Incorporating Dipolar Solvents with Variable Density in Poisson-Boltzmann Electrostatics

Cyril Azuara,* Henri Orland,† Michael Bon,† Patrice Koehl,‡ and Marc Delarue*

*Unité de Dynamique Structurale des Macromolécules, URA 2185 du Centre National de la Recherche Scientifique, Institut Pasteur, Paris, France; †Institut de Physique Théorique, Commissariat à l'Énergie Atomique-Saclay, Gif/Yvette, France; and ‡Department of Computer Science and Genome Center, University of California, Davis, California

ABSTRACT We describe a new way to calculate the electrostatic properties of macromolecules that goes beyond the classical Poisson-Boltzmann treatment with only a small extra CPU cost. The solvent region is no longer modeled as a homogeneous dielectric media but rather as an assembly of self-orienting interacting dipoles of variable density. The method effectively unifies both the Poisson-centric view and the Langevin Dipole model. The model results in a variable dielectric constant $\epsilon(\vec{r})$ in the solvent region and also in a variable solvent density $\rho(\vec{r})$ that depends on the nature of the closest exposed solute atoms. The model was calibrated using small molecules and ions solvation data with only two adjustable parameters, namely the size and dipolar moment of the solvent. Hydrophobicity scales derived from the solvent density profiles agree very well with independently derived hydrophobicity scales, both at the atomic or residue level. Dimerization interfaces in homodimeric proteins or lipid-binding regions in membrane proteins clearly appear as poorly solvated patches on the solute accessible surface. Comparison of the thermally averaged solvent density of this model with the one derived from molecular dynamics simulations shows qualitative agreement on a coarse-grained level. Because this calculation is much more rapid than that from molecular dynamics, applications of a density-profile-based solvation energy to the identification of the true structure among a set of decoys become computationally feasible. Various possible improvements of the model are discussed, as well as extensions of the formalism to treat mixtures of dipolar solvents of different sizes.

INTRODUCTION

A quantitative evaluation of the interactions of biological macromolecules with the surrounding solvent is required to understand their stability and functional properties as well as their interactions with other ligands (substrates, allosteric effectors...) (1). The solvent contribution is crucial for the evaluation of the electrostatic effects in enzymes, which have been recognized to be extremely important (2–4). This, in turn, makes it a key issue in structure-based drug-design studies. An atomic description and understanding of solvation is in general very difficult to obtain experimentally, so one often resorts to computer simulations to evaluate solvent interactions and solvation energies as described in many recent reviews (for example, see (5)).

Broadly speaking, currently available methods to compute solvation energy fall into two classes depending on whether they treat the solvent molecules explicitly or implicitly. While explicit representations such as those used in molecular simulations can provide an accurate treatment of solute-solvent interactions, they drastically increase the system size. In addition, these interactions need to be averaged over long simulation times before results become meaningful, especially as solvent and solute have very different dynamical properties. In response to these problems, simplified models have been developed that treat the solvent implicitly. In this second class, the Poisson-Boltzmann equation (PBE) has been recognized to give accurate results in a number of situations and can be solved numerically efficiently for solutes of arbitrary shapes using finite difference methods, as pioneered by Warwicker and Watson (6) and further developed by Gilson et al. (7,8), Nicholls and Honig (9), and Rocchia et al. (10). Several programs that solve either the linearized version of PBE, or directly the nonlinear PBE, with a variety of scientific computing techniques are now available (see (11,12) for recent reviews), among which the multigrid method is the fastest one (13–16).

Despite its success, PBE is only a mean-field approximation to the multibody problem of solvent-solute electrostatics interactions. It is based on several approximations that proved to be limitations in some cases. For example, PBE does not include effects due to ion size or ion-ion correlations in its treatment. Solutions have been proposed to account for at least ion size using either a single size (17) or two different sizes (18), yielding a size-modified Poisson-Boltzmann equation. This effectively overcomes one of the deficiencies of standard PBE methods, namely the point-charge approximation. Including ion-ion correlations, however, is a more difficult problem that is still the subject of active research. Current approaches are restricted to problems involving solutes of small sizes (19,20) or with very simple geometries such as a charged plate (21), or a linearly charged cylinder modeling DNA (22).

