Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions

Marco Scarsi, Joannis Apostolakis, and Amedeo Caflisch*

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

Received: April 28, 1997; In Final Form: August 1, 1997[⊗]

The efficient evaluation of electrostatic energies of macromolecules in aqueous solutions is useful for many problems in theoretical structural biology. A continuum method based on the generalized Born (GB) approximation is implemented here. It is shown that the choice of the dielectric discontinuity surface is critical for obtaining correct electrostatic energies of molecules in solution. In addition, it is demonstrated that an electrostatic model validated on solvation energies (vacuum to water transfer) might not be appropriate for energies in solution and might not yield correct energy ranking of ligand/protein complexes. The agreement between the GB approach and the finite difference solution of the Poisson equation is shown to be very good for both the molecular and the solvent accessible surface. The discrepancies between the GB and the finite difference approach are much lower than the ones due to the use of different surfaces.

1. Introduction

Accurate electrostatic energies in aqueous solution are needed to discriminate between near-native and nonnative conformations of a protein or to find the most probable binding mode of a ligand/protein or protein/protein complex. In most force fields, the electrostatics of the solute−solvent system are described as a distribution of point charges. An exact approach to the problem should consider the interactions among all possible pairs. This is computationally very expensive, not only because of the high number of interactions involved but also because of the computational costs due to the equilibration of the water molecules. Continuum electrostatic models were introduced to overcome such difficulties.[1−4]

A continuum model usually describes the solute as a region with low dielectric constant (between 1 and 4) with a certain spatial charge distribution usually obtained from the partial charges of the solute, and the solvent as a region with high dielectric constant (∼80). Assuming these conditions, an exact solution of the problem is obtained by solving the Poisson equation, or, in the presence of diluted salts, the Poisson−Boltzmann equation. These equations can be solved on a spatial grid with a finite difference approach. Results originating from this simplified model for electrostatics seem to agree well with molecular dynamics simulations that include explicit water molecules.[5] Even if the gain in computation time is remarkable with respect to explicit water calculations, for certain applications it is still too slow.

A continuum electrostatics method inspired by the work of Still et al.[6] is presented in this paper. Different descriptions of the solute−solvent boundary are allowed in the present implementation and this distinguishes it from previous studies.[6−9] The choice of the solute−solvent boundary is shown here to be extremely critical for obtaining accurate electrostatic energies.

It is also shown that for continuum models of the solvent a validation based on the *solvation energy* (vacuum to water transfer) might not be sufficient. Such validation is insufficient whenever an electrostatic continuum model of the solvent is used to supplement a vacuum Coulomb term in a force field to obtain the electrostatic *energy in solution* (i.e., solvation plus vacuum Coulomb energy) of a (macro)molecule. In cases where the solvation energy anticorrelates with the vacuum Coulomb energy, as in many molecular complexes analyzed in this study, it is necessary to validate the model also on energies in solution to obtain accurate energy rankings. Hence, a continuum electrostatic treatment that approximates well the solvation energies does not necessarily yield accurate electrostatic energies of macromolecules in aqueous solution. The latter are more relevant and by far the most critical in simulations of biologically relevant systems.

2. Theory

The electrostatic energy E of a continuum system can be expressed in terms of the electric displacement vector $\vec{D}(\vec{x})$ and of a location-dependent dielectric constant $\epsilon(\vec{x})$ as an integral over three-dimensional space R^3:[10]

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

Since $\vec{D}(\vec{x})$ is additive, for point charges it can be rewritten as a sum over all charges i in the solute molecules:

$$\vec{D}(\vec{x}) = \sum_i \vec{D}_i(\vec{x}) \tag{2}$$

In this way eq 1 can be split up into the self-energy and interaction energy terms:[11,12]

$$E = E^{\mathrm{self}} + E^{\mathrm{int}} = \sum_i E_i^{\mathrm{self}} + \sum_{i>j} E_{ij}^{\mathrm{int}} \tag{3}$$

$$E_i^{\mathrm{self}} = \frac{1}{8\pi}\int_{R^3} \frac{\vec{D}_i^{\,2}(\vec{x})}{\epsilon(\vec{x})}\, \mathrm{d}^3x \tag{4}$$

$$E_{ij}^{\mathrm{int}} = \frac{1}{4\pi}\int_{R^3} \frac{\vec{D}_i(\vec{x})\cdot\vec{D}_j(\vec{x})}{\epsilon(\vec{x})}\, \mathrm{d}^3x \tag{5}$$

The self-term originates from the energy stored in the electric field due to the charge itself, while the interaction term accounts for the interactions of all pairs of charges. Equation 4 gives the Born formula for the solvation of a spherical ion in water by assuming uniform dielectric constants for vacuum and water. If $\epsilon(\vec{x})$ is supposed to be constant over R^3, eq 5 yields the simple Coulomb law between charges i and j. In the case of molecules

* Corresponding author. Email: caflisch@bioc.unizh.ch.

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1997.

S1089-5639(97)01422-9 CCC: $14.00

in solution $\epsilon(\vec{x})$ can be assumed to be equal to ϵ_w in the solvent region and ϵ_p in the solute region.[13−15] Within this assumption eq 3 has been solved by a semianalytical approach[6] and by a completely analytical approach.[12]

In their semianalytical approach Still et al. neglected the reaction field (see below) and calculated the self-energies of every charge numerically.[6] For the interaction term they derived an analytical expression of the GB formula[16−18] (in the following this analytical expression will be referred to as the GB formula). It yielded the interaction energy of two charges as a function of their distance and the product of their self-energies.

In the analytical approach Schaefer and Karplus also neglected the reaction field in the calculation of the self-energies.[12] By introducing Gaussian functions to describe the atomic volumes, they obtained a formula for the self-energy of each atom as a function of the coordinates of all the other ones. It is an integral that does not have an analytical solution, but the authors provided an analytical approximation to its numerical solution (eq 22 in ref 12). For the interaction term they used the GB formula.

The main difference between the two methods is that the electrostatic energy in the analytical approach is fully differentiable and yields electrostatic forces that are consistent with the energy function, while the semianalytical approach assumes that the variations of the total electrostatic energy due to variations of the self-energies are small and can be discarded in the derivatives. However, the solute description in the semianalytical approach is more precise and different choices of the low dielectric volume are possible, while the volume description through the introduction of Gaussians does not allow a detailed representation of the solute−solvent boundary. In this paper it is shown that the choice of the solute−solvent boundary is critical for the correct evaluation of the electrostatic energy of molecules in solution.

3. Method

In the present implementation, eq 3 is treated in a semianalytical way within the Coulomb field approximation and the GB approach. In subsection 3.1 the Coulomb field approximation is used for the self-energy term (eq 4) and a numerical evaluation of the self-energy is presented. The GB formula is introduced in subsection 3.2 to evaluate the interaction energy term (eq 5).

