conformation as calculated by the AEI [fit on training set (a), r_{sphere} = 8.5 Å] and GB models against the corresponding fdP value. Data for 101 conformations along a high temperature unfolding trajectory of 1shg (top) and beta3s (bottom) are shown. Partial charges are used and 1-2 and 1-3 pairs are excluded from these calculations. Different symbols discriminate between different ranges for the radius of gyration. Circles and pluses represent the 20 conformations with small and large Rg, respectively, and triangles the 60 intermediate ones. The total electrostatic interaction energy of the native state is shown by the symbol x.

Table 5. For 101 Conformations along a High Temperature Unfolding Trajectory of 1shg and beta3s, the Total Electrostatic Interaction Energy $E_{\text{elec}}^{\text{inter}}$ Is Calculated by the AEI, GB, and SGM Models.

	No cutoff			Cutoff of 7.5 Å		
Model	Correlation	Slope	RMSD	Correlation	Slope	RMSD
1shg						
AEI	0.955	0.983	31.265	0.917	0.969	27.631
GB	0.941	1.040	81.515	0.935	1.038	83.598
SGM	0.706	0.674	139.198	0.661	0.672	124.610
beta3s						
AEI	0.939	1.006	15.159	0.916	0.958	15.567
GB	0.937	1.043	28.923	0.925	1.006	27.835
SGM	0.818	0.849	63.504	0.769	0.854	63.163

For the AEI model, the fit on training set (a), r_{sphere} = 8.5 Å is used. Correlation, slope, and RMSD with respect to fdP data are shown. Partial charges are used and 1-2 and 1-3 pairs are excluded from these calculations. No cutoff is applied for the calculation of the data on the left hand part of the table. The data on the right hand part show the total electrostatic interaction energy calculated with a cutoff of 7.5 Å in the AEI, GB, and SGM models, compared to the corresponding fdP values obtained with no cutoff. The unit of the RMSD is kcal/mol.

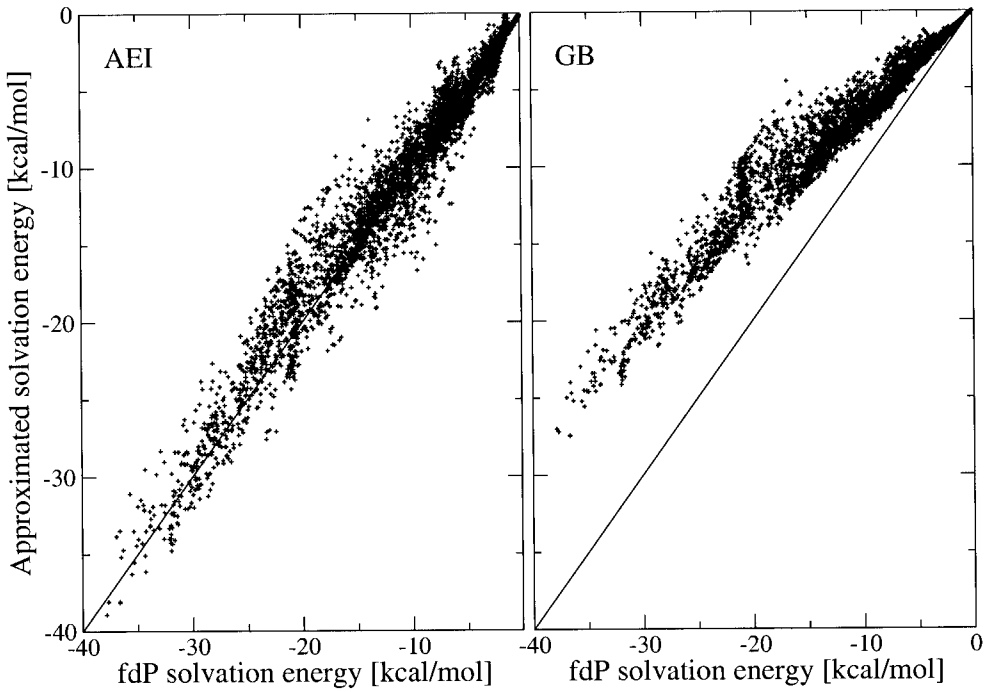

Figure 9. Atomic solvation energies calculated by the AEI (left) and GB (right) models for 10 proteins (6504 data points) are compared with the fdP values. Partial charges are assigned to all atoms.

demonstrating that solvation energies can be calculated with a measure of enclosure for single atoms that is similar to the one used for pairs.