Submitted February 22, 2008, and accepted for publication September 3, 2008.

Address reprint requests to Patrice Koehl, E-mail: koehl@cs.ucdavis.edu; or Marc H. Delarue, Tel.: 33-01-45-68-86-05; E-mail: delarue@pasteur.fr.

Cyril Azuara's present address is Equipe de Bioinformatique Génomique et Moléculaire, INSERM U-726, Université Paris 7, 75251 Paris cedex 05, France.

Editor: Ron Elber.

0006-3495/08/12/5587/19 $2.00

doi: 10.1529/biophysj.108.131649

More importantly, the PBE method contains a very rough approximation which consists in using a constant and somewhat arbitrary value for the dielectric constant of the protein (usually set at 2–4), that abruptly jumps to 80 at the interface between the protein and the solvent. Moreover, because of polarization effects in the vicinity of charges, it is expected that the representation of the solvent as a homogeneous dielectric medium is bound to be erroneous close to the interface. The need to have a smooth dielectric profile at the border of the solute has long been recognized and this problem still attracts a lot of attention and controversy (23), especially as the concept of dielectric constant is not, per se, a microscopic one (24). A number of attempts have been made to derive the function $\epsilon(\vec{r})$ from first principles. The most complete derivation comes from Ehrenson (25) for a dipolar fluid in the electric field of an ion or a dipole. This was applied to proteins by Hassan et al. (26) to derive a screened Coulomb potential with a distance-dependent $\epsilon(r)$ that can then be used in molecular dynamics simulations without explicit solvent (27).

In this article, we are interested in a solvent model with built-in $\epsilon(\vec{r})$ dependence that allows the rapid prediction of solvent density around macromolecular solutes. We show how to introduce (free) dipolar charges representing the solvent molecules in a Poisson-Boltzmann formalism. The system self-consistently adjusts its position-dependent solvent density and the electric potential is obtained numerically, not analytically. For a solute with a complex shape, the resulting dielectric profile and solvent density are not spherically symmetric but depend on the chemical nature and partial charge of the exposed nearby solute atom(s).

There is a variety of situations on both the experimental and computational sides where one would like to have access to the solvent density map, starting from just the PDB atomic coordinates and the partial charges of the molecule (28). On the experimental level, such a solvent map could be very helpful to interpret x-ray electron density maps in the final stages of model building, when water molecule assignment starts. It would also help interpreting both SAXS and SANS experimental data, where the hydration shell is usually modeled as a mere cushion of constant width around the solute (29).

On the modeling level, it would be very useful for quickly computing the electrostatic part of solvation energies with a more realistic model than PBE, especially as a molecular understanding of the nature of the hydrophobic effect is still lacking for macromolecules and remains a subject of active research (30–33). The structure of water around both polar and apolar solutes is also still actively studied (34–36). Knowledge of the solvent density profile has clear implications for the calculation of solvation energies and the modeling of the hydrophobic effect at different length scales (31). Indeed, as included in the van der Waals theory of capillarity (37), it has been proposed that the free energy contains a term proportional to the integral of the squared gradient of the solvent density profile (32,33).

Here we show how a simple solvent description based on an assembly of freely orienting and interacting dipoles on a grid can be readily incorporated into and solved within the Poisson-Boltzmann formalism. This is in effect a generalization of the Langevin Dipoles-Protein Dipoles model developed by Warshel and Levitt (38), Warshel and Russell (39), Russell and Warshel (40), and Warshel and Papazyan (41), with the key additional feature that the dipoles are now allowed to have a variable density at each allocation grid point around the solute. The use of a lattice ensures that size exclusion effects are included in this Poisson-Boltzmann-Langevin (PBL) model.

A preliminary account of the method and its implementation through a web site has been presented recently (42), and detailed numerical applications to the case of a charged planar surface have just been described by Abrashkin et al. (43). Here we give a full description of the method, starting with a phenomenological derivation of the free energy of the system. Minimization of the free energy then leads to a Poisson-Boltzmann-Langevin equation (PBLE), which we solve by rapid numerical methods.