3.1. Self-Energy. As mentioned in section 2, $\epsilon(\vec{x})$ can be assumed to be equal to ϵ_w in the solvent region and ϵ_p in the solute region. In this way eq 4 for the self-energy of a charge i can be rewritten as[12]

$$E_i^{\text{self}} = \frac{1}{8\pi\epsilon_w}\int_{\text{solvent}} \vec{D}_i^{\,2}(\vec{x})\,\mathrm{d}^3x + \frac{1}{8\pi\epsilon_p}\int_{\text{solute}} \vec{D}_i^{\,2}(\vec{x})\,\mathrm{d}^3x$$
$$= \frac{1}{8\pi\epsilon_w}\int_{R^3} \vec{D}_i^{\,2}(\vec{x})\,\mathrm{d}^3x + \frac{\tau}{8\pi}\int_{\text{solute}} \vec{D}_i^{\,2}(\vec{x})\,\mathrm{d}^3x \qquad (6)$$

where $\tau = (1/\epsilon_p) - (1/\epsilon_w)$.

In the presence of a spatial discontinuity in the dielectric constant, each charge i induces a polarization charge distribution at the discontinuity surface between solute and solvent. The field generated by this charge density, called reaction field and denoted in the following as $\vec{R}_i(\vec{x})$, should be added to the Coulomb field of the charge i itself to obtain the actual dielectric displacement:

$$\vec{D}_i(\vec{x}) = \frac{q_i}{(\vec{x} - \vec{x}_i)^2}\frac{(\vec{x} - \vec{x}_i)}{|\vec{x} - \vec{x}_i|} + \vec{R}_i(\vec{x}) \qquad (7)$$

In the self-energy term (eq 6) each $\vec{D}_i(\vec{x})$ will be approximated by the Coulomb field; i.e., the reaction field will be neglected (Coulomb field approximation). This approximation relies mainly on the short-range character of the Coulomb field energy density, that is proportional to $1/r^4$. The applicability of this approximation to the first and second integral in eq 6 has been addressed in refs 11 and 12, respectively. Neglecting the reaction field is more critical for large molecules, but for these the interaction term is dominant. In the case of a low dielectric molecule in a high dielectric medium, the Coulomb field approximation leads to an overestimation of the self-energies.[12]

In the Coulomb field approximation, eq 6 can be rewritten as

$$E_i^{\text{self}} = \frac{q_i^{\,2}}{2R_i^{\text{vdW}}\epsilon_w} + \frac{\tau q_i^{\,2}}{8\pi}\int_{\text{solute}\backslash\text{atom}_i} \frac{1}{(\vec{x}_i - \vec{x})^4}\,\mathrm{d}^3x \qquad (8)$$

where for each atomic charge i in the solute, a spherically symmetric distribution has been assumed inside its own van der Waals sphere. The integration domain in eq 8 is all of the solute volume apart from the volume of atom i. Contributions originating from the integration inside the sphere i itself can be discarded, since each individual charge distribution is assumed not to change during the system evolution.

In this way the evaluation of the electrostatic self-energy of molecules in solution in the Coulomb field approximation is reduced to the evaluation of the integral in eq 8 over the solute volume. This is done on a three-dimensional grid, as explained in the next sections.

3.1.1. Solute Volume. The definition of the solute volume is not unique. The most common definitions describe the molecular volume as the volume enclosed by the van der Waals surface (vdWS), the solvent-accessible surface (SAS), or the molecular surface (MS).

The vdWS is simply the external surface resulting from the convolution of all the van der Waals spheres of each atom. The SAS is obtained in the same way by previously increasing the van der Waals radii by an amount that can be interpreted as the radius of a hypothetical water sphere rolling over the van der Waals surface. The surface spanned by the *center* of such a rolling sphere describes the solvent-accessible surface.[19] The MS is spanned by the *surface* of this rolling sphere in contact with the vdWS.[20]

The first two volumes can be described by scanning every atom sphere (with radius R^{vdW} or $R^{\text{vdW}} + R^{\text{probe}}$, respectively) over a cubic grid and looking at which cube centers fall into such spheres. The procedure adopted for the MS consists of three steps. The first one is the same as for the SAS; in the second one a spherical grid is generated on the surface of each sphere (of radius $R^{\text{vdW}} + R^{\text{probe}}$) and all of the spherical grid points that do not fall inside any of these spheres are labeled (they describe the SAS). Finally, a sphere of radius R^{probe} is placed over each of the labeled points and all the cubic grid points falling inside the probe sphere are assigned to the solvent region. This procedure is also used in the program UHBD (University of Houston Brownian Dynamics).[21−23]

The volume enclosed by the vdWS appears to be unphysical for electrostatic calculations, since many small cavities between the atoms in the interior of a molecule would be assigned to the high dielectric region, even if they are not accessible to water molecules. It is still a matter of discussion whether in electrostatic continuum models the most appropriate volume to describe the solute region is the one enclosed by the MS or by the SAS. The present implementation allows the user to choose between any of the aforementioned surfaces, while in previous

implementations of the GB approach[6−9] it was not possible to define the dielectric discontinuity surface by the MS.

3.1.2. Numerical Evaluation of the Self-Energy. For the numerical evaluation of the integral in eq 8, a three-dimensional grid is used. Points inside the solute volume are labeled and a map of the solute grid points is built. Then the contribution of each of them is assigned to the self-energy of every charge:

$$\int_{\text{solute}\backslash\text{atom}_i} \frac{1}{(\vec{x}_i - \vec{x})^4} \mathrm{d}^3x \simeq \sum_{k\in K_i} \frac{1}{(\vec{x}_i - \vec{x}_k)^4} \Delta V_k \qquad (9)$$

where K_i contains all of the grid points occupied by all atoms but atom i (Figure 1). This definition of K_i will apply in all the following equations. The grid points belonging to K_i and falling at a distance from $\vec{x}_i$ larger than a fixed cutoff (a reasonable value is around 5 Å) are grouped in cubes containing 27 units each. This creates a second coarse grid beyond the given cutoff. Hence, in this region the grid spacing is 3 times larger than in the region inside the cutoff. This leads to a speed up of the calculations without loss of accuracy, since the main contribution to the integral are evaluated on the fine grid (without any grouping). In this way the self-energy of every charge i on a grid is

$$E_i^{\text{self}} = \frac{{q_i}^2}{2R_i^{\text{vdW}}\epsilon_{\text{w}}} + \frac{\tau {q_i}^2}{8\pi} \sum_{k\in K_i} \frac{1}{(\vec{x}_i - \vec{x}_k)^4} \Delta V_k \qquad (10)$$

