J|A|C|S
ARTICLES
Published on Web 03/21/2003

A Sphere-Based Model for the Electrostatics of Globular Proteins

Philipp Werner and Amedeo Caflisch*

Contribution from the Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received August 15, 2002; Revised Manuscript Received October 29, 2002; E-mail: caflisch@bioc.unizh.ch

Abstract: We propose a model for the electrostatics of globular proteins in which the low dielectric region is replaced by concentric spheres of the appropriate size. The method uses analytical formulas for the dielectric sphere and allows an efficient and accurate treatment of bulk charges. For surface charges, we propose a numerical determination of the sphere radius based on the solvent exposure of the individual atoms. The present implementation of the sphere model yields a good approximation of finite-difference Poisson solvation and interaction energies for a test set of 12 proteins.

1. Introduction

An accurate and efficient evaluation of electrostatic energies of macromolecules in aqueous solution is essential for many applications in computational structural biology, such as molecular dynamics simulations or structure prediction methods.[1] In the continuum dielectric approximation, a correct evaluation of the electrostatic energy of solvated macromolecules requires the solution of Poisson's equation. While iterative numerical calculations yield a good approximation of the potential,[2−5] their computational cost is prohibitively high for many important applications.[6] Thus, several simplified models have been developed. One of the most popular is the generalized Born (GB) approach,[7,8] which for every solute atom requires the evaluation of its effective Born radius by integration of the energy density over the solute volume. Several GB implementations have appeared in the literature with analytical[9−13] and numerical[7,14,15] evaluation of the effective Born radii. Numerical approaches are more accurate,[13,16] but only analytical implementations can be used in molecular dynamics simulations.[17,18] Most of the GB implementations make use of the Coulomb field approximation, which for macromolecules may result in a significant underestimation of the reaction field especially for atoms near the surface.[8,19] Recently, empirical corrections to this approximation have been proposed.[13,15]

Extremely crude approximations have been suggested for the electrostatic interaction energy mainly for efficiency reasons.[20,21] Although they cannot be derived from physical principles, distance dependent dielectric functions such as $\epsilon(r) \propto r$ have been used with some success in molecular dynamics simulations.[22−24] One could be tempted to develop a physically more accurate expression for the effective dielectric constant. It is, however, immediately clear that a function of r alone will not be adequate. To show this, one can consider a sphere of dielectric constant ϵ_m and radius R surrounded by water of dielectric constant ϵ_w. For a point charge at the center of the sphere, the potential ϕ is easily obtained, and the effective dielectric constant defined as $\epsilon(r) = 1/(r\phi(r))$ becomes

$$\epsilon(r) = \frac{\epsilon_m}{1 + \left(\frac{\epsilon_m}{\epsilon_w} - 1\right)\frac{r}{R}}, \quad r \le R \tag{1}$$

It follows that ϵ depends on the quantity (r/R), that is, on the position of the second charge relative to the radius R of the sphere.

(1) Lazaridis, T.; Karplus, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139−145.
(2) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.
(3) Gilson, M. K.; Honig, B. H. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7−18.
(4) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(5) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.
(6) Gilson, M. K.; McCammon, J. A.; Madura, J. D. *J. Comput. Chem.* **1995**, *16*, 1081−1095.
(7) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(8) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(9) Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(10) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(11) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(12) Dominy, B. N.; Brooks, C. L., III. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.
(13) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606−10614.
(14) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098−8106.
(15) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.
(16) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem. B* **1997**, *101*, 1190−1197.
(17) Bursulaya, B. D.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1999**, *121*, 9947−9951.
(18) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275−291.
(19) Schaefer, M.; Froemmel, C. *J. Mol. Biol.* **1990**, *216*, 1045−1066.
(20) Gelin, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2002−2006.
(21) Mehler, E. L. *Protein Eng.* **1990**, *3*, 415−417.
(22) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.
(23) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780−10785.
(24) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 24−33.

10.1021/ja021093i CCC: $25.00

While we do not wish to present the following developments as a search for an effective dielectric constant, it was the above example which inspired the idea to replace globular proteins by a sphere of appropriate volume and use this simplified framework to derive analytical expressions for the interaction and solvation energies. An analytical theory of electrostatic energies based on a spherical solute was proposed several decades ago[25] and later extended to account for differences in the solvent accessibility of side chains[26] and atoms.[27] The novel aspect in the present paper is the replacement of the solute volume by concentric spheres whose radii are determined using information about the location and structure of the molecular surface. An advantage of this approach is the simple and physically plausible treatment of bulk charges. For exposed charges, the method used to adjust the sphere radius is of critical importance. We show that a simple strategy based on the solvent exposure of the individual atoms yields good agreement with interaction and solvation energies obtained by finite-difference Poisson (fdP) calculations.

2. Theory

Consider a macromolecule of dielectric constant ϵ_m and volume V_m. If the shape of the molecule is almost spherical, it seems reasonable to approximate the low dielectric volume by a sphere of dielectric constant ϵ_m and radius R, such that $(4/3)\pi R^3 \approx V_m$ (the appropriate choice of R for arbitrary molecules will be the subject of section 3.1). The origin of this sphere coincides with the center of geometry of the molecule, and coordinates will henceforth be defined relative to this point. To avoid charges outside the sphere or account for the local surface structure, the radius R may have to be modified. The quality of our method will depend on the details of this rescaling, but the following calculations are not affected, so we postpone the heuristic derivation of a possible prescription to section 3.2.

