

# Distance and Exposure Dependent Effective Dielectric Function

BUDDHADEB MALLIK, ARTEM MASUNOV, THEMIS LAZARIDIS  
*Department of Chemistry, City College of CUNY, Convent Avenue & 138<sup>th</sup> Street,  
New York, New York 10031*

*Received 10 May 2001; Accepted 7 March 2002*

**Abstract:** In an effort to develop a dielectric screening function for molecular dynamics simulations of biomolecules in implicit solvent, effective dielectric constants ( $D_{\text{eff}}$ ) for a large number of atom pairs in a typical globular protein are calculated by continuum electrostatics. Plots of  $D_{\text{eff}}$  versus the interchange distance are in general sigmoidal with the characteristics of the curve depending on the distance of the two charges from the dielectric boundary and, secondarily, on the extent to which the area surrounding each charge is occupied by solvent (the “exposure”). The  $D_{\text{eff}}$  values were fitted to an empirical, analytical function of these parameters that reproduces the data reasonably well, although considerable scatter exists in the range of  $D_{\text{eff}}$  from 30 to 80. In the system used for parameterization, the mean square deviation of electrostatic interaction energies with this function is 0.48 kcal/mol, compared to 1.45 for an analytical Generalized Born model and 1.52 for the linear distance-dependent dielectric model. When tested in other proteins of varying size and compactness, the present function is superior to both of the above models, except for a fully unfolded polypeptide chain, where the Generalized Born model is superior.

© 2002 Wiley Periodicals, Inc. J Comput Chem 23: 1090–1099, 2002

**Key words:** continuum electrostatics; Poisson’s equation; effective dielectric constant; surface distance; exposure; implicit solvation

## Introduction

Electrostatic interactions are largely responsible for the structure and the diverse biological activities of proteins.<sup>1</sup> A large number of interdependent factors, such as position of fixed charges, rotational flexibility of internal dipoles and their magnitude, presence of ions, shielding from solvent molecules as well as within the protein, and so forth, govern the magnitude of these interactions. One fruitful approach has been to treat the protein interior as a low dielectric medium embedded in a high dielectric medium and use the equations of continuum electrostatics to obtain electrostatic energies.<sup>2</sup> Given the interior and exterior dielectric constants and a charge distribution, the Poisson equation can be solved to give the electrostatic potential at all points in space. The presence of salt in the solvent can be treated by the Poisson-Boltzmann equation. For simple boundary shapes, such as planes or spheres, the Poisson-Boltzmann equation can be solved analytically,<sup>3</sup> but for irregular solute shapes, it must be solved numerically.<sup>4</sup> This approach appears to be the most reliable method for estimating dielectric screening in biomolecules. However, it is computationally intensive. For example, solving the Poisson-Boltzmann equation at every step of a molecular dynamics simulation would require several hundred times more computer time than a vacuum simu-

lation.<sup>5–7</sup> Therefore, empirical dielectric screening functions have been pursued.

Warshel and Levitt<sup>8</sup> used a microscopic dielectric model and found that the effective dielectric can be described by the equation  $D_{\text{eff}} = 1 + r$ , where  $r$  is the distance between two charges. In another study of the energetics of bacteriorhodopsin Warshel<sup>9</sup> proposed a quadratic function [ $D_{\text{eff}} = 2 + (r - 1)^2$ ]. To calculate the electrostatic potential in the active site of papain, van Duijnen et al.<sup>10</sup> proposed a simple function for  $D_{\text{eff}}$  that varied between 1 and 2. In their study of the effect of dipeptide, protein, and solvent environment on side-chain torsional potentials, Gelin and Karplus<sup>11</sup> used a dielectric constant that is numerically equal to the distance in Å ( $D_{\text{eff}} = r$ ); this is the linear distance-dependent dielectric model that has been popular due to its computational efficiency (no need to take the square root of  $r^2$ ). In addition to using  $D_{\text{eff}} = r$ , Northrup et al.<sup>12</sup> approximated the dielectric screening due to the surrounding water by reducing the partial charges of atoms to an extent that depended on their distance from the protein center. Warshel et al.<sup>13</sup> used experimen-

**Correspondence to:** T. Lazaridis; e-mail: themis@sci.ccny.cuny.edu  
Contract/grant sponsor: National Science Foundation; contract/grant number: DBI-9974621

tal  $pK_a$  values and redox potentials to derive a distance-dependent dielectric function for surface charges and charges separated by more than 5 Å. However, they did not recommend this function for calculating the electrostatic interactions of entirely buried charges. Mehler and Eichele<sup>14</sup> introduced a sigmoidal distance-dependent dielectric permittivity function to calculate the electrostatic effects in water accessible regions of proteins. Their parameterization of the function was based on experimental and theoretical data of dielectric permittivity obtained from dissociation constants of bifunctional acids and bases. For metal ion-nucleic acid interactions, Hingerty et al.<sup>15</sup> proposed a modification of Debye's distance-dependent dielectric function, which is sigmoidal and rises rapidly with distance. An alternative mathematical expression for this function was proposed by Ramstein and Lavery.<sup>16</sup> Pickersgill<sup>17</sup> noted that the simple function  $D_{\text{eff}} = 4.5r$  reproduces Poisson's equation results in the active site of papain as well as more complex functions. Guarnieri et al.<sup>18</sup> proposed an extension of Mehler and Eichele's<sup>14</sup> work, where the position-dependence of the screening is partly accounted for by using different values of a parameter for buried and for exposed atoms.

All of the above functions have limitations. The simple  $D_{\text{eff}} = r$  model is clearly inadequate because it does not discriminate between buried and solvent-exposed sites. The functions of Warshel et al.,<sup>13</sup> Mehler and Eichele,<sup>14</sup> and Hingerty et al.<sup>15</sup> were designed for solvent-exposed regions and are not appropriate for buried atoms. In an attempt towards a more generally applicable function, Sandberg and Edholm<sup>19</sup> recently proposed a modification of the function of Warshel et al.,<sup>13</sup> with parameters depending on the mean distance of the atoms from the protein surface. In their model they divide the protein into two regions. The two outermost atomic layers ( $\leq 5.4$  Å from the protein surface) comprise the boundary region and the remainder of the protein comprises the core region. The parameters of their dielectric function depend on which region the two atoms belong to. The model was tested for its ability to predict  $pK_a$  values, but evaluation of the model is complicated by the fact that the self energies of the charges were neglected.<sup>20</sup>