THEORY

The PBL equation

Let us consider a fixed charged biomolecule (the solute) immerged in a solvent, and surrounded by an ion atmosphere. We represent the solvent as an assembly of freely orientable dipoles $\vec{p}_j$ in a z/z electrolyte with the free ions carrying a charge $\pm ze$, where e is the charge of the electron and z the valency of the free ions. The solute and the free dipoles and ions are embedded in a lattice where each site j bears a spinlike occupancy $d_j = 0, 1$ and $s_j = -1, 0, 1$ for the free dipoles and the free ions (see Fig. 1), with chemical potential μ_{dip} and μ_{ion}, respectively. The lattice allows for imposing directly steric hindrance between the different species, without recourse to an additional repulsion potential. We suppose for the moment that both ions and dipoles have the same diameter a (see Appendix B for the case where ions and dipoles have a different size). The dipoles and the ions create a local charge density in the solvent that is to be added to the fixed charge density $\rho_f(\vec{r}) = \sum q_i \delta(\vec{r} - \vec{r}_i)$ of the solute, all of them interacting solely through Coulomb potential $\int d\vec{r} d\vec{r}' \rho(\vec{r}) v(|\vec{r} - \vec{r}'|) \rho(\vec{r}')$ and $\rho(\vec{r}) = \rho_f(\vec{r}) + \rho^{elec}_{ions}(\vec{r}) + \rho^{elec}_{dip}(\vec{r})$.

As shown in Appendix A the free energy of the system described in Fig. 1 is a function of the electric potential $\Phi(\vec{r})$,

$$\beta\mathcal{F} = -\frac{\beta}{2}\int d\vec{r}\,\epsilon_0 |\vec{\nabla}\Phi(\vec{r})|^2 + \beta\int d\vec{r}\,\rho_f(\vec{r})\Phi(\vec{r}) - \frac{1}{a^3}\int_{Solvent} d\vec{r} \ln\left(1 + 2\lambda_{ion}\cosh(\beta e z \Phi(\vec{r})) + \lambda_{dip}\frac{\sinh(\beta p_o |\vec{\nabla}\Phi(\vec{r})|)}{\beta p_o |\vec{\nabla}\Phi(\vec{r})|}\right), \quad (1)$$

where we have set $\lambda_{ion} = e^{\beta\mu_{ion}}$ and $\lambda_{dip} = e^{\beta\mu_{dip}}$.

We can interpret the above free energy functional as follows. The first and second terms correspond to the internal electrostatic energy and the energy of interaction of the fixed charges of the solute with the electrostatic potential, respectively. The third term is the logarithm of the sum of three contributions, representing the vacant sites, the ions, and the dipoles in the solvent region, respectively. Note that this free energy functional simplifies to known free energies in the following two limiting cases:

1. $\lambda_{dip} = 0$: because $\lambda_{ion} \ll 1$, a Taylor expansion gives back the known free energy functional of PBE.
2. $\lambda_{ion} = 0$: as $\lambda_{dip} \gg 1$, we indeed recover the expected energy for the dipolar solvent term (44) namely, in its discrete form, $-kT\sum_i \ln(\sinh(u_i)/u_i)$, where $u_i = \beta p_0 E_i$.

Let us define

$$u = \beta p_0 |\vec{\nabla}(\Phi(\vec{r}))| = \beta p_0 E \tag{2}$$

and

$$D(\Phi(\vec{r})) = 1 + 2\lambda_{ion}\cosh(\beta ez\Phi(\vec{r})) + \lambda_{dip}\frac{\sinh(\beta p_o|\vec{\nabla}\Phi(\vec{r})|)}{\beta p_o|\vec{\nabla}\Phi(\vec{r})|}. \tag{3}$$