It follows that, for the transfer from a medium with dielectric constant ϵ_{p} to a medium with dielectric constant ϵ_w, the solvation energy of the charge i is

$$\Delta E_i = E_i^{\text{self}} - \frac{{q_i}^2}{2R_i^{\text{vdW}}\epsilon_{\text{p}}}$$

$$= \frac{\tau {q_i}^2}{8\pi} \sum_{k\in K_i} \frac{1}{(\vec{x}_i - \vec{x}_k)^4} \Delta V_k - \frac{\tau {q_i}^2}{2R_i^{\text{vdW}}} \qquad (11)$$

It is possible to define a quantity directly correlated to the integral in eq 9, the effective Born radius, R_i^{eff}:

$$\Delta E_i \equiv -\frac{{q_i}^2\tau}{2R_i^{\text{eff}}} \Rightarrow R_i^{\text{eff}} \equiv \frac{-{q_i}^2\tau}{2\Delta E_i} \qquad (12)$$

The Born formula for the solvation energy of a spherical ion (with charge q and radius R) in water is

$$\Delta E = -q^2\tau/(2R) \qquad (13)$$

Equations 12 and 13 show that the definition of the effective Born radius originates from the Born formula. By substituting eq 11 into eq 12 it is possible to rewrite R_i^{eff} as

$$R_i^{\text{eff}} \equiv \left(\frac{1}{R_i^{\text{vdW}}} - \frac{1}{4\pi} \sum_{k\in K_i} \frac{1}{(\vec{x}_i - \vec{x}_k)^4} \Delta V_k \right)^{-1} \qquad (14)$$

It is important to note that the effective Born radius of a charge does not depend on the dielectric constants or the charges of the system, but is just a geometrical parameter. R_i^{eff} will be used to calculate the interaction energy as explained in the next section.

3.2. Interaction Energy. The interaction term includes N^2 contributions and its numerical evaluation on a grid would be computationally too expensive, even in the Coulomb field approximation. A much more efficient way to evaluate the interaction term is provided by the GB formula.[6] It implies that the interaction energy between two charges in a molecule changes upon transfer from a medium of dielectric constant ϵ_{p} to a medium of dielectric constant ϵ_{w} according to:

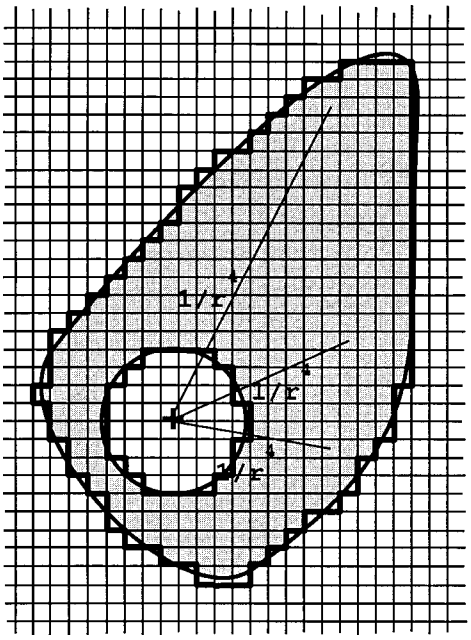

Figure 1. Schematic representation in 2D of the numerical evaluation of the integral in eq 9. The shaded region represents the molecule volume that contributes to the determination of the self-energy of an atom described by the white spherical region. Cubic grid elements are assigned to the integration domain if their center falls inside the solute volume and outside the sphere of the charge whose self-energy is evaluated.

$$\Delta E_{ij}^{\text{int}} = -q_iq_j\tau/R_{ij}^{\text{GB}} \qquad (15)$$

where

$$R_{ij}^{\text{GB}} = \left[{r_{ij}}^2 + R_i^{\text{eff}}R_j^{\text{eff}} \exp\left(\frac{-{r_{ij}}^2}{4R_i^{\text{eff}}R_j^{\text{eff}}} \right) \right]^{1/2} \qquad (16)$$

For Monte Carlo or molecular dynamics simulations in implicit water the electrostatic energy rather than the solvation energy is needed. Hence, it is necessary to add the interaction energy in the medium of dielectric constant ϵ_{p} to eq 15:

$$E_{ij}^{\text{int}} = \frac{q_iq_j}{\epsilon_{\text{p}}r_{ij}} - \frac{q_iq_j\tau}{R_{ij}^{\text{GB}}} \qquad (17)$$

In the case ${r_{ij}}^2 \gg R_i^{\text{eff}}R_j^{\text{eff}}$, eq 17 yields the Coulombic interaction energy in a medium with the dielectric constant of the solvent. In the opposite case $R_i^{\text{eff}}R_j^{\text{eff}} \gg {r_{ij}}^2$, it yields the Coulombic interaction energy in a medium with the dielectric constant of the solute. Equation 15 can be extended to the case $i = j$, where one obtains the solvation energy of the charge i (eq 12). In the intermediate cases, eq 15 gives an estimation of the screening on the $i \sim j$ interaction by the solvent.

3.3. Total Electrostatic Energy. From eqs 3, 10, and 17 it is possible to express the electrostatic energy of a solute of dielectric constant ϵ_{p} in a solvent with dielectric constant ϵ_{w} as:

$$E = \sum_i \left(\frac{q_i^2}{2R_i^{\mathrm{vdW}}\epsilon_w} + \frac{\tau q_i^2}{8\pi} \sum_{k \in K_i} \frac{1}{(\vec{x}_i - \vec{x}_k)^4} \Delta V_k \right) + \sum_{i>j} \left(\frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{\mathrm{GB}}} \right) \quad (18)$$

When the energy of different conformations of the same molecular system is evaluated the term $\sum_i q_i^2/(2R_i^{\mathrm{vdW}}\epsilon_w)$ is constant and can be neglected. The solvation energy for the transfer from a medium with dielectric constant ϵ_p to a medium with dielectric constant ϵ_w can be derived from eqs 3, 11, and 15:

$$\Delta E = \sum_i \frac{\tau q_i^2}{2} \left(\frac{1}{4\pi} \sum_{k \in K_i} \frac{1}{(\vec{x}_i - \vec{x}_k)^4} \Delta V_k - \frac{1}{R_i^{\mathrm{vdW}}} \right) - \sum_{i>j} \frac{q_i q_j \tau}{R_{ij}^{\mathrm{GB}}} \quad (19)$$

Using eq 12 one can rewrite eq 19 in a compact form

$$\Delta E = -\sum_{i,j} \frac{q_i q_j \tau}{2R_{ij}^{\mathrm{GB}}} \quad (20)$$

Equations 18 and 19 have been introduced in the CHARMM force field and provide the electrostatic energy in solution and the solvation energy, respectively.