For a system of point charges, the electrostatic energy is the sum of all pair interactions

$$E_{\mathrm{int},i,j} = \frac{1}{4\pi}\int_{\mathbb{R}^3}\epsilon(\vec{x})\vec{E}_i(\vec{x})\cdot\vec{E}_j(\vec{x})\mathrm{d}^3x \tag{2}$$

and the self-energy contributions of the individual charges

$$E_{\mathrm{self},i} = \frac{1}{8\pi}\int_{\mathbb{R}^3}\epsilon(\vec{x})\vec{E}_i^{\,2}(\vec{x})\mathrm{d}^3x \tag{3}$$

where $\epsilon(\vec{x})$ is a location-dependent dielectric constant, and $\vec{E}_i(\vec{x})$ is the electric field produced by charge i.[28] Because self-energies of point charges are infinite, we subtract the constant value of the self-energy in a homogeneous medium of dielectric constant ϵ_m to obtain a finite result, whose physical interpretation is the solvation energy of the sphere with charge i switched on:

$$E_{\mathrm{solv},i} = \frac{1}{8\pi}\int_{\mathbb{R}^3}(\epsilon(\vec{x})\vec{E}_i^{\,2}(\vec{x}) - \epsilon_m\tilde{\vec{E}}_i^{\,2}(\vec{x}))d^3x \tag{4}$$

$\vec{E}$ and $\tilde{\vec{E}}$ denote the electric fields in the inhomogeneous and homogeneous medium, respectively.

Analytical expressions can be derived for (2) and (4) because of the spherical geometry of the solute region. The calculation is carried out in Appendix A. If we denote the Cartesian coordinates of atom i by $\vec{x}_i$, its charge by q_i, and the relative positions of atoms i and j by their respective radial distances r_i and r_j, as well as an angle θ, the resulting formulas read (r_i, $r_j \le R$)

$$E_{\mathrm{int},i,j} = \frac{q_iq_j}{\epsilon_m|\vec{x}_i - \vec{x}_j|} - \frac{q_iq_j}{R}\left(\frac{1}{\epsilon_m} - \frac{1}{\epsilon_w}\right)\sum_{l=0}^{\infty}C_l\left(\frac{r_ir_j}{R^2}\right)^l P_l(\cos\theta) \tag{5}$$

and

$$E_{\mathrm{solv},i} = -\frac{q_i^2}{2R}\left(\frac{1}{\epsilon_m} - \frac{1}{\epsilon_w}\right)\sum_{l=0}^{\infty}C_l\left(\frac{r_i}{R}\right)^{2l} \tag{6}$$

where P_l denotes the Legendre polynomial of rank l, and the coefficient C_l is

$$C_l = \frac{1}{1 + \frac{l}{l+1}\frac{\epsilon_m}{\epsilon_w}} \tag{7}$$

The infinite series in (5) and (6) may be simplified using the fact that $\epsilon_m \ll \epsilon_w$. Setting

$$C_l \approx \frac{1}{1 + \frac{\epsilon_m}{\epsilon_w}} \quad \text{for} \quad l \ge 1 \tag{8}$$

and using the relation $\sum_{l=0}^{\infty}x^lP_l(y) = (1 - 2xy + x^2)^{-1/2}$, one can express the results (5) and (6) in the closed forms

$$E_{\mathrm{int},i,j} = \frac{q_iq_j}{\epsilon_m|\vec{x}_i - \vec{x}_j|} - \frac{q_iq_j}{R}\frac{\frac{1}{\epsilon_m} - \frac{1}{\epsilon_w}}{1 + \frac{\epsilon_m}{\epsilon_w}}\left(\frac{\epsilon_m}{\epsilon_w} + \frac{1}{\sqrt{1 - 2\frac{r_ir_j}{R^2}\cos\theta + \left(\frac{r_ir_j}{R^2}\right)^2}}\right) \tag{9}$$

and

$$E_{\mathrm{solv},i} = -\frac{q_i^2}{2R}\frac{\frac{1}{\epsilon_m} - \frac{1}{\epsilon_w}}{1 + \frac{\epsilon_m}{\epsilon_w}}\left(\frac{\epsilon_m}{\epsilon_w} + \frac{1}{1 - \left(\frac{r_i}{R}\right)^2}\right) \tag{10}$$

In fact, slightly more accurate expressions could be obtained using the approximation $(1 + l/(l+1)(\epsilon_m/\epsilon_w))^{-1} \approx 1 - l/(l+1)(\epsilon_m/\epsilon_w)$ instead of (8) as shown in Appendix B. The accuracy of (9) and (10) is sufficient, however, and we will therefore build our model on the basis of these simple formulas. They allow a new approach to the modeling of electrostatic solvation and interaction energies which seems particularly suitable for large, globular proteins.

The sphere radius R is a quantity whose precise value for each of the (pairs of) charges has yet to be defined. A careful determination of R would, in principle, allow one to reproduce the fdP interaction and solvation energies by means of (9) and (10), respectively. Strategies of varying degree of sophistication

(25) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351−361.
(26) Shire, S. J.; Hanania, G. I. H.; Gurd, F. R. N. *Biochemistry* **1974**, *13*, 2967−2974.
(27) Havranek, J. J.; Harbury, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11145−11150.
(28) Jackson, J. D. *Classical Electrodynamics*; John Wiley & Sons: New York, 1975.