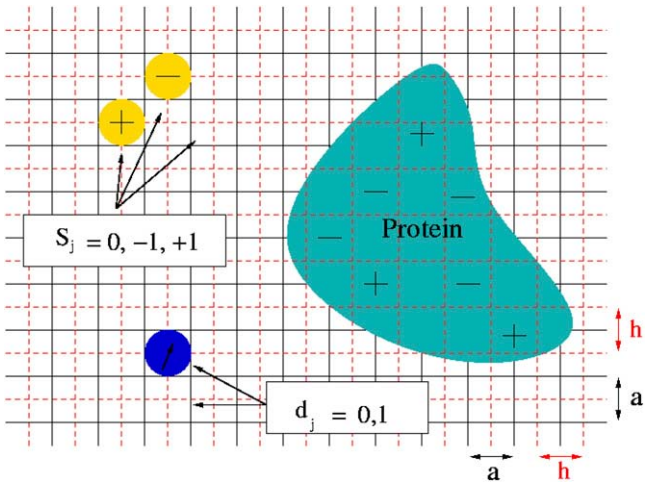

FIGURE 1 The PBL model and its lattice representation. The atmosphere around the solute includes free ions of charge $+ze$ or $-ze$ and dipoles representing the water. Each lattice point is associated with a spinlike variable indicating the presence of a free ion (S_j) or a spherical dipole (d_j) of the same size. The grid size (a) used to place these entities and ensure exclusion effects is not necessarily the same as the grid size (h) used for discretizing the free energy equation. The interface between the solute interior and the solvent lattice is defined as either the accessible surface (using $R_{probe} = 1.4$ Å) or the molecular surface (using $R_{shrink} = 1.0$ Å).

Setting $\partial\mathcal{F}/\partial\Phi = 0$ we obtain, after some algebra, the PBLE, which contains four more terms compared to PBE. Several terms can be factorized, leading to

$$\epsilon_0\left(1 + \frac{\lambda_{dip}\beta p_0^2 F_1(u)}{\epsilon_0 a^3 D(\Phi(\vec{r}))}\right)\Delta\Phi(\vec{r}) + \rho_f(\vec{r}) = \frac{2\lambda_{ion}ez\sinh(\beta ez\Phi(\vec{r}))}{a^3 D(\Phi(\vec{r}))}(1 + \lambda_{dip}u^2 F_1(u)/D(\Phi(\vec{r}))) + \frac{\lambda_{dip}\beta^3 p_0^4 \vec{\nabla}\Phi(\vec{r}).(\vec{\nabla}\Phi(\vec{r}).\vec{\nabla})\vec{\nabla}\Phi(\vec{r})}{a^3 D(\Phi(\vec{r}))}(-F_2(u)/u + \lambda_{dip}F_1(u)^2/D(\Phi(\vec{r}))) \tag{4}$$

with

$$F_1(u) = \frac{\sinh(u)}{u^2}\mathcal{L}(u), \quad F_2(u) = \frac{\sinh(u)}{u^2}\left(1 - \frac{3}{u}\mathcal{L}(u)\right), \tag{5}$$

where $\mathcal{L}(u) = 1/\tanh(u) - 1/u$ is the Langevin function and where $\sinh(u)/u$, $F_1(u)$, and $F_2(u)$ are related to the spherical Bessel functions j_0, j_1, and j_2. Note that $F_1(u) \to 1/3$ and $F_2(u)/u \to 1/15$ when $u \to 0$. In general, however, the approximation $u \ll 1$ is not justified (see, for example, Fig. 3 *c* below), leading to substantial nonlinear effects.

The term $D(\Phi(\vec{r}))$ in Eqs. 3 and 4 enforces steric avoidance of the different free species in solution, as it will discourage ($D > 1$) the placement of any species at lattice points that are already occupied either by a free ion or a dipole solvent molecule.

Finally, we need to specify how the fugacities are related to the ions and dipole concentrations, C_{ion} and C_{dip}, respectively. Following Borukhov et al. (45) and setting $\mathcal{N}_a$ as the Avogadro number, we find

$$\lambda_{dip} = \frac{\mathcal{N}_a C_{dip} a^3}{1 - 2\mathcal{N}_a C_{ion} a^3 - \mathcal{N}_a C_{dip} a^3}, \tag{6}$$

$$\lambda_{ion} = \frac{\mathcal{N}_a C_{ion} a^3}{1 - 2\mathcal{N}_a C_{ion} a^3 - \mathcal{N}_a C_{dip} a^3}. \tag{7}$$

The volume correction due to ions is usually very small. At $C_{dip} = 55$ M, the value of λ_{dip} varies numerically from 2.7 to 8.2, depending on the diameter $a = 2.8$ or 3.0 Å of the solvent sphere.