The goal of the present work is to design a function describing the dielectric screening of the interaction between any pair of charges for use in implicit-solvent molecular dynamics simulations of biomolecules. The screening of electrostatic interactions has contributions from electronic polarization, solvent orientational polarization, and relaxation of protein moieties.<sup>21</sup> Because our function is to be used in dynamics simulations, in which the last contribution is included explicitly, we only need to consider the first two contributions. Electronic polarization is usually taken into account by setting the interior dielectric constant equal to 2. The solvent contribution can be calculated by treating the solvent as a medium of dielectric constant 80. Our approach is then the following: we consider a typical protein and use continuum electrostatics to obtain effective dielectric constants for a large number of atom pair interactions. The resulting  $D_{\text{eff}}$  values are fitted to an analytical function, which depends on the distance between the two points, the distance of each point from the protein surface, and the "exposure" of each point to the solvent (see below). Our ultimate goal is to incorporate this function into an implicit solvation model for biomolecules by combining it with the self energies of the EEF1 model.<sup>22</sup>

## Methods

Continuum electrostatics calculations were performed using DELPHI,<sup>23</sup> a package that solves numerically the Poisson-Boltzmann equation for solutes of arbitrary shape and charge distribution. In this work we assume zero ionic strength, that is, we solve the Poisson equation

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla \cdot \phi(\mathbf{r})] + \frac{4\pi\rho(\mathbf{r})}{kT} = 0 \quad (1)$$

where  $\phi(r)$  is the dimensionless electrostatic potential in units of  $kT/e$ ,  $k$  is Boltzmann's constant,  $T$  the absolute temperature,  $\epsilon(r)$  the dielectric constant, and  $\rho(r)$  the charge density. All calculations were performed at a resolution of 1.4 grids per Å.

The molecular surface was defined with a water probe radius of 1.8 Å. A dielectric constant of 2 was used for the protein interior (some calculations with  $\epsilon_i = 1$  were also performed for comparison), and the value 80 was used for the solvent. We placed a unit charge at 25 different points in the protein, and for each we calculated the potential at a large number of other points. The interaction between two charges is equal to the potential generated by the first charge times the second charge:

$$G_{ij} = 0.5925\phi_j Q_j = 332 \frac{Q_i Q_j}{D_{\text{eff}} r_{ij}} \text{ kcal/mol} \quad (2)$$

where 0.5925 kcal/mol is  $kT$  at 298 K. The above equation defines the effective dielectric constant.<sup>24,25</sup> Therefore

$$D_{\text{eff}} = \frac{561.2Q_i}{\phi_j r_{ij}} \quad (3)$$

where  $Q_i$  is the reference charge at a selected site  $i$ ,  $\phi_j$  is the potential (in  $kT/e$ ) at the uncharged site  $j$  due to  $Q_i$ , and  $r_{ij}$  is the distance in Å between sites  $i$  and  $j$ .

We then examined the dependence of the resulting  $D_{\text{eff}}$  values on two parameters. The first is the distance of the two points from the protein surface ( $d_i, d_j$ ). The surface generated by DELPHI is represented by the Cartesian coordinates of a large number of more or less evenly spaced points. The distances from any atom to all these surface points were calculated, and the minimum of these distances was taken as the atom-to-surface distance. The second parameter is a measure of the exposure of each point to the solvent. For that we used the ratio of the solvation free energy of an atom, as calculated by EEF1,<sup>22</sup> to its reference solvation free energy:

$$e_i = \frac{\Delta G_i^{\text{solv}}}{\Delta G_i^{\text{ref}}} \quad (4)$$

According to EEF1, the solvation free energy of any atom in a polyatomic solute is

$$\Delta G_i^{\text{solv}} = \Delta G_i^{\text{ref}} - \sum_{i \neq j} f_i(\mathbf{r}_{ij}) V_j \quad (5)$$

where  $f_i(r)$  is the solvation free energy density of atom  $i$  at point  $r$ ,  $V_j$  is the volume of a neighboring atom that displaces solvent molecules around  $i$ , and  $r_{ij}$  is the distance between  $i$  and  $j$ .  $\Delta G_i^{\text{ref}}$  is the reference free energy of atom  $i$  in a small molecule where the atom  $i$  is fully exposed. The solvation free energy density is assumed to be a Gaussian function of the dimensionless distance from the atom:

$$f_i(r)4\pi r^2 = \alpha_i \exp(-x_i^2); \quad x_i = \frac{r - R_i}{\lambda_i} \quad (6)$$

where  $R_i$  is the van der Waals radius of  $i$ ,  $\lambda_i$  is a correlation length, and  $\alpha_i$  is a proportionality coefficient

$$\alpha_i = 2\Delta G_i^{\text{free}}/\sqrt{\pi}\lambda_i \quad (7)$$

where  $\Delta G_i^{\text{free}}$  is the solvation free energy of the “isolated” atom  $i$ .

From eqs. (5), (6), and (7) we get

$$e_i = 1 - \sum_j \frac{\Delta G_i^{\text{free}}}{\Delta G_i^{\text{ref}}} \cdot \frac{\exp(-x_i^2)}{2\sqrt{\pi}\pi\lambda_i r_{ij}^2} \quad (8)$$

Because the ratio  $\Delta G_i^{\text{free}}/\Delta G_i^{\text{ref}}$  is approximately the same for different atoms, the value of the exposure is essentially independent of the chemical type of the atom. Because  $f$  is a short-ranged function,  $e_i$  gives a measure of the short-range exposure of the atom, that is, how much of the immediate neighborhood of the atom is occupied by solvent. Thus, this parameter tends to be larger at convex points and smaller at concave points of the protein surface. The exposure is correlated with the distance from the surface, but the correlation is not perfect. One disadvantage of  $e_i$  is that it is influenced by the presence of small voids. Thus, even deeply buried atoms do not have  $e_i$  exactly zero as they ought to.

## Results and Discussion

To calculate the effective dielectric constants we chose the crystal structure of subtilisin (pdb code 1sbt except Ala99 is replaced by Asp). This protein has 275 amino acid residues and a roughly spherical shape with an average diameter of about 45 Å. There are concave parts on its surface so that the maximum distance of any atom from the surface is about 11.1 Å. We grouped the atoms according to their distance from the protein surface. The first set (R1) consists of all surface atoms ( $d_i = 1.6$ – $1.9$  Å), the second set (R2) consists of atoms with surface distances 1.9–3.6 Å, the third set (R3) with  $d_i = 3.6$ – $5.4$  Å, and the rest of the atoms (R4) with  $d_i > 5.4$  Å. The 25 reference points are distributed over these four regions (five in R1, eight in R2, five in R3, and seven in R4). Figure 1 shows the calculated effective dielectric constant as a function of distance from the reference charge for four of the 25 reference points. The plots have a sigmoidal character, in agreement with previous studies of solvent-exposed sites,<sup>14</sup> but the exact shape of the curve and the position of the inflection point depend on the position of the two charges. Over a certain range of

distances the effective dielectric constant can be higher than 80, as has been noted before.<sup>24,26</sup>