4. Test Systems

The approximations which have led to the expression for the electrostatic energy in solution (eq 18) and for the solvation energy (eq 19) generate both a systematic and a statistical error. The former originates from the Coulomb field approximation introduced to calculate the self-energy and from the GB formula used for the interaction energy. The statistical error due to the discrete grid in the self-energy evaluation can be asymptotically eliminated by reducing the grid spacing.

To check the systematic error, the electrostatic energies in solution and the solvation energies of different chemical compounds evaluated by eqs 18 and 19 are compared with the corresponding values calculated by the finite difference solution of the Poisson equation, provided by the program UHBD. The distinction between solvation energy and energy in solution is important and in the following they will be analyzed separately.

The test systems consist of a set of 40 small organic compounds, and a set of structures of proteins or ligand/protein complexes generated independently of this work. They include a set of complexes between small ligands and one monomer of HIV-1 aspartic proteinase (ligand/HIV PR; Caflisch, A., unpublished results), a set of complexes between small ligands and thrombin (ligand/thrombin; ref 24), a series of snapshots taken every 5 ps of a molecular dynamics (MD) trajectory of barnase at 300 K (barnaseT300; ref 25), and a series of snapshots of the same protein taken every 10 ps of an MD denaturation simulation at 360 K in an 8 M urea solution (barnaseT360; Caflisch, A., unpublished results). For the small organic compounds the comparison between the present method and the finite difference calculation is done for the solvation energies, while for all other test cases it involves also the energies in solution.

Charges and van der Waals radii for the ligand/HIV PR and ligand/thrombin complexes, and for the barnaseT300 trajectory are assigned according to the polar hydrogen parametrization of CHARMM (param19).[26] The all-hydrogen parametrization of CHARMM (param22),[27] was employed for charges and radii of the small organic compounds and of the barnaseT360 trajectory. As already mentioned, these structures were generated independently of this work, and this is the reason for the different parametrizations. In all the calculations the solvent dielectric constant is assumed to be 78.5 and the one of the solute 1. The value for the solute is consistent with the partial charges of the two aforementioned parameter sets that were derived from *ab initio* calculations with a dielectric constant of 1. The radius of the probe water molecule used to generate the MS or the SAS is 1.4 Å. The grid spacing of the final focusing in the finite difference calculation was 0.25 Å for the small organic compounds and 0.5 Å for the other test cases. The grid spacing used for the evaluation of the self-energies in the present method was 0.06 Å for the small organic compounds and 0.35 Å for the other test cases. The difference in the CPU time required by the two electrostatic methods for every energy evaluation of the macromolecular systems ranged from a factor of 31 for the ligand/HIV PR complexes (800 s for UHBD, and 26 s for the present method) to a factor of 14 for the barnaseT360 trajectory (1000 and 71 s for the two calculations, respectively). The calculations were performed on SGI workstations (processor R4400, clock frequency 200 MHz). As shown in subsection 5.5 the grid spacing in the self-energy calculation can be increased to values of 0.5−0.6 Å without significant loss in accuracy but with a major gain in speed.

In the next sections, the continuum electrostatic model which led to eqs 18 and 19 will be called MS/GB, SAS/GB, or vdWS/GB, according to the surface used to define the solute boundary. Following the same convention, the finite difference solution of the Poisson equation will be referred to as MS/FDP, SAS/FDP, or vdWS/FDP.

5. Results

The solvation energy values of the small organic compounds are discussed in subsection 5.1, while the results for the ligand/HIV PR complexes, the ligand/thrombin complexes, and the barnase trajectories are presented in subsection 5.2, 5.3, and 5.4, respectively. Unless otherwise specified, the *x*-axis in the plots refers to the present method and the *y*-axis to the finite difference calculations. The error due to the grid approximation in the self-energy evaluation is estimated in subsection 5.5 for a conformation of barnase from the barnaseT360 trajectory.

5.1. Solvation Energies of Small Organic Compounds. This is a typical test case for electrostatic continuum models.[6,12,28−30] The main reason is that experimental values of solvation free energies are available.[31−33] In addition, their structure does not undergo large rearrangements during solvation. The solvation free energy of a molecule can be divided into an electrostatic contribution and an apolar contribution that is due to the free energy needed for the cavity formation and to the solute−solvent dispersion energy. In the present calculations only the electrostatic term is evaluated. The apolar term is positive and of the order of 1−2 kcal/mol for the small molecules analyzed in this work. It accounts for almost the entire solvation energy of the apolar compounds.

Forty small organic molecules are analyzed. They include 10 charged (acetate, ethanolate, methanolate, phenolate, prolineamide, guanidinium, imidazolium, methylammonium, ethylthiolate, methylthiolate), 15 polar (acetic acid, acetamide, formamide, *N*-acetylproline-NH_2, methylamine, imidazole, indole, ethanol, methanol, phenol, 2-propanol, (*Z*)-*N*-methylacetamide, alanine dipeptide in the $C7_{eq}$, C5, and α_R conformations), and 15 apolar compounds (two conformations of 1-butene, two conformations of 2-butene, ethene, propane,

TABLE 1: Comparison of Experimental Solvation Energies with Finite Difference Calculations

molecule	ΔG^{solv} (kcal/mol) MS/FDP[a]	SAS/FDP[a]	experiments[b]
ethane	−0.20	−0.082	1.8
n-butane (anti)	−0.26	−0.12	2.3
ethene	−1.8	−0.14	1.3
propane	−1.5	−0.15	1.3
1-butene	−1.5	−0.13	1.4
methanethiol	−1.9	−0.30	−1.2
benzene	−2.0	−0.28	−0.9
ethanol	−7.0	−0.99	−5.0
methanol	−7.3	−1.1	−5.1
2-propanol	−6.4	−0.89	−4.8
phenol	−7.3	−0.97	−6.6
acetic acid	−8.0	−0.82	−6.7
acetate ion	−76	−44	−75
methylammonium ion	−80	−46	−68

[a] Electrostatic contribution to the solvation energy. [b] All experimental data were taken from Cabani et al.[31] except acetate ion and methylammonium ion which were taken from Pearson.[33]

benzene, butane, isobutane, two conformations of ethane, ethyl sulfide, methyl ethyl sulfide, dimethyl disulfide, methanethiol). In Table 1 the electrostatic contributions to the solvation energies calculated by the MS/FDP and SAS/FDP methods are compared with the experimental solvation energies for 14 of the 40 small molecules for which experimental data are available. The apolar contributions are not evaluated but they are always positive. In the MS/FDP calculations, the electrostatic contribution to the solvation energy is in general more negative or close to the experimental solvation energy, while in the SAS/FDP calculations the electrostatic solvation energy is always larger than the experimental solvation energy. The larger volume enclosed by the SAS reduces the electrostatic effects significantly and does not seem to be appropriate to describe the electrostatics of molecules in solution, at least for the CHARMM parametrizations.