In Appendix B, we briefly show how this formalism can be extended to ions of different valency, dipoles with a size different from the ion size or mixtures of dipoles with different dipole moments and/or sizes.

Relationship with standard electrostatics equations in matter

The general form of the PBL equation (Eq. 4) is not standard, but it can be rewritten in a more familiar form. First, we note that the ion density surrounding the solute is given by $-\partial\mathcal{F}/\partial\mu_{ion}$,

$$\rho_{\pm}(\vec{r}) = \frac{1}{a^3}\left(\frac{\lambda_{ion}e^{\mp\beta ez\Phi(\vec{r})}}{D(\Phi(\vec{r}))}\right), \tag{8}$$

which is identical to the ion density with a size exclusion effect (45). The resulting electrostatic density of charge then reads $\rho_{ions}^{elec} = ze(\rho_+ - \rho_-)$.

Similarly, the dipole density is given by $-\partial\mathcal{F}/\partial\mu_{\rm dip}$,

$$\rho_{\rm dip}(\vec{r}) = \frac{1}{a^3}\left(\frac{\lambda_{\rm dip}\sinh(u)/u}{D(\Phi(\vec{r}))}\right), \tag{9}$$

where u and $D(\Phi(\vec{r}))$ are given by Eqs. 2 and 3, respectively. This is similar to equations derived for comparable models (20,46) but now includes explicitly hard-core repulsion between the solvent dipoles (and free ions) through the $D(\Phi(\vec{r}))$ term.

The local dipole moment per unit volume $\vec{P}(\vec{r})$ (polarization density) is given by $-\partial\mathcal{F}/\partial\vec{E}$. Performing the derivation, we find an expression very similar to previous work (20), with the additional $D(\Phi(\vec{r}))$ ensuring self-avoidance of dipoles and free ions,

$$\vec{P} = \langle p\rangle\rho_{\rm dip}(\vec{r})\frac{\vec{E}}{|\vec{E}|} = \beta p_0^2 F_1(u)\frac{\lambda_{\rm dip}}{a^3 D(\Phi(\vec{r}))}\vec{E}, \tag{10}$$

where $\langle p\rangle = p_0\mathcal{L}(u)$ is the average dipole strength. $\vec{P}(\vec{r})$ is always parallel to $\vec{E}(\vec{r})$, the total electric field, but not to $\vec{E}_0(\vec{r})$, the electric field created by the solute only, as proposed in the Langevin Dipoles-Protein Dipoles model (40).

Setting as usual (47) $\vec{P} = \epsilon_0\chi(\vec{r})\vec{E}$, a position-specific electric susceptibility $\chi(\vec{r})$ is derived:

$$\chi(\vec{r}) = \frac{\beta p_0^2 F_1(u)}{\epsilon_0}\frac{\lambda_{\rm dip}}{a^3 D(\Phi(\vec{r}))}. \tag{11}$$

Using these definitions, the last term of PBL Eq. 4 can be recast in terms of $-\beta^2 p_0^2\vec{E}(\vec{r}).(\vec{P}(\vec{r})\cdot\vec{\nabla})\vec{E}(\vec{r})$, where $(\vec{P}\cdot\vec{\nabla})$ is formally equivalent to a local dipolar charge density operator and $(\beta p_0 E)^2 = u^2$ is dimensionless.

The PBL equation now reads

$$\begin{aligned}\epsilon_0(1+\chi(r))div\vec{E} = \rho_{\rm f} + \rho_{\rm ions}^{\rm elec}(\Phi)(1+\lambda_{\rm dip}u^2F_1(u)/D(\Phi)) \\ + ((\vec{P}\times\vec{\nabla})\vec{E}\times\vec{E})\frac{u^2}{E^2} \\ \times\left(-\frac{F_2(u)}{uF_1(u)} + \lambda_{\rm dip}\frac{F_1(u)}{D(\Phi)}\right),\end{aligned} \tag{12}$$

where ion-dipole and dipole-dipole steric exclusion couplings are clearly present and mediated by $F_1(u)/D(\Phi)$. This can be further simplified into

$$-div(\epsilon_0\vec{E} + \vec{P}) + \rho_{\rm f} + \rho_{\rm ions}^{\rm elec} = 0, \tag{13}$$

or simply $div(\vec{D}) = \rho_{\rm f} + \rho_{\rm ions}^{\rm elec}$, where $\vec{D} = \epsilon_0\vec{E} + \vec{P}$ is the electric displacement vector. We therefore recover the usual Maxwell equation in matter, as expected (47).