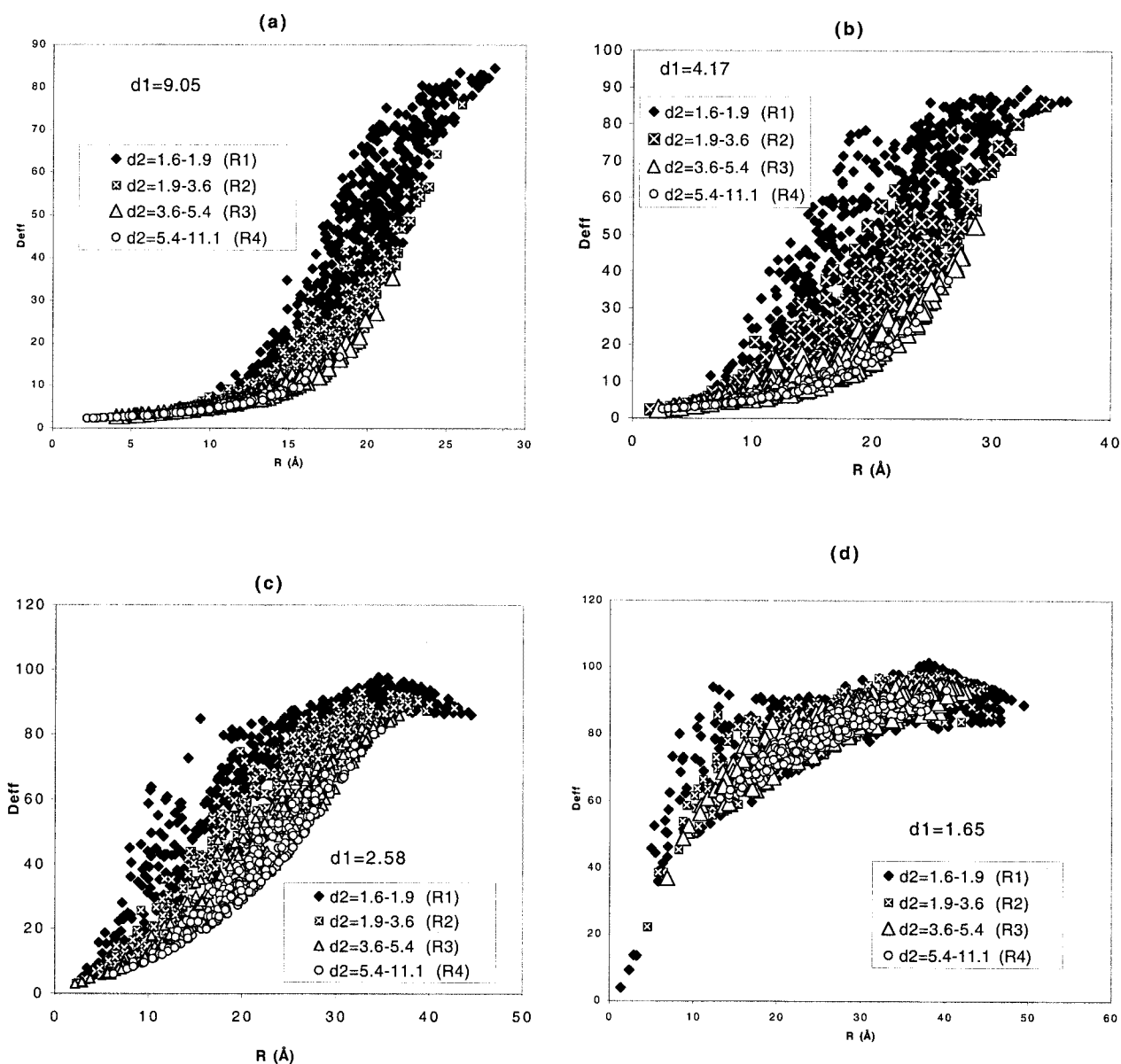
The data in Figure 1a are for a reference charge that is deeply buried ( $d_1 = 9.05$  Å). The  $D_{\text{eff}}$  of core atoms (regions R4 and R3) rises very slowly with distance and is relatively small up to a long distance. At longer distances,  $D_{\text{eff}}$  correlates clearly with the position of the second atom; it is higher the closer the second atom is to the protein surface, although there is considerable scatter. A similar plot is observed in Figure 1b for a reference charge in region R3 ( $d_1 = 4.17$  Å), except there is more scatter in the values of  $D_{\text{eff}}$  for surface (R1) atoms. The variation of  $D_{\text{eff}}$  with separation distance is shown in Figure 1c for a reference charge in R2 ( $d_1 = 2.58$  Å). The rise of  $D_{\text{eff}}$  for core charges (R4 and R3) is now faster and more gradual than in Figures 1a and 1b (the sigmoidal character is diminished). In region R2, almost a linear dependence of  $D_{\text{eff}}$  on distance is found. For a charge located in the outer boundary region (R1) with  $d_1 = 1.65$  Å, the variation of  $D_{\text{eff}}$  with R is shown in Figure 1d. This plot is reminiscent of two charges in aqueous medium.  $D_{\text{eff}}$  rises very quickly to 40 even at 5–6 Å distance and then approaches the asymptotic value at a slower pace. The sigmoidal nature of  $D_{\text{eff}}$  is barely visible. The  $D_{\text{eff}}$  values for atoms in all four regions (R1, R2, R3, R4) merge together at a distance 15 Å and beyond.

In these figures one can see that there is considerable scatter in the  $D_{\text{eff}}$  values even for points with similar distances to the surface for surface atoms. Some of this scatter can be explained by different exposure values. The dependence of  $D_{\text{eff}}$  on exposure for the reference charge of Figure 1b ( $d_1 = 4.17$  Å) interacting with atoms in region R1 is shown in Figure 2. Although significant scatter still exists, a correlation between low exposures and low  $D_{\text{eff}}$  is visible. As expected, for buried atoms, small variations in  $e_i$  (due to the presence of small voids) do not have any effect on  $D_{\text{eff}}$  (results not shown).