Acknowledgments

We thank two anonymous referees for very useful comments.

Appendix

Interaction Energy

Let r_k be the van der Waals radius of atom k and $v_k = \frac{4}{3}\pi r_k^3$ its van der Waals volume. The van der Waals radii are taken from the CHARMM parameter set PARAM19 and depend only on the atom type. The measure of enclosure u_{ij}^{AEI} is defined by eqs. (5) and (7). Let r_{ij} be the distance between atoms i and j and define $p_{ij} = u_{ij}^{\mathrm{AEI}}/r_{ij}$. The reciprocal of the effective dielectric function in the AEI model is denoted by g_k, where $k = 1$ for 1-2 pairs, $k = 2$ for 1-3 pairs, and $k = 3$ for all but 1-2 and 1-3 pairs. Due to the close relationship between the AEI and GB approaches, the g_k are chosen to be of the same functional form as the reciprocal of the effective dielectric function in the GB model [see f in eq. (4)]:

$$g_k(p_{ij}) = a_{1,k} - (a_{1,k} - a_{2,k}) \times (1 + (a_{4,k}(p_{ij} + a_{3,k}))^2 e^{-[1/(a_{5,k}(p_{ij}+a_{3,k}))^2]})^{-(1/2)} \quad \text{(A1)}$$

All five parameters $a_{i,k}$ appearing in g_k have a well defined meaning. The functions g_k are of sigmoidal shape with the maximum and minimum value $a_{1,k}$ and $a_{2,k}$, respectively, if $a_{1,k} > a_{2,k}$. This condition is always satisfied in the AEI model. The parameter $a_{3,k}$ translates the function g_k parallel to the abscissa axis, and $a_{4,k}$ and $a_{5,k}$ are scaling factors. The parameters $a_{i,k}$ are determined by fitting g_k to the inverse of fdP-derived effective dielectric constants extracted from all 52 conformations [training set (a)]. The sphere radius used is $r_{\mathrm{sphere}} = 8.5$ Å. The parameters are determined to be

$$(a_{i,k}) = \begin{pmatrix} +0.113 \cdot 10^{+01} & +0.133 \cdot 10^{+01} & +0.100 \cdot 10^{+01} \\ +0.253 \cdot 10^{+00} & -0.980 \cdot 10^{-01} & +0.127 \cdot 10^{-01} \\ +0.145 \cdot 10^{+07} & +0.124 \cdot 10^{+07} & +0.000 \cdot 10^{+00} \\ +0.451 \cdot 10^{-06} & +0.541 \cdot 10^{-06} & +0.135 \cdot 10^{-05} \\ +0.998 \cdot 10^{+00} & +0.967 \cdot 10^{+00} & +0.581 \cdot 10^{-05} \end{pmatrix} \quad \text{(A2)}$$

The functions g_1 and g_2 have five parameters each whereas the function g_3, which is the most relevant for molecular mechanics and dynamics, has in effect only two parameters because $a_{1,3}$, $a_{2,3}$, and $a_{3,3}$ are set to the standard GB values, that is, $a_{1,3} = 1/\varepsilon_m = 1$, $a_{2,3} = 1/\varepsilon_s = 1/78.5$, and $a_{3,3} = 0$.