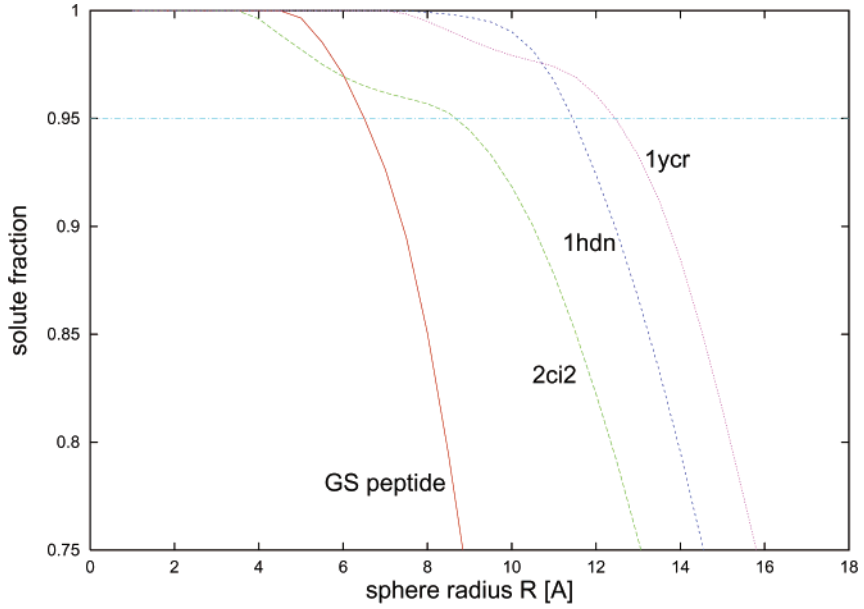

Figure 1. Definition of the sphere radius for the core region. A calculation based on ellipsoid semi-axes yields results comparable to $R_{0.95}$ obtained by choosing $solute\_fraction(R_{0.95}) = 0.95$.

and computational efficiency could be imagined. In this paper, we will propose a fairly simple, semianalytical method based on the solvent exposure of the individual atoms.

An extension of the sphere model to a spheroidal model would reduce the need for and extent of rescaling. In fact, analytical expressions can also be derived for a low dielectric region of spheroidal shape, as shown in Appendix C. For surface charges, a careful adjustment of the dielectric boundary position as a function of solvent exposure (or some other parameter) would, however, still be inevitable. To keep matters simple, we decided to stick with the sphere model.

3. Implementation

3.1. Definition of the Sphere Radius for Charges in the Core Region. In the core region of the molecule, the detailed structure of the molecular surface may be ignored, and a constant value of R may be used. To avoid systematic shifts in energy, the value of R has to be chosen correctly. Because most globular molecules are not spherical, but rather ellipsoidal in shape, good results for the bulk electrostatics are obtained by choosing the sphere radius roughly equal to the minor semi-axis. This value may be obtained either by diagonalizing the mass of inertia tensor of the molecule (assuming a uniform density) and rescaling the axes to preserve the volume or numerically by choosing R such that a certain percentage of the sphere volume is occupied by the solute. Figure 1 illustrates the fraction of volume occupied by solute as a function of R for the four molecules 1ycr, 1hdn, 2ci2, and a three-stranded β-sheet (GS peptide[29]). While the first three examples (636−1002 atoms) are fairly large and of roughly spheroidal shape, the 20-residue GS peptide does not really fall into the class of globular proteins and is considered here to assess the quality of the method when applied to cases for which it was originally not designed.

A calculation based on ellipsoid semi-axes yields values of R which approximately correspond to a solute fraction of 0.95. Hence, we used the latter criterion to determine the sphere radius for the core region ($R_{0.95}$).

3.2. Definition of the Sphere Radii for Surface Charges. 3.2.1. Charges Exposed to the Solvent. A major disadvantage of replacing a molecule by a geometrically simple object such as a sphere is the loss of information about the local surface structure. Even small water-filled cavities have a large screening effect and dramatically reduce/enhance the interaction/solvation energies of nearby solute charges. To take this effect into account, we calculate the "solvent exposure" ρ_i of all atoms i in the molecule. For surface atoms ($\rho_i > 0$), we determine the sphere radius R in such a way that charges, whose solvent exposure is large, will end up close to the dielectric discontinuity surface, as illustrated in Figure 2.

Simple analytical considerations yield some guidance in the search for an appropriate prescription. We define the solvent exposure ρ_i of atom i as the fractional volume of a shell around atom i, which is occupied by solvent. For the thickness Δ of the shell, we used $\Delta = r_{vdW,i}$. Similar prescriptions (and results of comparable quality) may be obtained by choosing a constant value such as $\Delta = 1$ Å.

First, consider a completely solvated sphere of radius r_{vdW} and unit charge. Its solvation energy (assuming $\epsilon_m = 1$, $\epsilon_w = \infty$) is $-(2r_{vdW})^{-1}$. To reproduce this result with (10), we have to adjust the sphere radius R such that

$$-\frac{1}{2R}\frac{1}{1-\left(\frac{r}{R}\right)^2} = -\frac{1}{2r_{vdW}} \tag{11}$$

Supposing $(r_{vdW}/r) \ll 1$, it follows from (11) that

$$R \approx r + \frac{r_{vdW}}{2} \quad \text{for} \quad \rho \approx 1 \tag{12}$$

For small values of ρ, the distance of the charge from the surface of the sphere may be estimated from the volume fraction of the penetrating shell segment of height $h = r + 2r_{vdW} - R$. This

(29) De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854−865.