From these results we conclude that the primary determinant of  $D_{\text{eff}}$  (other than  $R_{ij}$ ) is the distance of the two points from the surface,  $d_i$  and  $d_j$ . For points close to the surface, a secondary determinant is the “exposure” of the two points. Therefore, we seek to determine an analytical function of the form

$$D_{\text{eff}} = D_{\text{eff}}(R_{ij}; d_i, d_j, e_i, e_j)$$

which is symmetric with respect to interchange of  $i$  and  $j$ . To separate the problem of determining an approximate “mixing rule” (obtaining an “average”  $d$  from  $d_1, d_2$  and  $e$  from  $e_1, e_2$ ) from determining the dependence of  $D_{\text{eff}}$  on depth and exposure, we first focus on data for which  $d_1 \approx d_2 = d$  and  $e_1 \approx e_2 = e$ ; that is, we first consider only data where surface distances are within  $\pm 0.5$  Å and exposures are within  $\pm 0.05$  from the surface distance and exposure of the reference charge, respectively. The variation of  $D_{\text{eff}}$  with R for these data is shown in Figure 3a, grouped into four regions as before. At all separation distances,  $D_{\text{eff}}$  is lowest for the entirely buried atoms and it increases with decreasing surface distance. The largest scatter of data points is observed for region R2. This is partly due to the sensitivity of  $D_{\text{eff}}$  to  $d$  in this range of  $d$ . Figure 3b shows the dependence of  $D_{\text{eff}}$  on exposure for R1 atoms, where it can be seen that high  $D_{\text{eff}}$  tends to correspond to high exposures.



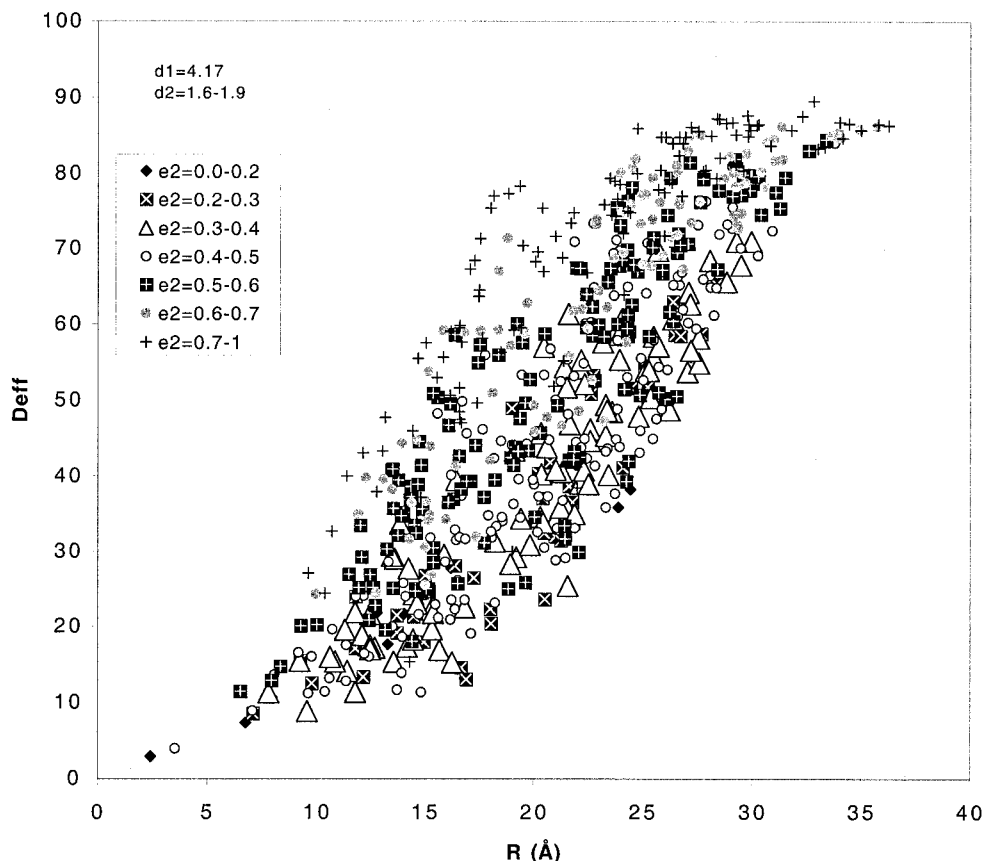
**Figure 1.** Plot of  $D_{\text{eff}}$  versus  $R$  for four reference charges at different distances from the protein surface: (a)  $d_1 = 9.05$  Å, (b)  $d_1 = 4.17$  Å, (c)  $d_1 = 2.58$  Å, (d)  $d_1 = 1.65$  Å. The data points are grouped according to the distance of the second atom from the surface (see legend).

These results suggest a functional form for  $D_{\text{eff}}(R_{ij}; d_i, d_j, e_i, e_j)$  that is sigmoidal and allows us to adjust independently the position of the inflection point and the slope at that point. Most sigmoidal functions proposed so far<sup>14–16</sup> cannot do that because they have only one adjustable parameter. An alternative sigmoidal function, obtained in studies of cooperative binding, is the following:

$$f = \frac{r^n}{1 + r^n} \quad (9)$$

This function is sigmoidal for  $n > 1$ . It goes from 0 to an asymptotic value of 1 with an inflection point at  $r = 1$ . The exponent  $n$  determines the steepness (“cooperativity”) of the curve. This suggests the following function for  $D_{\text{eff}}$ .

$$D_{\text{eff}} = 2 + 76 \frac{\left(\frac{r}{A}\right)^n}{1 + \left(\frac{r}{A}\right)^n} \quad (10)$$



**Figure 2.** Plot of  $D_{\text{eff}}$  versus  $R$  for a reference charge of surface distance  $d_1 = 4.17 \text{ \AA}$  interacting with atoms in region R1 (surface distances between 1.6 and 1.9  $\text{\AA}$ ). The points are grouped according to the exposure of the second atom (see legend).

The above function goes from 2 to 78 and has an inflection point at  $r = A$ . (Such a function cannot capture the overshoot,  $D_{\text{eff}} > 80$ , observed in our plots. However, this is of essentially no energetic consequence.) From the data it is evident that  $A$  must be a function of  $(d_i, d_j, e_i, e_j)$ . By examining the inflection points as a function of  $d$  and  $e$ , we arrived at the following functional form:

$$A = c_1 \ln[1.5 + c_2(d - 1.59) + c_3 f] \quad (11)$$

where  $f = 1 - e$  and approximate values are  $c_1 \approx 7.8$ ,  $c_2 = 5.0$ ,  $c_3 = 4.4$ . The  $\ln$  function was chosen because it changes rapidly for small values of the argument and is less sensitive for larger values. The  $f$  term is added to the  $d$  term so that it has a significant effect when  $d$  is small but is much less important when  $d$  is large. For a fully exposed atom pair ( $d = 1.59$ ,  $e = 1$ ),  $A$  should be about 3  $\text{\AA}$ . This fixes the value of the constant term in the brackets to about 1.5. An additional adjustment of the functional form in eq. (2) was made by noting that the  $D_{\text{eff}}$  curves are not symmetric; that is, the upper part of the curve is usually steeper than the lower part (see the low envelope of the data points in Figure 1a). This can be captured by making the exponent  $n$  dependent on  $r$ . The following functional form was found to work well for deeply buried groups:

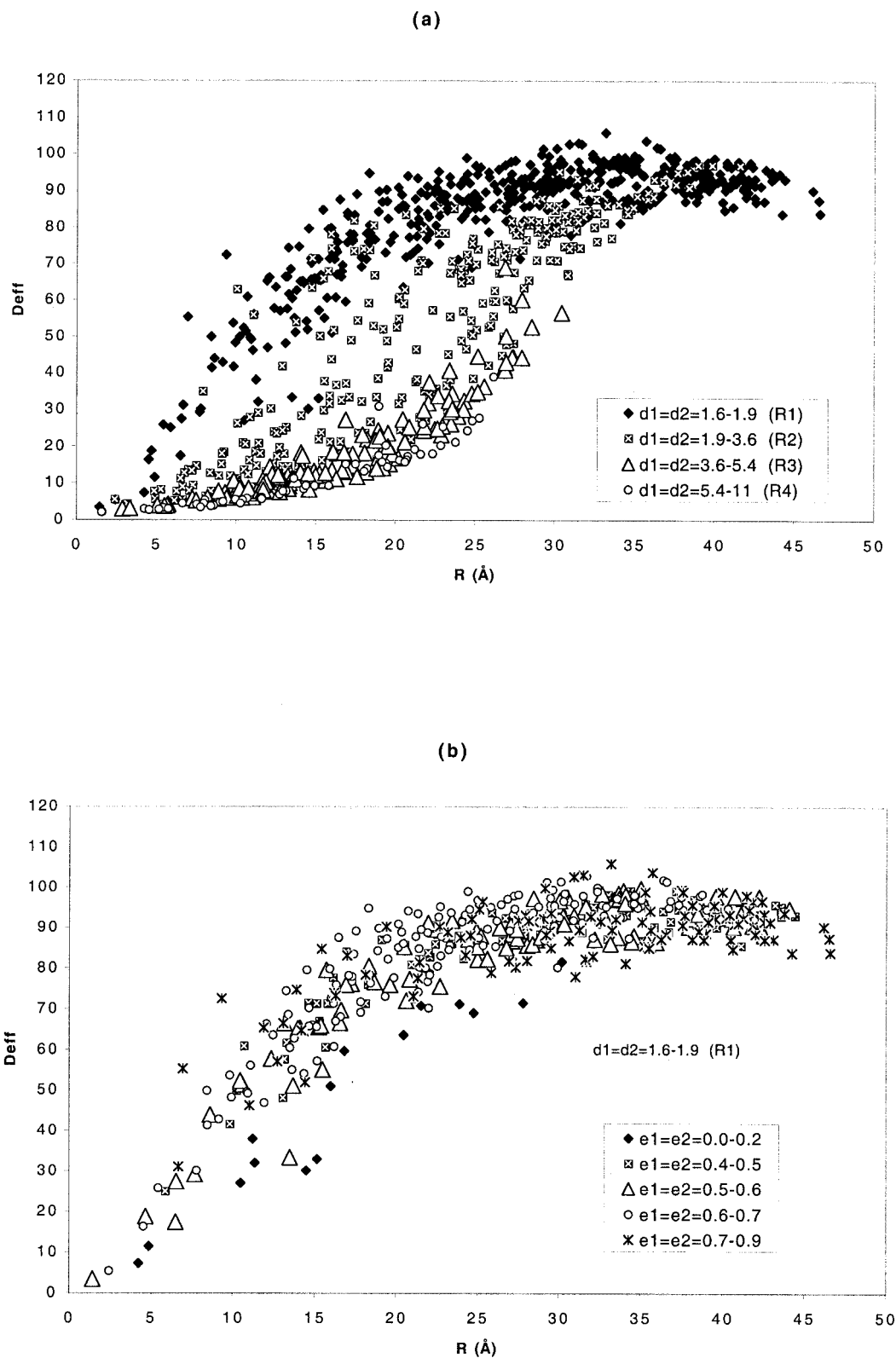
$$n = c_4 + \frac{r^2}{c_5} \quad (12)$$

with approximate values  $c_4 \approx 2$ ,  $c_5 \approx 100$ .

With this model and starting with the above approximate values for the constants we performed a nonlinear least squares fit of all the data (40177 data points) using the Levenberg-Marquardt algorithm<sup>27</sup> to determine the values of  $c_1$ – $c_5$  that minimize  $\chi^2 = \sum (1/D_{\text{eff}} - 1/D_{\text{eff}}^{\text{cal}})^2$ . This choice of  $\chi^2$  biases the fit towards small values of  $D_{\text{eff}}$  and is made because the energy is proportional to  $1/D_{\text{eff}}$ . The fitting was performed using different choices for the “mixing rules” (the dependence of  $d$  and  $f$  on  $d_i, d_j$ , and  $e_i, e_j$ , respectively). Each of the four possible combinations of geometric and arithmetic means for  $d$  and  $f$  was tried. It was found that geometric means for both  $d$  and  $f$  give the lowest  $\chi^2$  (harmonic means work equally well). Therefore

$$d = \sqrt{d_i d_j} \quad \text{and} \quad f = \sqrt{(1 - e_i)(1 - e_j)} \quad (13)$$

The best values for the parameters were found to be  $c_1 = 7.6$ ,  $c_2 = 4.4$ ,  $c_3 = 2.8$ ,  $c_4 = 2.28$ ,  $c_5 = 96$ . Consistent with the



**Figure 3.** (a) Plot of  $D_{\text{eff}}$  versus  $R$  for all data points where the surface distance of the second atom ( $d_2$ ) is within  $\pm 0.5$  Å of that of the reference charge ( $d_1$ ), and the exposure of the second atom ( $e_2$ ) is within  $\pm 0.05$  of that of the reference charge ( $e_1$ ). (b) The data in region R1 grouped according to their exposure.