Solvation Energy

The measure of enclosure w_i^{AEI} for a single atom i is defined by eq. (13). Solvation energies are calculated by $h_p(w_i^{\mathrm{AEI}}) = b_{1,p} + b_{2,p} w_i^{\mathrm{AEI}} + b_{3,p}(w_i^{\mathrm{AEI}})^2$, where $p = 1$ for van der Waals radii in

the range from 0.5–1.0 Å, $p = 2$ for the range from 1.5–2.0 Å (there are no van der Waals radii in PARAM19 with a value between 1.0 and 1.5 Å), and $p = 3$ for van der Waals radii larger than 2.0 Å. The sphere radius used is $r_{\text{sphere}} = 8.5$ Å. The parameters $b_{i,p}$ are determined by fitting the functions h_p to fdP-derived atomic solvation energies for unit charges of one single protein (1a2p, 1,073 atoms). The parameters are determined to be

$$(b_{i,p}) = \begin{pmatrix} -0.220 \cdot 10^{+03} & -0.108 \cdot 10^{+03} & -0.766 \cdot 10^{+02} \\ +0.209 \cdot 10^{+00} & +0.502 \cdot 10^{-01} & +0.169 \cdot 10^{-01} \\ -0.326 \cdot 10^{-04} & +0.193 \cdot 10^{-04} & +0.250 \cdot 10^{-04} \end{pmatrix} \tag{A3}$$

References

1. Gilson, M. K. Curr Opin Struct Biol 1995, 5, 216.
2. Roux, B.; Simonson, T. Biophys Chem 1999, 78, 1.
3. Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027.
4. Cramer, C. J.; Trulhar, D. G. Chem Rev 1999, 99, 2161.
5. Orozco, M.; Luque, F. J. Chem Rev 2000, 100, 4187.
6. Warwicker, J.; Watson, H. C. J Mol Biol 1982, 157, 671.
7. Gilson, M. K.; Honig, B. H. Proteins Struct Funct Genet 1988, 4, 7.
8. Bashford, D.; Karplus, M. Biochemistry 1990, 29, 10219.
9. Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Comput Phys Commun 1991, 62, 187.
10. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
11. Scarsi, M.; Apostolakis, J.; Caflisch, A. J Phys Chem A 1997, 101, 8098.
12. Bashford, D.; Case, D. A. Annu Rev Phys Chem 2000, 51, 129.
13. Lee, M. S.; Salsbury Jr, F. R.; Brooks III, C. L. J Chem Phys 2002, 116, 10606.
14. Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. Chem Phys Lett 1995, 246, 122.
15. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J Phys Chem A 1997, 101, 3005.
16. Dominy, B. N.; Brooks III, C. L. J Phys Chem B 1999, 103, 3765.
17. Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227.
18. Gelin, B. R.; Karplus, M. Biochemistry 1979, 18, 1256.
19. Mehler, E. L.; Eichele, G. Biochemistry 1984, 23, 3887.
20. Mehler, E. L. Protein Eng 1990, 3, 415.
21. Hassan, S. A.; Guarnieri, F.; Mehler, E. L. J Phys Chem B 2000, 104, 6478.
22. Wang, L.; Hingerty, B. E.; Srinivasan, A. R.; Olson, W. K.; Broyde, S. Biophys J 2000, 83, 382.
23. Mallik, B.; Masunov, A.; Lazaridis, T. J Comput Chem 2002, 23, 1090.
24. Hirschfelder, J. O.; Curtiss, C. F.; Bird, R. B. Molecular theory of gases and liquids; John Wiley & Sons: New York, 1964.
25. Jackson, J. D. Classical Electrodynamics; John Wiley & Sons: New York, 1975.
26. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
27. Ferrara, P.; Apostolakis, J.; Caflisch, A. Proteins Struct Funct Genet 2002, 46, 24.
28. Im, W.; Beglov, D.; Roux, B. Comput Phys Commun 1998, 111, 59.
29. Onufriev, A.; Case, D. A.; Bashford, D. J Comput Chem 2002, 23, 1297.
30. De Alba, E.; Santoro, J.; Rico, M.; Jimenez, M. A. Protein Sci 1999, 8, 854.