Figure 2a shows the results for the electrostatic contribution to the solvation energy, calculated by the MS/FDP and MS/GB methods for the 40 small organic compounds. The agreement is very good. The slope of the fitting line is 1.06 and the correlation coefficient (R) is 0.998 with all the compounds. Including only the nonionic compounds, one obtains a slope of 1.47 and an R of 0.986. The MS/GB solvation energies are shifted on average by 2.7 kcal/mol with respect to the ones from the MS/FDP calculations. This effect is due mainly to the Coulomb field approximation. It is possible to reduce this shifting to around 0.8 kcal/mol by using in eq 19 a solute−solvent boundary defined by a surface similar to the vdWS (it is defined as the vdWS, with vdW radii decreased by 0.1 Å; in the following it will be referred as vdWS-01). This choice is similar to the one of Still et al. (they reduced the vdW radii by 0.09 Å to reproduce solvation energies calculated by free energy perturbations)[6] and even if it can appear unphysical for the reason given in subsection 3.1.1 it yields solvation energy values close to the MS/FDP ones as shown in Figure 2b. The correlation coefficients are 0.999 and 0.992, and the slopes of the fitting lines 0.99 and 1.15, with and without the ions, respectively. In the following it will be shown that in general this agreement does not hold for the energies in solution.

5.2. Ligand/HIV PR Complexes. The second test case consists of a set of 1050 conformations of small molecules docked in the dimerization interface of the HIV-1 aspartic proteinase monomer by minimization of the CHARMM energy with a distance-dependent dielectric. Minimized positions of the ligands on the surface of the rigid proteinase monomer had

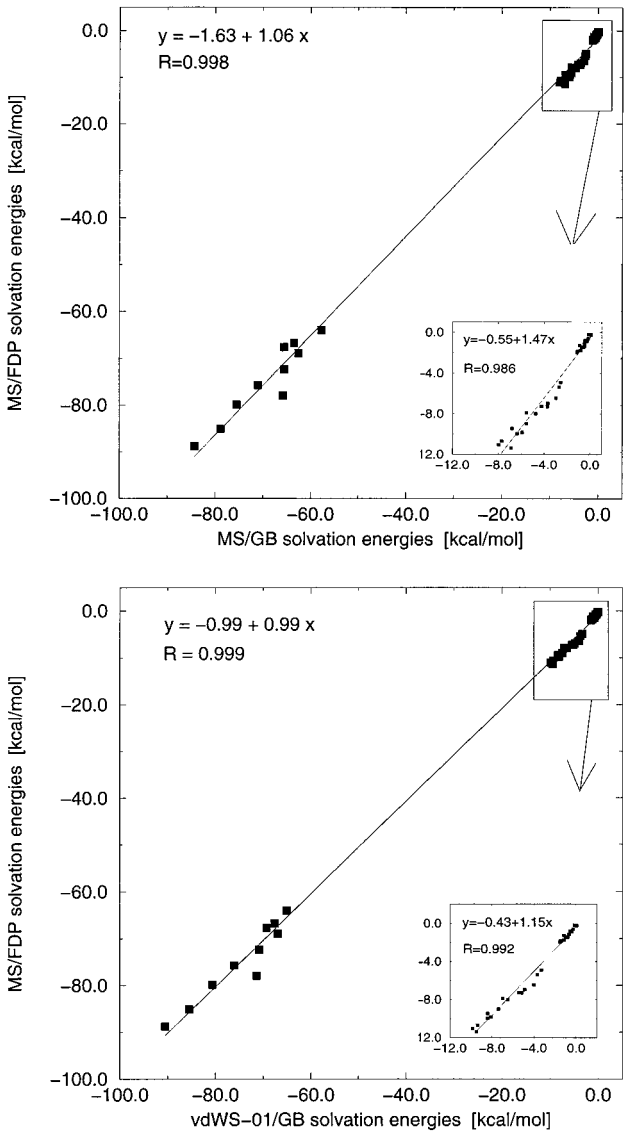

Figure 2. (a, top) Electrostatic contribution to the solvation energy of 40 small molecules calculated by the MS/GB and the MS/FDP methods. The fit has been done with the ionic compounds (full plot), and without the ionic compounds (inset). (b, bottom) Same as in (a), for the vdWS-01/GB and the MS/FDP methods.

been generated with the program MCSS (Multiple Copy Simultaneous Search).[34] For all of them the solvation energies and the energies in solution calculated by UHBD and by eqs 18 and 19 are compared. Different surfaces are used to define the solute volume. The ligands are benzene, cyclohexane, *N*-methylacetamide, benzamidine, imidazole, diketopiperazine, acetate ion, and methylammonium ion.

5.2.1. Choice of the Solute−Solvent Boundary Surface. The choice of the solute−solvent boundary surface strongly affects the electrostatic energies in solution. Figure 3a compares the energies in solution of the diketopiperazine/HIV PR complexes calculated by the MS/FDP and SAS/FDP methods, respectively. The correlation coefficient between the two sets of data is very low (0.269). Low correlations can be observed also in the *N*-methylacetamide/HIV PR ($R = 0.463$) and imidazole/HIV PR ($R = 0.398$) complexes. This shows that different descriptions of the solute−solvent boundary yield completely different ranking of the energies in solution for this kind of ligand/protein complexes.

5.2.2. Solvation Energies and Energies in Solution. The solvation energies of the diketopiperazine/HIV PR complexes calculated by MS/FDP and SAS/FDP show a correlation of