Dielectric properties of the bulk solvent and relationship with the Debye-Langevin formula

The dielectric profile is given by $\epsilon_{\rm sol}(\vec{r}) = 1 + \chi(\vec{r})$ and Eq. 11. If $u \ll 1$, we can set $F_1(u) = 1/3$ and approximate $D(\Phi(\vec{r}))$ by $1 + \lambda_{\rm dip}$ so that $\lambda_{\rm dip}/a^3 D(\Phi(\vec{r})) = \mathcal{N}_a C_{\rm dip}$ is independent of a. This gives the following expression for the dielectric constant in bulk solvent:

$$\epsilon_{\rm bulk} - 1 = \frac{\beta p_0^2\mathcal{N}_a C_{\rm dip}}{3\epsilon_0}. \tag{14}$$

For $p_0 = 1.85D$ (gas phase) and $C_{\rm dip} = 55$ M, we find $\epsilon_{\rm sol} = 1.0 + 11.5$; alternatively, for $p_0 = 2.35D$ (liquid phase), we find $\epsilon_{\rm sol} = 1.0 + 18.5$.

In the formalism described above, we have neglected the polarizability of the solvent dipoles. In Appendix B, we give the expression of the free energy if the dipoles have an intrinsic polarizability α_0. If we apply again the relationship $\vec{P} = -\partial\mathcal{F}/\partial\vec{E}$ and set $\lambda_{\rm ion} = 0$, we get

$$\epsilon_0(\epsilon_{\rm bulk} - 1) = \left(\alpha_0 + \frac{\beta p_0^2}{3}\right)\mathcal{N}_a C_{\rm dip}. \tag{15}$$

This is the celebrated Debye-Langevin formula (48). To see whether we may neglect the intrinsic polarizability of the water dipole, it is useful to set $\alpha_s = \beta p_0^2/3$. Numerically $\alpha_0/\alpha_s = 1/20$, and therefore we can neglect α_0.

MATERIALS AND METHODS

Solving the PBL equation

We solve the PBL equation on a regular grid, using a finite difference approach. The specifics of our solver are:

1. *Input file*. The structure of the biomolecule is read from the corresponding PDB file and is first processed to build all possible missing atoms. For proteins, this involves generating missing loops (49) and/or missing side chains (50). All polar hydrogen atoms are added to the resulting complete structure. Charges and radii are assigned according to CHARMM19 (51); they are written with the coordinates of the atoms in a pseudo PDB file in the occupancy and B-factors columns, respectively, following the PQR format of APBS. We chose to follow the PQR format as it provides us compatibility with the service PDB2PQR (52). In this way, other sets of partial charges are possible (AMBER, PARSE...).
2. *Setting the grid*. Since the solver we use is based on finite differences, we build a regular grid around the solute of interest. The user provides the number of grid points. The grid size is then computed in such a way that there is at least a distance of $2l_B$ (l_B is the Bjerrum length in water at 300 K, i.e., ~7 Å) from any point on the surface of the solute to the closest edge of the grid. The grid is positioned such that its center matches the center of the solute. Currently, the program can handle regular symmetric grids with either 65, 129, 193, or 257 points in each dimension.
3. *Defining the solute surface*. We have implemented calculation of both the accessible surface and the molecular surface of the molecule. Note that even if the solvent is treated explicitly in the PBL theory, its precision and accuracy still depends strongly on our ability to generate accurate representations of these surfaces.

The accessible surface is obtained as the envelope of the hydrated spheres representing the atoms, whose radii are the vdW radii increased by $R_{\rm probe}$ = 1.4 Å, where $R_{\rm probe}$ is the radius of a water molecule (53). The molecular surface is the lower envelope obtained by rolling a water probe of radius $R_{\rm probe}$ on the vdW surface of the molecule. We map this surface on the mesh using the method proposed by Jiang and Brünger (54). Unless otherwise stated, the surface of the solute was chosen as the accessible surface area.