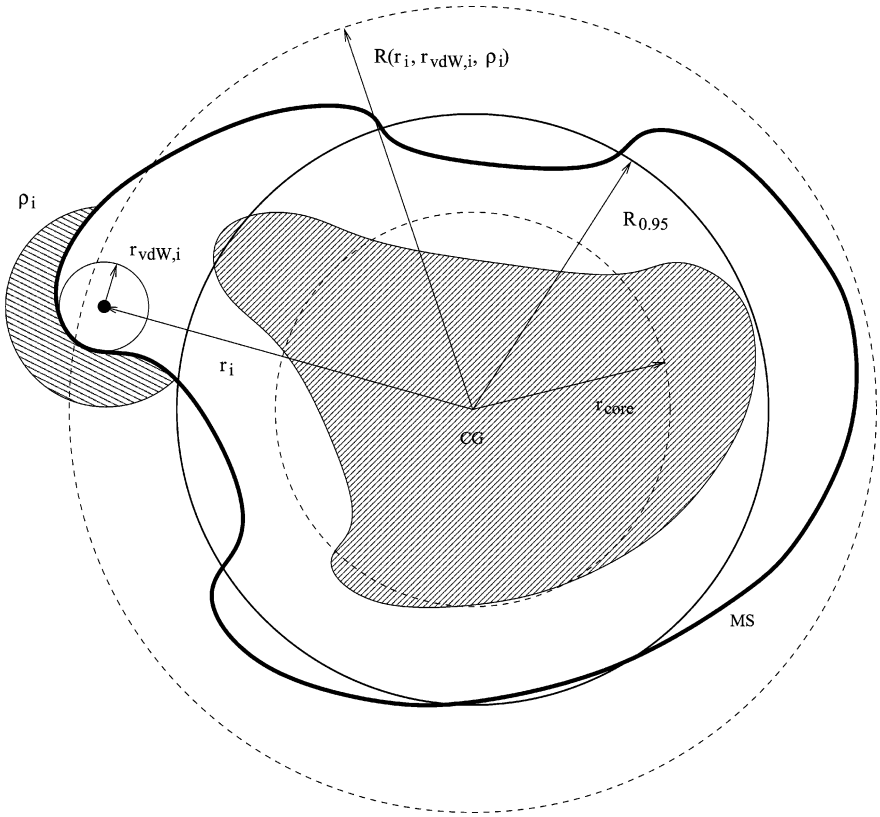

Figure 2. Illustration of the strategy used to adjust the sphere radius R according to the shape of the molecular surface (MS) and the position of the charges. $R_{0.95}$ is the radius of the sphere, centered on the center of geometry (CG), whose volume is occupied to 95% by solute. The hatched region inside the molecule indicates bulk atoms. Electrostatic energies of bulk atoms in the core region (small dashed circle) are calculated using $R_{0.95}$. For charges which are exposed to the solvent ($\rho_i > 0$), R is modified in such a way that the charge position ends up close to the surface of the sphere (large dashed circle). The solvent exposure ρ_i of atom i is indicated by the hatched region outside of the molecular surface.

yields

$$R \approx r + 2r_{\mathrm{vdW}}(1 - \alpha\sqrt{\rho}) \quad \text{for} \quad \rho \approx 0 \tag{13}$$

with $\alpha = \sqrt{7/6}$. To define a function $R(\rho)$ valid in the interval $0 \le \rho \le 1$, which smoothly approaches (12) for $\rho \to 1$, we used the modification

$$\alpha\sqrt{\rho} \to \frac{3}{4}\sqrt{\frac{1+\delta}{\delta}}\sqrt{\rho - \frac{1}{1+\delta}\rho^{1+\delta}} \tag{14}$$

in (13). The resulting $R(\rho,\delta)$ satisfies $\partial R/\partial\rho(\rho = 1) = 0$, and the parameter δ determines how rapidly $R(\rho,\delta)$ decreases as a function of ρ. We found that small values of δ produced the best results. In the limit $\delta \to 0$, one obtains the formula

$$R = r + 2r_{\mathrm{vdW}}\left\{1 - \frac{3}{4}\sqrt{\rho - \rho\ln(\rho)}\right\} \tag{15}$$

which works reasonably well for the interval $0 \le \rho \le 1$.

Even though the previous arguments apply to solvation energies only, slight modifications of the semiempirical formula (15) (see Results section) were implemented as the prescriptions for both solvation and interaction energies. In the case of interaction energies, the situation is somewhat complicated by the fact that two charges i_1 and i_2 have to be considered. If both charges are exposed to the solvent, we first calculate

$$\delta r = \min_{i=i_1,i_2}\left(2r_{\mathrm{vdW},i}\left\{1 - \frac{3}{4}\sqrt{\rho_i - \rho_i\ln(\rho_i)}\right\}\right) \tag{16}$$

and then adjust the radius according to

$$R = \max_{i=i_1,i_2}(r_i) + \delta r \tag{17}$$

If only one charge is exposed to the solvent, r_{vdW} and ρ of the exposed atom are used to calculate δr.

Alternative strategies to account for solvent accessibility in sphere-based models can be found in the work by Shire et al.[26] and in a more recent paper by Havranek and Harbury.[27]

3.2.2. Bulk Charges Outside the Core Region. Even if an atom is not exposed to the solvent, an R larger than $R_{0.95}$ may still be required. In Figure 2, these atoms correspond to that part of the hatched bulk region, which lies outside the core region, defined by $r_{\mathrm{core}} = R_{0.95} - 2r_{\mathrm{vdW}}$. Increasing the sphere radius according to

$$R = \max(R_{0.95}, r_i + 2r_{\mathrm{vdW},i}) \tag{18}$$

works quite well for interaction energies. In the case of solvation energies and small van der Waals radii, the relative error can be high. A simple way to more accurately adjust the sphere radius consists of evaluating the solvent fraction in several shells of varying thicknesses. Our results were obtained using shell thicknesses of $\Delta = 1$ Å and $\Delta = 2$ Å in addition to $\Delta = r_{\mathrm{vdW}}$.

4. Preparation of Data Sets

To test our model, we used 12 proteins for which fdP data were available from an unrelated investigation. The molecular surface was used to define the low dielectric boundary in the