visual inspection of the plots, the fitting shows that the most important determinant of  $D_{\text{eff}}$  is the distances of the two atoms from the surface. If we eliminate  $f$  ( $c_3 = 0$ ) and refit, we obtain  $c_1 = 6.2$ ,  $c_2 = 10.2$ ,  $c_4 = 2.35$ ,  $c_5 = 97$  with  $\chi^2$  only slightly worse. But if we eliminate  $d$  and make  $A$  a function of only  $f$ ,  $\chi^2$  is much larger. In summary, the proposed function (referred to as DEDIE) is

$$D_{\text{eff}} = 2 + 76 \frac{\left(\frac{r}{A}\right)^n}{1 + \left(\frac{r}{A}\right)^n} \quad (14)$$

where

$$A = 7.6 \ln[1.5 + 4.4(d - 1.59) + 2.8f] \quad (15)$$

and

$$n = 2.28 + \frac{r^2}{96} \quad (16)$$

A comparison of the Poisson  $D_{\text{eff}}$  and the values calculated with this model is shown in Figure 4a. The data are scattered around the diagonal with the scattering increasing as  $D_{\text{eff}}$  increases. At first sight, this plot is discouraging. It is very difficult to predict reliably values of  $D_{\text{eff}}$  between 20 and 80. However, one should keep in mind that even large errors in  $D_{\text{eff}}$  in this range translate to rather small errors in energy. In addition, because there is as much underestimation as overestimation of  $D_{\text{eff}}$ , these errors tend to cancel out when summed up to obtain the total energy. Because the energy is inversely proportional to  $D_{\text{eff}}$ , a more practically meaningful plot is  $1/D_{\text{eff}}$  calculated versus  $1/D_{\text{eff}}$  from Poisson (Figure 4b). This plot exhibits less scatter and a higher correlation coefficient (0.97) than the data in Figure 4a (0.9).

Figure 5a plots electrostatic energies ( $G_{ij} = 332/R_{ij}D_{\text{eff}}$ ) calculated from the model against the Poisson values. The few points with very large interaction energies correspond to bonded atoms. Such interactions are typically excluded in molecular mechanics force fields. The mean square deviation in  $G_{ij}$  values, excluding all 1-2 and 1-3 interactions, is 0.48 kcal/mol.

#### Effect of Interior Dielectric Value

Although the value  $\epsilon_i = 2$  is typical in continuum electrostatics calculations, most molecular mechanics force fields have been parameterized with  $\epsilon_i = 1$ . Therefore, a model derived with  $\epsilon_i = 1$  might be more appropriate to combine with such force fields. To assess the effect of the interior dielectric, we repeated the calculations in subtilisin with  $\epsilon_i = 1$ . From the results (not shown) it is evident that the inflection point of the curve moves to larger distances. Nonlinear squares fit of the data to the model

$$D_{\text{eff}} = 1 + 77 \frac{\left(\frac{r}{A}\right)^n}{1 + \left(\frac{r}{A}\right)^n} \quad (17)$$

gave  $A = 7.6 \ln[1.5 + 5.8(d - 1.59) + 4f]$  and  $n = 2.47 + r^2/89$ .

#### Comparison with Other Functions

The linear distance-dependent model,  $D_{\text{eff}} = r$ , is clearly inadequate, not so much for the fact that it is not sigmoidal and has the wrong asymptotic behavior, but for the fact that the rise in  $D_{\text{eff}}$  with  $r$  is independent of solvent exposure. This model overestimates the screening for buried groups and underestimates it for exposed groups. One might argue that it is thus correct on average. However, if an accurate energetic analysis of burying charges or the relative energy of different conformations of the same protein is desired, this model is deficient. Figure 5b plots the electrostatic interaction energies calculated with  $D_{\text{eff}} = r$  against the Poisson values. It clearly shows larger deviations than our proposed function (mean square deviation = 1.52 kcal/mol, excluding bonded atom pairs). Most of the more sophisticated functions proposed so far have been designed for exposed groups and cannot describe the dielectric screening of buried groups. The function of Sandberg and Edholm,<sup>19</sup> although not sigmoidal, is qualitatively reasonable. Quantitatively, it is good for exposed groups but overestimates  $D_{\text{eff}}$  for buried groups.

A popular alternative to continuum electrostatics is the Generalized Born (GB) model.<sup>28</sup> In its most practical incarnation, this model is entirely analytical.<sup>29–31</sup> The  $D_{\text{eff}}$  implied by the GB model is given by

$$\frac{1}{D_{\text{eff}}} = 1 - \left(1 - \frac{1}{\epsilon}\right) \frac{r_{ij}}{f_{\text{GB}}} \quad \left[ \text{for interior dielectric of 2: } \frac{1}{D_{\text{eff}}} = \frac{1}{2} - \left(\frac{1}{2} - \frac{1}{\epsilon}\right) \frac{r_{ij}}{f_{\text{GB}}} \right] \quad (18)$$

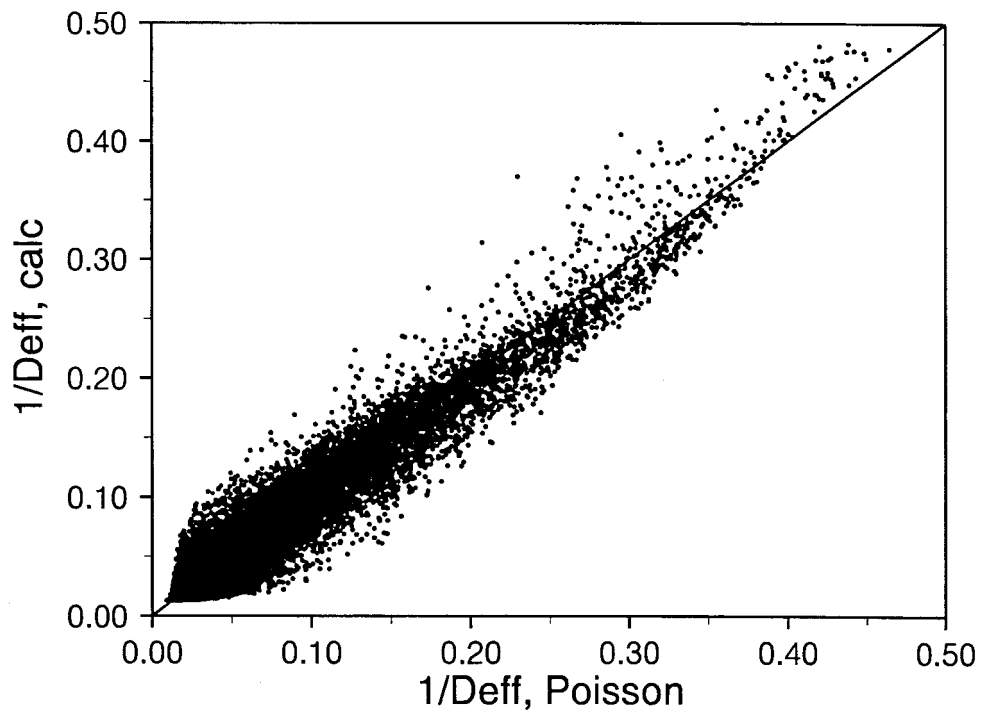
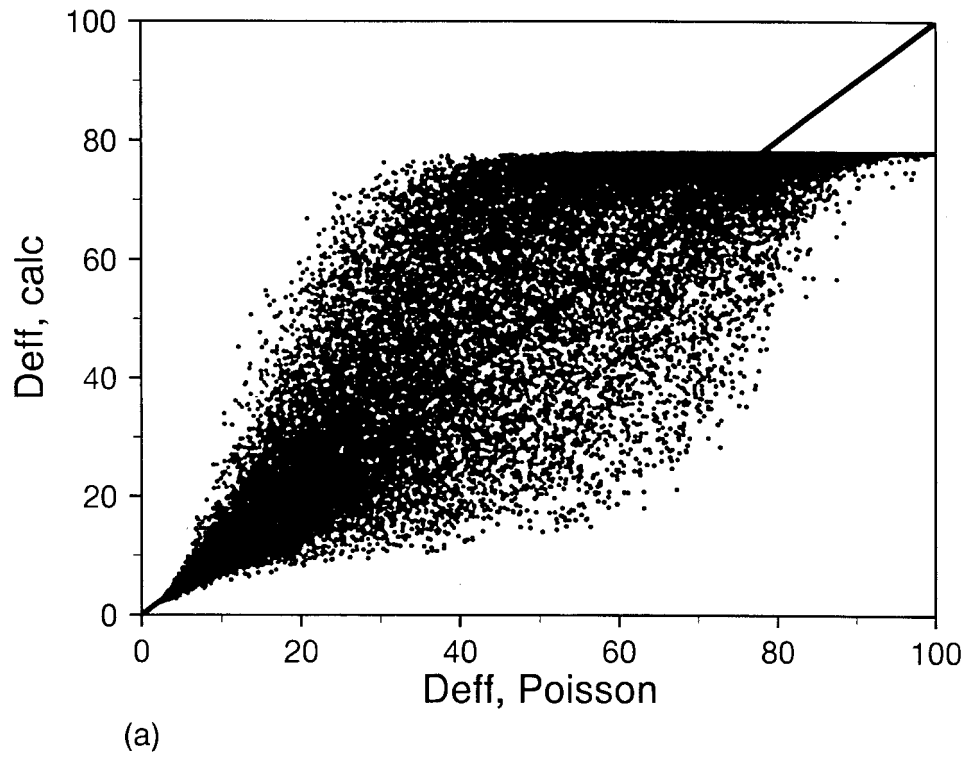
where