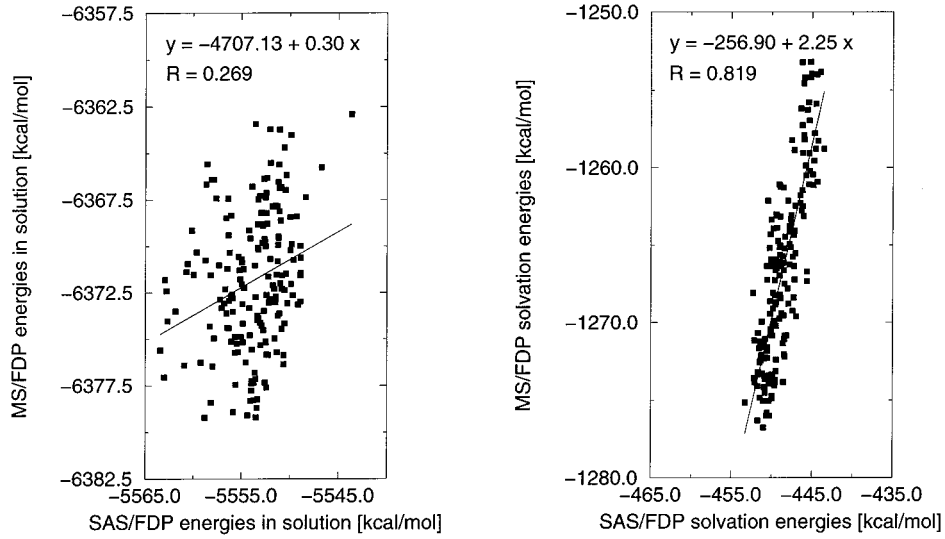

Figure 3. Electrostatic energies of the diketopiperazine/HIV PR complexes evaluated by the SAS/FDP (*x*-axis) and the MS/FDP (*y*-axis) methods. Energies in solution (a, left) and solvation energies (b, right) show a different behavior.

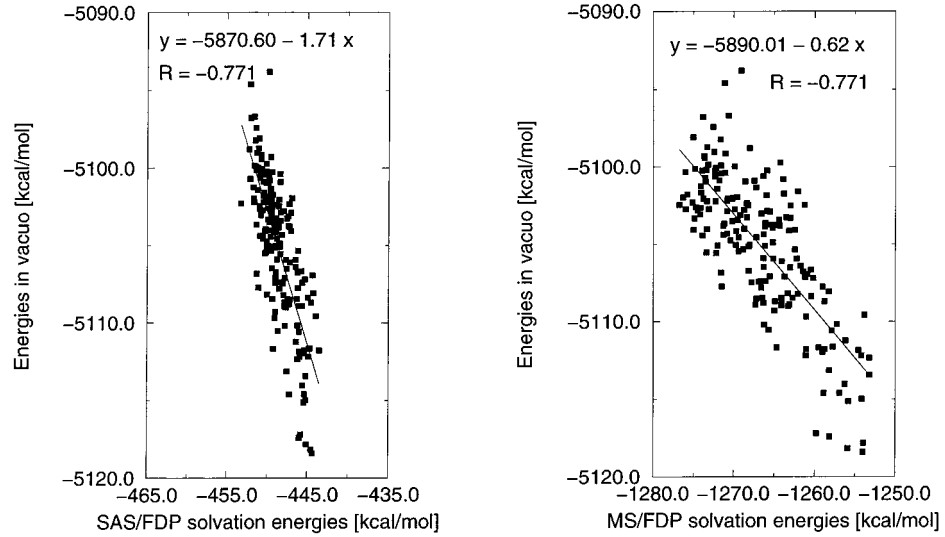

Figure 4. Electrostatic solvation energies calculated by a finite difference approach vs electrostatic energies in vacuo (Coulomb law) for the same complexes as in Figure 3. In (a, left) the SAS is assumed as the dielectric discontinuity surface in the solvation energy calculations, while in (b, right) the MS is used.

0.819 (Figure 3b). Also the solvation energies of the *N*-methylacetamide/HIV and imidazole/HIV PR complexes show good correlations ($R = 0.874$ and $R = 0.854$, respectively). Yet, as shown in the previous subsection, the energies in solution calculated by MS/FDP and SAS/FDP show a poor correlation. The reason for this is an anticorrelation between the solvation energies (calculated with any surface) and the energies in vacuo (correlation coefficients of -0.771 with both the SAS and the MS), as it can be seen in Figure 4, a and b. Interestingly, a multiplicative factor times the vacuum Coulomb energy could well reproduce the solvation energies of these systems. This means that the correlation between the MS/FDP and SAS/FDP solvation energy values is determined mainly by the contribution of the energy in vacuo and only marginally by the energy in solution. Consequently, in general it will not be sufficient to check the solvation energy to validate an electrostatic model.

5.2.3. Comparison with the Finite Difference Calculations. Electrostatic energies calculated by the vdWS/GB method are not in agreement with the corresponding values obtained by the vdWS/FDP calculations. When the solute−solvent boundary is defined by the vdWS charges are much more exposed to the solvent than in the case of the MS or the SAS and the Coulomb approximation is not valid any more.

By contrast, both solvation energies and energies in solution calculated by the MS/GB method correlate well with the results from the MS/FDP method. Results of the fits for these quantities are shown in Table 2. The energies calculated by the two methods are plotted in Figure 5 for the methylammonium ion/HIV PR complex. The slopes of the fitting lines for energies in solution of different complexes are between 0.51 (*N*-methylacetamide) and 0.83 (diketopiperazine) and the correlation coefficients vary between 0.701 (*N*-methylacetamide) and 0.966 (methylammonium ion). Hence, the ranking of the energies in solution is preserved. The slope can eventually be corrected analytically by means of an appropriate scaling factor. The solvation energies show a better correlation than the energies in solution. Both solvation energies and energies in solution calculated by the present method are significantly shifted in magnitude with respect to the finite difference calculations. This is an effect of the Coulomb field approximation that overestimates the self-energies. If the MS/FDP results are compared with the ones from the vdWS-01/GB method, solvation energies are not shifted any more and show satisfactory

TABLE 2: Comparison of Electrostatic Energies Calculated by the MS/GB and the MS/FDP Methods

		energy in solution			solvation energy	
	N^a	R^b	α^c	Δ^d	R^b	α^c
Ligand/HIV PR Complexes						
benzene	58	0.812	0.80	442	–[e]	–
cyclohexane	117	0.812	0.80	445	–	–
N-methylacetamide	192	0.701	0.51	446	0.901	1.04
imidazole	178	0.718	0.61	445	0.877	1.00
diketopiperazine	183	0.832	0.83	445	0.929	1.02
acetate ion	88	0.942	0.61	452	0.969	1.15
benzamidine	163	0.937	0.78	454	0.956	1.13
methylammonium ion	71	0.966	0.82	458	0.966	1.10
Ligand/Thrombin Complexes						
benzene	32	0.920	1.15	1269	–	–
propane	84	0.871	1.32	1270	–	–
methanol	78	0.572	0.56	1272	0.862	1.51
imidazole	81	0.556	0.92	1268	0.856	1.22
methylammonium ion	52	0.728	0.90	1234	0.936	1.28
pyrrolidine	68	0.803	0.82	1229	0.919	1.30
Barnase MD Trajectories						
barnaseT300	63	0.814	0.54	591	0.995	1.21
barnaseT360	126	0.931	0.68	498	0.947	1.27