4. *Parameters*. The PBL considers explicitly the solvent in the form of an assembly of solvent dipoles; it is therefore crucial to parameterize these dipoles correctly. We have tried different values for the solvent dipole

moment and its radius, i.e., $\mu_{dip} = 1.85$ or 2.1 or 2.35 or even $4.8D$ and a grid half-size of $a/2 = 1.2$, 1.4 or 1.6 Å, respectively.

5. *Solving the nonlinear differential equation*. The PBL equation is much more complex than the classical PBE equation, as the coefficients of the Laplacian operator depend on the electrostatics potential. There is also an additional differential operator $\vec{E}.(\vec{P}.\vec{\nabla})\vec{E}$, also with nonlinear coefficients. We have tried both a standard relaxation technique (55) and a Newton method with multigrid preconditioner to solve the PBL (14). For all systems described in this article, both methods reached convergence to the same electrostatic potential, with the latter being significantly faster. This is therefore the method implemented on the PDB_hydro server (42).

We have checked that solving the PBL equation with $p_0 = 0$ yields the same results as PBE in APBS (56) with the dielectric constant set as uniform and equal to 1 ($\epsilon = 1$). For a protein of average size (250 amino acids), a grid of 129^3 points leads to a grid size of ~0.7 Å. The calculation used to take ~40–45 mn CPU on a desktop workstation (42). We have recently implemented a new version of the algorithm using Eq. 13 together with Gauss theorem instead of Eq. 4 (23). The whole code has been optimized and is ~20 times faster. The new program, AquaSol, is freely available upon request (M. Delarue and P. Koehl, unpublished).

Analyzing the outputs of the PBL solver

Solution of the PBL equation provides the values for the electrostatics potential at all grid points. This potential can be used to derive ion and dipole density maps, to compute radial density profiles and the partial solvation energy $\mathcal{F}_1$ (see below), and to place water molecules around the molecule.

1. *Density map format*. We have chosen to write all maps generated by our solver (namely maps of the electrostatic potential, of the solvent density, of the ion densities, and of the electric field) using the electron density map format of either CNS (57), EZD (CCP4), UHBD (APBS), or OpenDX (PyMol). Maps in this format can be visualized either with O (58), Coot (59), or Pymol (http://www.pymol.org). Other map formats can be obtained by using MapMan from the Rave suite of programs (http://alpha2.bmc.uu.se/usf/rave.html).
2. *Mapping the electrostatic potential at the surface of the molecule*. For each atom at the surface of the molecule, we first find all grid points that are contained within a sphere of radius R_{vdw} + 3 Å centered on that atom. The mean value of the electrostatics potential of these grid points is then mapped as a color on the surface of the atom, using Pymol.
3. *Placing water molecules or ions around the molecule*. First we sort the $\rho_{dip}(\vec{r})$ values obtained from Eq. 9 in descending order. Water molecules are placed by walking down the list until the desired N_w number of water molecules is reached, as calculated by the known hydration level h (in g/g), or until the desired ρ_{thres} has been reached; each time a water molecule is placed, we eliminate points within 1.5 Å of this position from the list. We use the electrostatic field to orient the dipole.

A similar approach is used for placing ions, this time using $\rho_{ions}(\vec{r})$ as a guide.

4. *Computing solvent density profiles*. We compute dipole radial density profiles for the different atom types C, N, S, O, P. First we generate probe points on the surface of these atoms as described by Shrake and Rupley (60). For each probe point, we draw a (radial) line passing through the center of the corresponding atom and this probe surface point. A segment of 12 Å on this line outside the atom is kept if it does not cross any other atom of the solute. The values of $\rho_{radial}(r)$ along each segment are obtained by interpolation of $\rho_{dip}(\vec{r})$ using 0.1 Å steps; these profiles are then averaged over all accessible probe points.
5. *Calculating solvation energy*. The partial solvation energy $\mathcal{F}_1$ (see Eq. 16) of each atom can be computed from solvent radial density profiles using the integral of its squared derivative. We then divide (normalize) the $\mathcal{F}_1$ free energy of each atom by its accessible surface area. The mean value of this ratio, which can be calculated and averaged for each atom type, is then compared to the Eisenberg and McLachlan (61) atomic solvation parameters (ASP) coefficients, for the same atom types C, S, O−, N+, and polar O/N. This amounts to calculate a free energy of solvation with a surface tension coefficient that is not only atom-type dependent, as in Eisenberg and McLachlan method (61), but also position-dependent, as dictated by the local solvent profile.