$$f_{\text{GB}} = \sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right)} \quad (19)$$

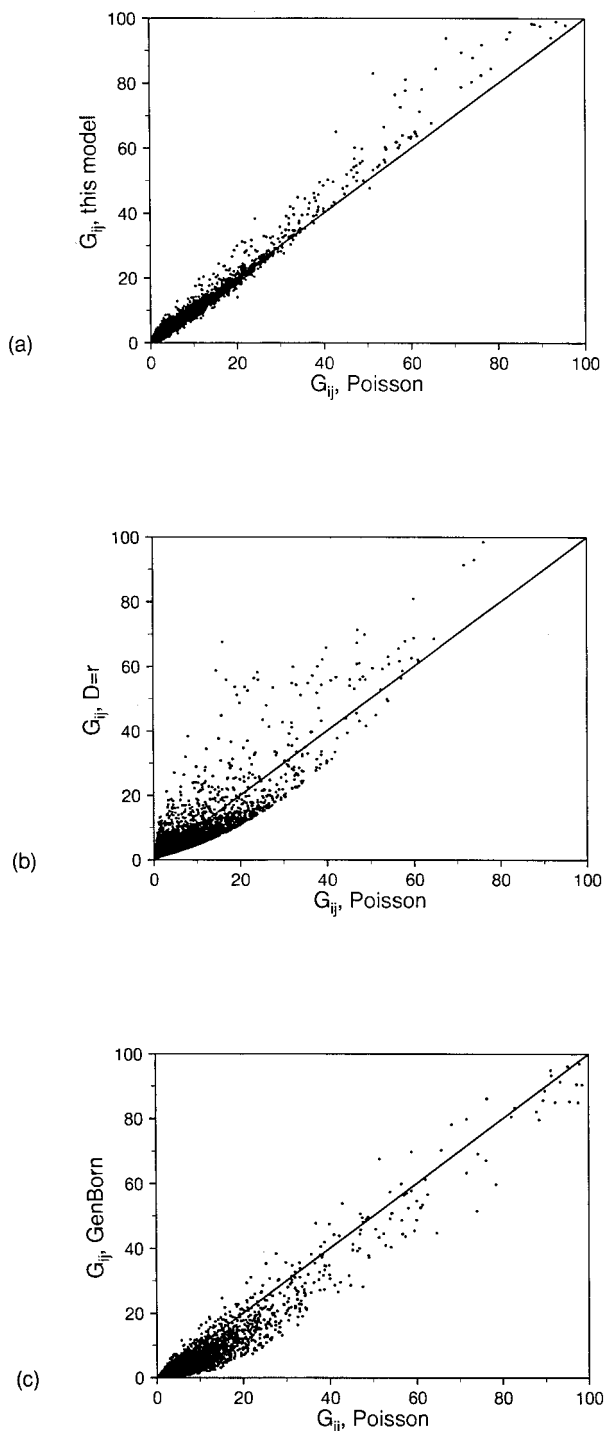
$\epsilon$  is the bulk dielectric constant of the solvent, and  $\alpha_i$  and  $\alpha_j$  are the effective Born radii obtained indirectly from an analytical expression for the polarization energy. As  $r_{ij} \rightarrow \infty$ ,  $f_{\text{GB}} \rightarrow r_{ij}$  and  $D_{\text{eff}} \rightarrow \epsilon$ . We calculated  $D_{\text{eff}}$  values in subtilisin using the version of GB proposed by Still and coworkers<sup>31</sup> and implemented in the CHARMM program.<sup>32</sup> Figure 5c shows electrostatic energies calculated from this model plotted against those obtained from the Poisson equation. The mean square deviation (excluding bonded atom pairs) was 1.45 kcal/mol.

#### Testing on Proteins of Varying Size and Shape

The above function was obtained by fitting data for a typical globular protein of 275 residues. One might wonder whether this function is transferable to other proteins of different size and shape. To examine this issue we calculated  $D_{\text{eff}}$  values in proteins not used in the parameterization: phospholipase c (pdb code 1AH7, 245 residues); barnase (pdb code 1A2P, 110 residues); BPTI (pdb code 4PTI, 58 residues); and a “random coil” structure obtained by



**Figure 4.** Plot of  $D_{\text{eff}}$  (a) and  $1/D_{\text{eff}}$  (b) calculated by the analytical function versus that obtained from the Poisson equation.