[a] *N*: number of conformations analyzed. [b] *R*: correlation coefficient for the fit of the MS/GB to the MS/FDP energies. [c] α: slope of the fitting line of the MS/GB to the MS/FDP energies. [d] Δ: difference (in kcal/mol) between average energy from the MS/GB and from the MS/FDP calculations. [e] Apolar compounds such as benzene, cyclohexane, and propane have no partial charges in the polar hydrogen parametrization of CHARMM param19. For all the ligand/protein complexes involving these molecules the parameters for the fit of the solvation energies are exactly the same as for the fit of the energies in solution.

correlations (in the example of the ligand *N*-methylacetamide, *R* has a value of 0.546). Also the energies in solution are not shifted any more, but for many ligand/protein complexes they do not correlate with the finite difference calculations (for *N*-methylacetamide, *R* is 0.151). This means that, although the choice of a smaller solute volume in eq 18 yields accurate solvation energies for the small organic compounds (see Figure 2b) and for many ligand/protein complexes, in the general case of large molecules it is not suited for reproducing the MS/FDP energies in solution.

When compared with the finite difference calculations, the MS/GB energies in solution are shifted up and their scaling is slightly enhanced with respect to MS/FDP. The enhancement of the scaling is reflected in the value of the slope, which is lower than 1. As already mentioned the shifting is of no importance, while the slope can be corrected analytically, since it is similar for the different complexes.

The results from the SAS/FDP method have been reproduced by the SAS/GB calculations with a very good correlation. The mean value of *R* for the energies in solution of all the ligand/HIV PR complexes is 0.88 ± 0.13 and the one for the solvation energies is 0.83 ± 0.10. The reason for this good agreement is that, in the case of the SAS, the induced polarization charge density is lower than in the case of smaller solute volumes, like the ones enclosed by the MS or the vdWS.

5.3. Ligand/Thrombin Complexes. The calculations of the last subsection were performed also on a set of 395 ligand/thrombin complexes involving the ligands benzene, propane, methanol, pyrrolidine, imidazole, and methylammonium ion. The minimized positions of the ligands on the surface of the rigid thrombin molecule had been generated by MCSS with a distance dependent dielectric and have been described in detail.[24]

As shown in Table 2, the MS/FDP and MS/GB calculations show a reasonable correlation with correlation coefficients varying from 0.556 (imidazole) to 0.920 (benzene). The solvation energies and the energies in solution calculated by the two methods are plotted in Figure 6 for the methylammonium ion/thrombin complexes. The slopes of the fits vary more than in the ligand/HIV PR complexes, but for thrombin the number of minima of each ligand type is in general smaller and this leads to poorer statistics. The shift in the energy is enhanced and seems to grow proportionally with the number of atoms of the complex (922 atoms in the HIV PR monomer and 2559 in thrombin).

For the ligand/HIV PR and ligand/thrombin complexes the energy ranking obtained by MS/GB is substantially correct (Table 2) and can be used for structure-based ligand design.[24] If more accurate energies are needed, for the ligand/HIV PR complexes one could scale the energies calculated by the MS/GB approach by a factor corresponding to the average value of the slope of the fitting lines. This might be accurate enough because the slopes have similar values (ranging from 0.51 to 0.83) for different ligand types. For the ligand/thrombin complexes the slopes of the fitting lines show a larger variability (from 0.56 to 1.32) and it might be more difficult to reparametrize the model with a simple scaling factor. The larger variability in the slope might be a consequence of the smaller number of conformations as mentioned above.

5.4. MD Trajectories of Barnase. In the test cases presented above, the structures of the different complexes were the result of energy minimizations. However, the actual conformation of a molecule in solution is determined by the free energy. For this reason, snapshots from MD trajectories of barnase with explicit solvent molecules are considered (barnaseT300 and barnaseT360). BarnaseT300 is a control trajectory of 310 ps at 300 K, pH 7 with explicit water, saved every 5 ps.[25] BarnaseT360 is a simulation of 1.25 ns at 360 K, pH 7 with explicit water and urea (8 M), saved every 10 ps (Caflisch, A., unpublished results).

For every barnase conformation the electrostatic energy in solution and solvation energy are evaluated by the MS/FDP and MS/GB methods. The results of the fits are shown in Table 2. There is a linear correlation between the results of the two methods. The correlation coefficient for energies in solution is 0.814 and 0.931 for barnaseT300 and barnaseT360, respectively. Thus the present method approximates well the continuum electrostatic energies in aqueous solution for structures minimized with a distance-dependent dielectric as well as for conformations obtained from MD simulations with explicit solvent molecules.

5.5. Estimation of the Grid Error. The error due to the grid approximation in the evaluation of the self-energies is estimated by subsequently rotating the protein barnase by half a degree for a total of 720 rotations (360°). The conformation of barnase is the one assumed after 600 ps of the barnaseT360 MD trajectory. For every orientation the electrostatic energy is evaluated by the MS/GB method. This procedure is repeated with different grid spacing of 0.3, 0.4, 0.5, and 0.6 Å.

The analysis has been done on the solvation energy, since this quantity is directly affected by the statistical error. The electrostatic energy in vacuo is not calculated on a grid and is consequently independent of rotations or translations. The results are shown in Table 3. The average solvation energy is approximately the same for the three different grid spacings, while the standard deviation ranges from 1.2 kcal/mol for a grid spacing of 0.3 Å to 3.4 kcal/mol for a grid spacing of 0.6 Å. The CPU time required for one energy evaluation ranges from about 86 s (spacing of 0.3 Å) to about 38 s (spacing of 0.6 Å).

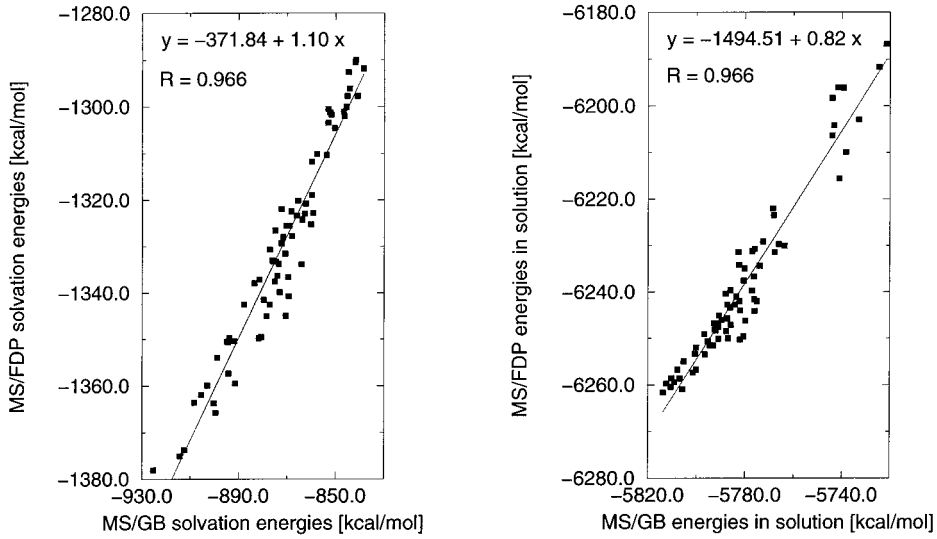

Figure 5. Electrostatic solvation energies (a, left) and energies in solution (b, right) calculated by the MS/GB and the MS/FDP methods for the 71 methylammonium ion/HIV PR complexes.