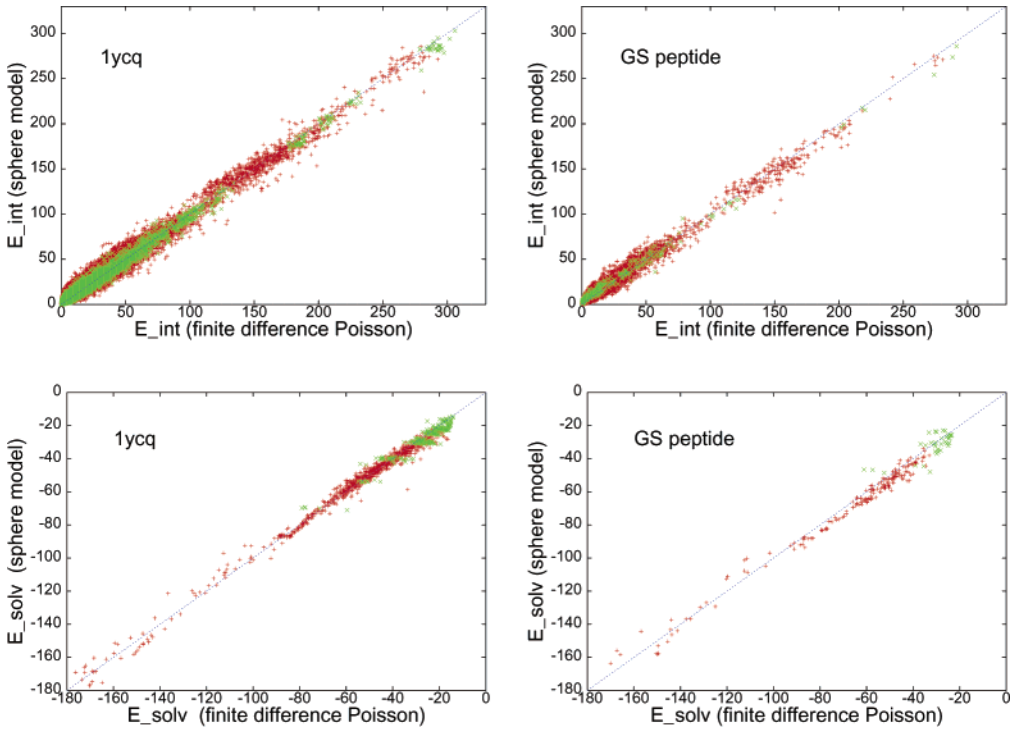

Figure 3. Comparison of electrostatic energies obtained from the sphere model and from the numerical solution of Poisson's equation for $\epsilon_w = 80$, $\epsilon_m = 1$, and unit charges. Green crosses and red plus signs represent charges in the core and surface region, respectively. All energy values are in kcal/mol.

sphere model, GB, and fdP calculations. The molecular surface is spanned by the *surface* of a water probe of radius 1.4 Å; it is preferred over other surface definitions (e.g., solvent accessible surface) because it gives the best agreement between fdP results and experimental values.[14] We set $\epsilon_m = 1$ and $\epsilon_w = 80$ in all of the calculations. Van der Waals radii were taken from the CHARMM param19 set[30] and ranged from 0.6 to 2.4 Å.

4.1. Sphere Model. Shells of thickness r_{vdW} were used to calculate the solvent exposures ρ_i of the individual atoms. $R_{0.95}$ was determined by increasing the sphere size in steps of 0.25 Å until the solute fraction fell below 0.95. All numerical integrations were performed by summation on a 0.25 Å grid.

4.2. Finite-Difference Poisson. The fdP calculations were performed with the PBEQ module[31] in CHARMM.[30] Pair interaction energies were obtained by calculating the potential for each charge separately and evaluating this potential at the positions of the other charges. Solvation energies were calculated by subtracting the vacuo self-energy from the self-energy in solution. To obtain the self-energies, the fdP potential was evaluated at the charge position itself (which yields a finite value because of the charge discretization on the grid) and multiplied by a factor of $^1/_2$. For each charge, both self-energy values were calculated using the same grid to eliminate effects resulting from the distribution of the charge onto the lattice sites. The grid spacing used in the fdP calculations was 0.2 Å.

4.3. Generalized Born. The program SEED[32] was used for the GB calculations. SEED performs a numerical integration of the energy density on a cubic grid to evaluate the effective Born radii.[14] A grid spacing of 0.1 Å was used.

5. Results

To judge the quality of the present implementation of the sphere model, the interaction and solvation energies of 12 proteins were compared to the values obtained by fdP. The PDB[33] codes of the 12 proteins are 1a2p (1073 atoms, $R_{0.95} =$ 10.0 Å), 1ycr (1002, 12.5), 1ycq (979, 11.3), 1pht (813, 10.3), 1hdn (785, 11.5), 1ubq (746, 11.3), 2ci2 (636, 9.0), 2ptl (575, 9.3), 1bpi (568, 8.0), 1pgb (535, 8.8), 1crn (396, 7.0), and the GS peptide (215, 6.5).

No systematic optimization of parameters was performed. The prescriptions for the adjustment of the sphere radius R were based on formula (15). Slight modifications ($^3/_4 \rightarrow 0.77$ and a lower bound at $r + 0.6r_{vdW}$ for solvation energies, $^3/_4 \rightarrow 0.79$ for interaction energies) were made to improve the results for one protein (1hdn) and then kept fixed for the evaluation of the other molecules.

All pair interaction energies and solvation energies were calculated for unit charges. The quality of our approximation turned out to be similar for all 12 molecules (Table 1). We therefore only show the results for the large, roughly spheroidal structure 1ycq (axes ratio $\approx$ 2:3) and the small, clearly nonspherical GS peptide. The upper panels in Figure 3 compare the interaction energies, while the lower panels show the results for solvation energies. The red data points in these figures indicate atoms (or pairs of atoms) exposed to the solvent, while green data points show the results for bulk atoms. Two different colors were chosen because of the distinct nature of the prescriptions. The discrete levels in the bulk solvation energies originate from bulk charges outside the core region (see section 3.2.2). Whenever two such atoms have the same van der Waals

(30) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(31) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(32) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88−105.
(33) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.