**Figure 5.** Electrostatic energies from the Poisson equation versus those obtained by (a) the present function, (b) the linear distance-dependent dielectric model, (c) the analytical GB model.

high temperature unfolding simulations of the villin headpiece domain (pdb code 1vii, 36 residues). Table 1 shows the mean square deviations in electrostatic interaction energies (excluding

1-2 and 1-3 bonded atom pairs) for these systems with the proposed function (DEDIE), the Generalized Born model, and the  $\epsilon = r$  model. These results show that DEDIE is reasonably transferable to proteins of different size and compactness. The deviations tend to increase as the protein size decreases. A likely reason for this deviation is the fact that  $D_{\text{eff}}$  does not depend only on the minimum distance to the surface but also on the average distance to the surface. For instance, a point 6 Å from the protein boundary is less shielded from solvent in a 12 Å diameter protein than in a 40 Å diameter protein. For “random coil” villin the deviations decrease again, probably because the electrostatic interactions in this highly solvent-exposed structure are very small. DEDIE is substantially better than the GB model for large, compact proteins and essentially similar in performance for BPTI. For the unfolded chain GB is superior. The linear distance-dependent dielectric model gives consistently the largest deviations.

#### Limitations of the Model

The proposed model implemented into molecular dynamics simulations will be more computationally expensive than previous distance dependent functions, primarily because it requires the calculation of the distance of every atom from the protein surface. In certain cases, this calculation could be done only once in the beginning of the simulation, as, for example, simulations of native proteins normally do not lead to large changes in atom positions. Also, in simulations of small peptides where all atoms are almost fully exposed, one could use simplified functions, because for exposed atoms the effective dielectric is essentially linear with distance at small distances ( $D_{\text{eff}} \sim 6R$  in Fig. 1d). However, in applications such as evaluation of large numbers of decoys for protein structure prediction,<sup>33</sup> use of the present function will add considerable computational cost. A preliminary estimate from our ongoing implementation in CHARMM is a factor of 4 over a simulation in vacuum.

The dielectric screening between two charges is a complex function of the position of the two charges and the size and shape of the low dielectric region in which they reside. This dependence cannot be fully captured by the five independent variables of our proposed function and this is the cause of the scatter observed in Figure 4 (although some of the scatter may result from inaccuracies of the numerical solution of the Poisson equation). One can easily think of other parameters that might correlate with  $D_{\text{eff}}$ . One example, mentioned above, is the *average* distance to the surface. Other geometrical parameters could be devised that describe the distribution of electric field lines between the two points. These might improve the analytical fit of the data, but would increase the

**Table 1.** Mean Square Deviations in Electrostatic Interaction Energies (kcal/mol).

	SBT	PLC	Barnase	BPTI	Unf. Villin
DEDIE	0.48	0.50	0.79	1.25	0.70
GB	1.45	1.10	1.13	1.26	0.28
RDIE	1.52	1.36	1.80	2.33	4.77

computational cost. We feel that the proposed function strikes a good balance between accuracy and computational efficiency.

## Conclusions

Effective dielectric constants obtained by continuum electrostatics strongly depend on the distance between the two charges and the distance of each atom from the surface. A secondary determinant for solvent exposed atoms was found to be the "exposure," that is the proportion of the surrounding space that is occupied by solvent. An analytical function of these five variables was designed that reproduces the data reasonably well, especially for lower values of  $D_{\text{eff}}$ . While most earlier functions have difficulties when they are applied to buried groups, the present function provides a smooth transition from the low dielectric screening pertaining to buried groups to the high screening of surface groups and is a good candidate for implementation in molecular dynamics simulations. The function is expected to be transferable to proteins of different size and compactness.

## Acknowledgments

Helpful comments on the manuscript by Prof. Marilyn Gunner and an informative communication with Dr. L. Sandberg are gratefully acknowledged.

## References

1. Honig, B.; Nicholls, A. *Science* 1995, 268, 1144.
2. Sharp, K.; Honig, B. *Ann Rev Biophys Biophys Chem* 1990, 19, 301.
3. Tanford, C.; Kirkwood, J. G. *J Am Chem Soc* 1957, 79, 5333.
4. Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* 1986, 1, 47.
5. Roux, B. Personal communication.
6. Sharp, K. *J Comp Chem* 1991, 12, 454.
7. Gilson, M. K.; McCammon, J. A.; Madura, J. D. *J Comp Chem* 1995, 16, 1081.
8. Warshel, A.; Levitt, M. *J Mol Biol* 1976, 103, 227.
9. Warshel, A. *Photochem Photobiol* 1979, 30, 285.
10. van Duijnen, P. T.; Thole, B. T.; Hol, W. G. *J Biophys Chem* 1979, 9, 273.
11. Gelin, B. R.; Karplus, M. *Biochemistry* 1979, 18, 1256.
12. Northrup, S. H.; Pear, M. R.; Morgan, J. D.; McCammon, J. A.; Karplus, M. *J Mol Biol* 1981, 153, 1087.
13. Warshel, A.; Russel, S. T.; Churg, A. K. *Proc Natl Acad Sci USA* 1984, 81, 4785.
14. Mehler, E. L.; Eichele, G. *Biochemistry* 1984, 23, 3887.
15. Hingerty, B. E.; Ritchie, R. H.; Ferrell, T. L. *Biopolymers* 1985, 24, 427.
16. Ramstein, J.; Lavery, R. *Proc Natl Acad Sci* 1988, 85, 7231.
17. Pickersgill, R. W. *Prot Engin* 1988, 2, 247.
18. Guarnieri, F.; Schmidt, A. B.; Mehler, E. L. *Int J Quantum Chem* 1998, 69, 57.
19. Sandberg, L.; Edholm, O. *Proteins* 1999, 36, 474.
20. Mehler, E. L.; Warshel, A. *Proteins* 2000, 40, 1.
21. Simonson, T. *J Am Chem Soc* 1998, 120, 4875.
22. Lazaridis, T.; Karplus, M. *Proteins* 1999, 35, 133.
23. Nicholls, A.; Honig, B. *J Comp Chem* 1991, 12, 435.
24. Gilson, M. K.; Rashin, A.; Fine, R.; Honig, B. *J Mol Biol* 1985, 183, 503.
25. Alexov, E. G.; Gunner, M. R. *Biophys J* 1997, 74, 2075.
26. Hill, T. L. *J Phys Chem* 1956, 60, 253.
27. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes, Fortran version*; Cambridge University Press: Cambridge, 1989.
28. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J Am Chem Soc* 1990, 112, 6127.
29. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem* 1996, 100, 19824.
30. Schaefer, M.; Karplus, M. *J Phys Chem* 1996, 100, 1578.
31. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J Phys Chem A* 1997, 101, 3005.
32. Dominy, B. N.; Brooks III, C. L. *J Phys Chem B* 1999, 103, 3765.
33. Lazaridis, T.; Karplus, M. *J Mol Biol* 1999, 288, 477.