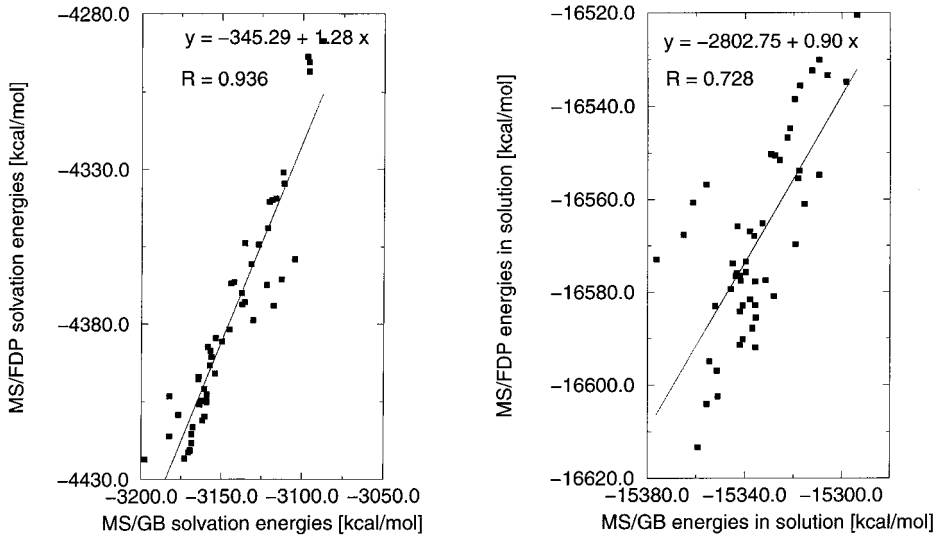

Figure 6. Electrostatic solvation energies (a, left) and energies in solution (b, right) calculated by the MS/GB and the MS/FDP methods for the 52 methylammonium ion/thrombin complexes.

TABLE 3: Statistical Error Due to the Grid Approximation: Electrostatic Solvation Energy of Barnase

	ΔE [kcal/mol]				
grid spacing (Å)	average	std dev	min	max	CPU time (s)
0.3	−860.0	1.2	−863.0	−856.9	85.9
0.4	−859.9	1.8	−864.3	−854.4	54.2
0.5	−859.9	2.5	−866.6	−852.9	42.8
0.6	−860.1	3.4	−870.9	−848.7	37.5

6. Conclusions

An efficient method for the evaluation of electrostatic energies of molecules in pure solution in the continuum approximation has been presented. The effect of the introduced approximations has been evaluated by comparison with finite difference calculations. It has been shown that the errors originating from them are low and the present method yields a correct description of the electrostatics of macromolecules in solution. These errors are smaller than the differences originating from various descriptions of the dielectric discontinuity surface. There is a significant gain in speed with respect to the finite difference calculations.

The molecular systems used for the validation included not only the standard case of the solvation energies of small organic compounds but also some macromolecular complexes in order to investigate the possible applications of the method. These test cases have highlighted two important features: The fundamental role of the discontinuity surface between the two dielectrics, and the importance of a distinct analysis of solvation energies and energies in solution. In this sense the typical test case of the solvation energies of the small molecules cannot be considered sufficient. It would suggest to use the vdWS-01/GB method in order to reproduce the results obtained by the MS/FDP calculations, which are the closest to the experimental values. But this choice is shown to be inadequate for macromolecules, where only the use of the same surface in UHBD and in eq 18 provides energies in solution that correlate well, even if shifted. Such a shift is not problematic, since in general one is interested in energy differences. The analysis of the solvation energies is not sufficient to validate an electrostatic model. When compared with an exact calculation, solvation energies and energies in solution behave differently. This is particularly evident for molecular complexes that show a strong anticorrelation between solvation energies and energies in vacuo.

Acknowledgment. We thank Prof. J. A. McCammon for providing the UHBD program, which was used for all the finite difference calculations. This work was supported by the Swiss Federal Office for Public Health (Nationales AIDS-Forschungs-Programm, grant no. 3139-043652.95) and the Swiss National Science Foundation (Schweizerischer Nationalfonds Grant 3100-043423.95).

References and Notes

(1) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.
(2) Gilson, M. K.; Honig, B. H. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7.
(3) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219.
(4) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187.
(5) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439.
(6) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(7) Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.
(8) Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(9) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(10) Jackson, J. D. *Classical Electrodynamics*; John Wiley & Sons: New York, 1975.
(11) Schaefer, M.; Froemmel, C. *J. Mol. Biol.* **1990**, *216*, 1045.
(12) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578.
(13) Born, M. *Z. Phys.* **1920**, *1*, 45.
(14) Kirkwood, J. G. *J. Chem. Phys.* **1939**, *7*, 911.
(15) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.
(16) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1.
(17) Kozaki, T.; Morihashi, K.; Kikuchi, O. *J. Mol. Struct.* **1988**, *168*, 265.
(18) Kozaki, T.; Morihashi, K.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547.
(19) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(20) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(21) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1989**, *10*, 386.
(22) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1990**, *11*, 401.
(23) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 909.
(24) Caflisch, A. *J. Comput.-Aided Mol. Design* **1996**, *10*, 372.
(25) Caflisch, A.; Karplus, M. *J. Mol. Biol.* **1995**, *252*, 672.
(26) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(27) MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Roux, B.; Schlenkrich, M.; Smith., J. C.; Stote, R.; Straub, J.; Wiorkiewicz-Kuczera, J.; Karplus, M. *FASEB J.* **1992**, *6*, A143.
(28) Jean-Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. H.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454.
(29) Mohan, V.; Davis, M. E.; McCammon, J. A.; Pettitt, B. M *J. Phys. Chem.* **1992**, *96*, 6428.
(30) Sharp, K.; Jean-Charles, A.; Honig, B. *J. Phys. Chem.* **1992**, *96*, 3822.
(31) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(32) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(33) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.
(34) Miranker, A.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29.