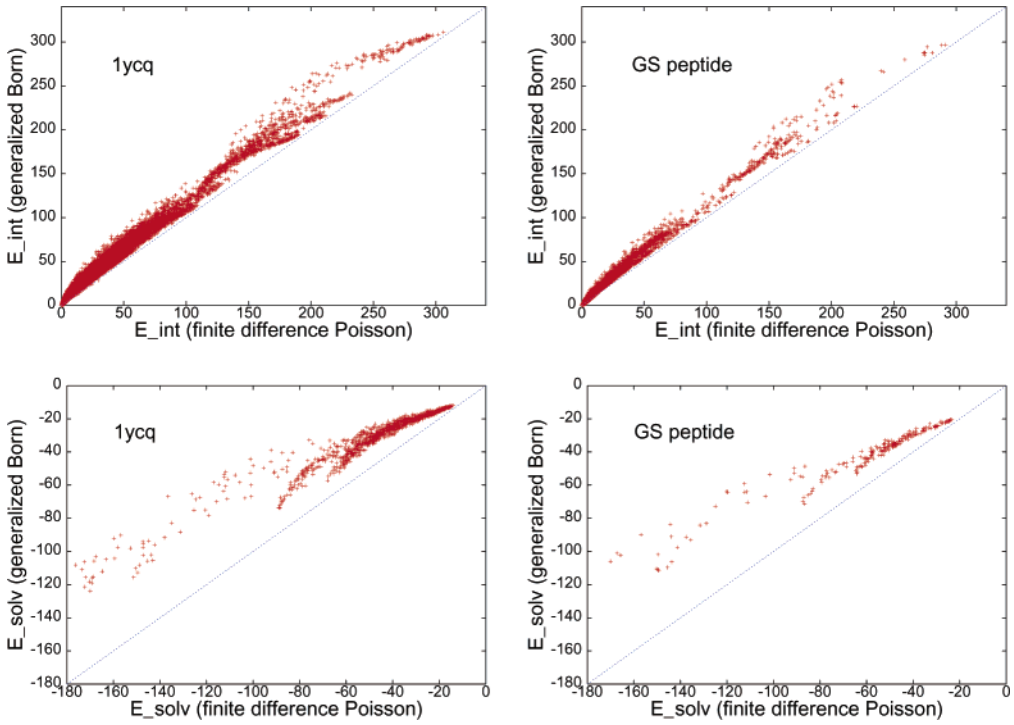

Figure 4. Comparison of electrostatic energies obtained from the generalized Born model and from the numerical solution of Poisson's equation for $\epsilon_w = 80$, $\epsilon_m = 1$, and unit charges. All energy values are in kcal/mol.

Table 1. Comparison of the Sphere Model and GB with fdP

	E_{int}				E_{solv}			
	sphere[a]		GB[b]		sphere[a]		GB[b]	
	slope	correl	slope	correl	slope	correl	slope	correl
1a2p	0.988	0.984	1.203	0.989	0.960	0.992	0.607	0.967
1bpi	0.993	0.984	1.201	0.991	0.972	0.994	0.652	0.971
1crn	0.991	0.987	1.187	0.993	0.977	0.992	0.590	0.962
1hdn	0.994	0.988	1.199	0.990	0.980	0.996	0.637	0.975
1pgb	1.000	0.987	1.200	0.992	0.967	0.993	0.611	0.968
1pht	0.991	0.986	1.198	0.990	0.959	0.993	0.622	0.966
1ubq	0.991	0.988	1.195	0.991	0.975	0.995	0.618	0.970
1ycq	0.991	0.986	1.205	0.990	0.968	0.994	0.640	0.973
1ycr	0.995	0.988	1.193	0.991	0.960	0.994	0.614	0.977
2ci2	0.996	0.986	1.200	0.991	0.964	0.993	0.624	0.963
2ptl	0.997	0.987	1.195	0.991	0.970	0.994	0.631	0.975
GS	0.972	0.989	1.168	0.995	0.981	0.994	0.609	0.974

[a] Slopes and correlation coefficients obtained by comparing the sphere model and fdP. [b] Slopes and correlation coefficients obtained by comparing the GB approach and fdP.

radius, definition (18) of the sphere radius yields identical distances from the dielectric boundary and hence similar solvation energies.

The solvent-exposure parameter ρ yields information on the local structure of the solute−solvent boundary. As it appears in Figure 3, this information proves sufficient to quite accurately reproduce solvation energies even for the GS peptide which consists of only 20 residues. The quality of the interaction energies is somewhat less impressing. This is evident also in Table 1, which contains correlation values and slopes of the fitting lines, and is partly due to the fact that interaction energies may depend on the global shape of the molecule and hence cannot be adequately treated on the basis of local surface information alone. A parameter derived from the solute distribution in a larger neighborhood of the charges as well as a prescription specifically tailored to interaction energies would be required to improve the estimates for weak (highly screened) interactions.

Figure 4 compares the interaction and solvation energies calculated by the GB method to the values obtained by fdP. The GB approach used here is based on the Coulomb field approximation. This leads to an overestimation of the effective radii and hence to a systematic overestimation of interaction energies and underestimation of solvation energies. The significant deviation in the slope of the GB solvation energies is consistent with analytical results for simple solute geometries.[8] The values in Table 1 indicate that the sphere model reproduces fdP solvation energies more accurately than GB. On the other hand, GB yields a slightly better correlation with fdP for interaction energies due to the smaller spread of points in the highly populated low energy region of the plot.

Both the GB method and our sphere model are clearly superior to the simple approximation $\epsilon(r) = r$, which completely fails to evaluate high energy interactions correctly (results not shown).

The computation time in the present implementation of the sphere model is dominated by the few minutes needed to calculate the molecular surface. Because the sphere model only involves integrations over shells around each atom and not over the entire solute volume, it is somewhat more efficient than numerical GB.

6. Concluding Discussion

Our aim was to develop a simple but physically plausible model for the electrostatics of macromolecules in solution. The replacement of the solute/solvent boundary by a sphere yields analytical formulas which, for an appropriate choice of the sphere radius, produce good results in the bulk of the molecule. For charges located near the surface of the protein, information on the local surface structure is inevitable for an accurate

evaluation of solvation and interaction energies. We chose to consider a shell around the atom and used the volume fraction occupied by solvent as the parameter which measures solvent exposure. Simple analytical considerations then lead to a formula for the sphere radius as a function of this parameter.

Our method yields a reasonably accurate description of electrostatic effects at a fraction of the computational cost of a numerical solution of Poisson's equation. In its current implementation, the method is somewhat simpler and faster than a numerical GB calculation but yields solvation and pair interaction energies of comparable accuracy. The strengths of our model are the good results in the high energy region and the absence of systematic errors, both indications that the essential physics is correctly captured.

For macromolecules composed of several roughly spherical clusters, such as 1a2p, a more successful strategy could be to replace the individual clusters by spheres of appropriate radius. Future developments will include the use of a simple analytical expression for the solvent exposure parameter. The real test for any energy function is its application in molecular dynamics simulations and its ability to discriminate between correctly folded and misfolded structures, and we plan to investigate this in the future.

Acknowledgment. We thank Dr. Andrea Cavalli for interesting discussions, and Dr. Nicolas Majeux and Urs Haberthür for performing the GB and fdP calculations. This work was supported by the Swiss National Competence Center in Structural Biology and the Swiss National Science Foundation (grant no. 31-64968.01 to A.C.).

Appendix A

The analytical solution of the Poisson−Boltzmann equation for the dielectric sphere was obtained by Kirkwood.[25] Here we derive the solution for the Poisson equation, which is of interest for the present approach. We want to calculate the potential at Cartesian position $\vec{x}$ due to a unit charge located at a point $\vec{x}'$ inside a spherical cavity centered on the origin, of radius R and dielectric constant ϵ_m, which is surrounded by water (ϵ_w), that is, solve the Poisson equation

$$\nabla_{\vec{x}}^2\phi(\vec{x},\vec{x}') = -4\pi\delta(\vec{x}-\vec{x}')\begin{bmatrix}\frac{1}{\epsilon_m} & |\vec{x}| \le R\\ \frac{1}{\epsilon_w} & |\vec{x}| > R\end{bmatrix} \tag{19}$$

Changing to spherical coordinates (r,θ,ϕ) compatible with the geometry of the dielectric boundary and using the symmetry of revolution around the Ox' axis, we expand the potential and delta function in Legendre polynomials[28]

$$\phi(\vec{x},\vec{x}') = \sum_{l=0}^{\infty} g_l(r,r')P_l(\cos\theta) \tag{20}$$

$$\delta(\vec{x}-\vec{x}') = \frac{1}{r^2}\delta(r-r')\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}P_l(\cos\theta) \tag{21}$$

Substitution of (20) and (21) into (19) yields the following equation for $g_l(r,r')$

$$\frac{1}{r}\frac{d^2}{dr^2}(rg_l(r,r')) - \frac{l(l+1)}{r^2}g_l(r,r') = -\frac{2l+1}{r^2}\delta(r-r')\begin{bmatrix}\frac{1}{\epsilon_m} & r \le R\\ \frac{1}{\epsilon_w} & r > R\end{bmatrix} \tag{22}$$

which for $r \neq r'$ allows the two solutions r^l and $(1/r^{l+1})$. Because the potential for $\vec{x} \neq \vec{x}'$ must be finite and symmetric under interchange of $\vec{x}$ and $\vec{x}'$, $g_l(r,r')$ is of the form

$$g_l(r,r') = \begin{bmatrix}\frac{A_l}{\epsilon_m}r_<^l\left(B_l r_>^l + \frac{1}{r_>^{l+1}}\right) & r \le R\\ \frac{C_l}{\epsilon_w}\frac{r_<^l}{r_>^{l+1}} & r > R\end{bmatrix} \tag{23}$$

where $r_< = \min(r,r')$ and $r_> = \max(r,r')$. The delta function in (22) implies

$$\frac{d}{dr}(rg_l(r,r'))|_{r=r'_+} - \frac{d}{dr}(rg_l(r,r'))|_{r=r'_-} = -\frac{2l+1}{\epsilon_m r} \tag{24}$$

from which it follows that

$$A_l = 1 \tag{25}$$

The coefficients B_l and C_l can now be determined from the boundary conditions at $r = R$. Continuity of the parallel component of the electric field implies

$$g_l(r,r')|_{r=R_-} = g_l(r,r')|_{r=R_+} \tag{26}$$

and the continuity of the normal component of the displacement field

$$\epsilon_m\frac{\partial g_l(r,r')}{\partial r}\bigg|_{r=R_-} = \epsilon_w\frac{\partial g_l(r,r')}{\partial r}\bigg|_{r=R_+} \tag{27}$$

For the interior solution, one needs the expression of the coefficient B_l, which is

$$B_l = -\frac{1}{R^{2l+1}}\frac{1-\frac{\epsilon_m}{\epsilon_w}}{1+\frac{l}{l+1}\frac{\epsilon_m}{\epsilon_w}} \tag{28}$$

From (28), (25), (23), and (20), we finally obtain the potential

$$\phi(\vec{x},\vec{x}') = \sum_{l=0}^{\infty}\frac{r_<^l}{\epsilon_m}\left(\frac{1}{r_>^{l+1}} - \frac{r_>^l}{R^{2l+1}}\frac{1-\frac{\epsilon_m}{\epsilon_w}}{1+\frac{l}{l+1}\frac{\epsilon_m}{\epsilon_w}}\right)P_l(\cos\theta) \tag{29}$$

which, using the identity $\sum_{l=0}^{\infty}(r_<^l/r_>^{l+1})P_l(\cos\theta) = 1/(|\vec{x}-\vec{x}'|)$,

can be expressed in the form

$$\phi(\vec{x},\vec{x}') = \frac{1}{\epsilon_m|\vec{x}-\vec{x}'|} - \frac{1}{R}\left(\frac{1}{\epsilon_m}-\frac{1}{\epsilon_w}\right)\sum_{l=0}^{\infty}\frac{1}{1+\frac{l}{l+1}\frac{\epsilon_m}{\epsilon_w}}\left(\frac{rr'}{R^2}\right)^l P_l(\cos\theta) \quad (30)$$

Formulas (5) and (6) for the interaction and solvation energies immediately follow.

Appendix B

Slightly more accurate closed form analytical expressions for (5) and (6) can be derived using the approximation

$$C_l \approx 1 - \frac{l}{l+1}\frac{\epsilon_m}{\epsilon_w} \quad (31)$$

instead of (8). From the relation

$$\sum_{l=0}^{\infty} x^l P_l(y) = \frac{1}{\sqrt{1-2xy+x^2}} \quad (32)$$

one finds by integration and subsequent derivation with respect to x

$$\sum_{l=0}^{\infty}\frac{l}{l+1}x^l P_l(y) = \frac{1}{\sqrt{1-2xy+x^2}} - \frac{1}{x}\ln\left[\frac{\sqrt{1-2xy+x^2}+x-y}{1-y}\right] \quad (33)$$

Hence, the interaction energy may be approximated by

$$E_{int,i,j} = \frac{q_i q_j}{\epsilon_m|\vec{x}_i-\vec{x}_j|} - \frac{q_i q_j}{R}\left(\frac{1}{\epsilon_m}-\frac{1}{\epsilon_w}\right)\left[\frac{1-\frac{\epsilon_m}{\epsilon_w}}{\sqrt{1-2\left(\frac{r_i r_j}{R^2}\right)\cos\theta+\left(\frac{r_i r_j}{R^2}\right)^2}} + \frac{\frac{\epsilon_m}{\epsilon_w}}{\left(\frac{r_i r_j}{R^2}\right)^2}\ln\left(\frac{\sqrt{1-2\left(\frac{r_i r_j}{R^2}\right)\cos\theta+\left(\frac{r_i r_j}{R^2}\right)^2}+\left(\frac{r_i r_j}{R^2}\right)-\cos\theta}{1-\cos\theta}\right)\right] \quad (34)$$

and from a similar calculation for the solvation energy one finds

$$E_{solv,i} = -\frac{q_i^2}{2R}\left(\frac{1}{\epsilon_m}-\frac{1}{\epsilon_w}\right)\left[\frac{1-\frac{\epsilon_m}{\epsilon_w}}{1-\left(\frac{r_i}{R}\right)^2} - \frac{\frac{\epsilon_m}{\epsilon_w}}{\left(\frac{r_i}{R}\right)^2}\ln\left(1-\left(\frac{r_i}{R}\right)^2\right)\right] \quad (35)$$

For $\epsilon_w = 80$ and $\epsilon_m = 1$, the accuracy of the simpler formulas (9) and (10) is fully sufficient for our purpose. One could in fact even use the formulas for the conducting sphere ($\epsilon_w = \infty$).

Appendix C

Because many proteins are roughly spheroidal in shape, improved estimates for solvation and interaction energies could be obtained by approximating the solute−solvent boundary by a spheroid instead of a sphere. The orientation and axes ratios of the spheroid are found from a diagonalization of the molecule's mass of inertia tensor (assuming a uniform density). In spheroidal coordinates, analytical results for interaction and solvation energies may again be derived. We will give them here for prolate spheroids (see the Appendix in Hill's paper[34] for details on how to obtain the electrostatic potential).

The Cartesian coordinates x, y, and z are related to the prolate spheroidal coordinates ξ, η, and ϕ by

$$x = a\sinh(\xi)\sin(\eta)\cos(\phi) \quad (36)$$

$$y = a\sinh(\xi)\sin(\eta)\sin(\phi) \quad (37)$$

$$z = a\cosh(\xi)\cos(\eta) \quad (38)$$

where the parameter a is fixed by the semi-axes a_{large} and a_{small} of the spheroid as

$$a^2 = a_{large}^2 - a_{small}^2 \quad (39)$$

The surface of the spheroid corresponds to a value ξ_s defined by

$$\tanh(\xi_s) = \frac{a_{small}}{a_{large}} \quad (40)$$

A somewhat lengthy calculation yields the following expression for the interaction energies between two charges

$$E_{int,i,j} = \frac{q_i q_j}{\epsilon_m|\vec{x}_i-\vec{x}_j|} - \frac{q_i q_j}{a}\left(\frac{1}{\epsilon_m}-\frac{1}{\epsilon_w}\right)\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^m \left[\frac{(l-m)!}{(l+m)!}\right]^2 (2l+1)C_{lm}P_l^m(\cosh\xi_i)P_l^m(\cosh\xi_j) \times P_l^m(\cos\eta_i)P_l^m(\cos\eta_j)\cos m(\phi_i-\phi_j) \quad (41)$$

where

$$C_{lm} = \left(\frac{P_l^m(\cosh\xi_s)}{Q_l^m(\cosh\xi_s)} - \frac{\epsilon_m}{\epsilon_w}\frac{P_l^{m\prime}(\cosh\xi_s)}{Q_l^{m\prime}(\cosh\xi_s)}\right)^{-1} \quad (42)$$

The expression for the solvation energy becomes

$$E_{solv,i} = -\frac{q_i^2}{2a}\left(\frac{1}{\epsilon_m}-\frac{1}{\epsilon_w}\right)\sum_{l=0}^{\infty}\sum_{m=-l}^{m}(-1)^l\left[\frac{(l-m)!}{(l+m)!}\right]^2 \times (2l+1)C_{lm}[P_l^m(\cosh\xi_i)P_l^m(\cos\eta_i)]^2 \quad (43)$$

The P_l^m and Q_l^m denote the associated Legengre polynomials of the first and second kind, respectively. Recurrence relations for the associated Legendre polynomials would, in principle, allow

(34) Hill, T. L. *J. Chem. Phys.* **1944**, *12*, 147−156.

an efficient calculation of these sums up to a given l_{max}, but precautions must be taken to control the loss of significant figures. A numerically robust implementation of these series would provide more accurate results for the bulk electrostatics of proteins. For charges exposed to the solvent, however, a prescription analogous to (15) would still be required.

JA021093I