Decoys and test sets

The decoys of Park and Levitt (62) were used for the L7/L12 ribosomal protein C-terminal domain (1CTF). For lysozyme (1LZR), Nseq-1 circularly permuted mutants were generated by decorating the same 1LZR backbone with a sequence shifted by 1, 2, . . ., $N - 1$ and then reconfiguring, each time, the side chains using mean-field optimization theory (42,50). We also computed the ENVIRON energy for each of these decoys (63), as a check.

The list of protein structures used for density profiles and surface tensions statistics is the following: 1ARB, 2EXO, 5NLL, 1SRP, 1CP4, 1EDB, 1PHP, 2CTB, 2DRI, 2APR, 2FCR, 2ACS, 1BL8, 1BXW, 1C3W, 1FEP, 1MAL, 1OPF, 1PRN, 1XIO, 2FCP, 1AQ6, 1AUO, 1BBH, 1BIS, 1BKP, 1BSL, 1BXG, 1COZ, 1IMB, 1ISA, 1PGT, 1R2F, 1TC1, 2CCY, 3SDH, 8PRK, 4TMK, 2PTC, 1CHO, 1PPF, 2KAI, 1CSE, 3TPI, 1GLA, 1BRS, 1UDI, 1DHK, 1FSS, 1YDR, 1DFJ, 2PCC, 1JHL, 1NCA, 1OSP, and 1KB5.

Molecular dynamics

The GROMACS package was used (64) with the SPC/E water model and force-field GROMOS 53A6. The protein was 1TIM. The system was heated and then equilibrated at 1 bar and 310 K before subjecting it to molecular dynamics. The protein is held fixed with harmonic constraints on all positions with an elastic constant of 1000 kJ/Mol per Å^2. The box size is 74.9 Å in every direction, with periodic boundary conditions. Temperature is held at 310 K. Concentration of NaCl is 0.1 M. The box contains 12,548 H_2O molecules, 26 Na^+ and 26 Cl− ions. Particle mesh Ewald is used for treating electrostatic interactions. The simulation is run for 4,000,000 steps (time step is 2 fs) and the list of neighbors is updated every five steps. A total of 8000 frames were written (each picosecond) and used to calculate time-averaged dipole density $\rho_{MD}(\vec{r})$ as well as the $\vec{P}(\vec{r})$ vector field. Three different trajectories were generated with different initial velocities, and it was checked that the resulting solvent density was consistent for the different runs, therefore ensuring that the simulation had converged. The correlation coefficients between the solvent density maps of the different runs were 0.95 for grids of 1.4 Å size and 0.99 for grids of 2.1 Å size. This protocol is similar to the one described in typical references in the field such as Simon et al. (65), Higo and Nakasako (66), and Higo et al. (67).

RESULTS

Calibrating the method with simple solutes

The PBE method is usually calibrated using the Born formula for the electrostatic energy of spherical ions of different radii and charges, which is exact for an ion in a dielectric medium of constant dielectric constant ϵ. However, there is a priori no reason to expect that the solvation energy in a solvent of variable density and variable dielectric constant should be exactly equal to the Born energy. Still, by analogy with PBE calibration, we compare the Born and PBL free energy for free ions of different radii and charges, and scan different values of p_0 and a (see Fig. 2, *top left* and *top right*). We see that $p_0 = 2.35D$ with $a/2 = 1.4$ Å give excellent results over a wide range of both ionic radii and charges. For higher values of a, the set of PBL curves with different p_0 go up in Fig. 2 *top left*, increasing the gap with the PB curve, unless the value of p_0 is also further increased. The same

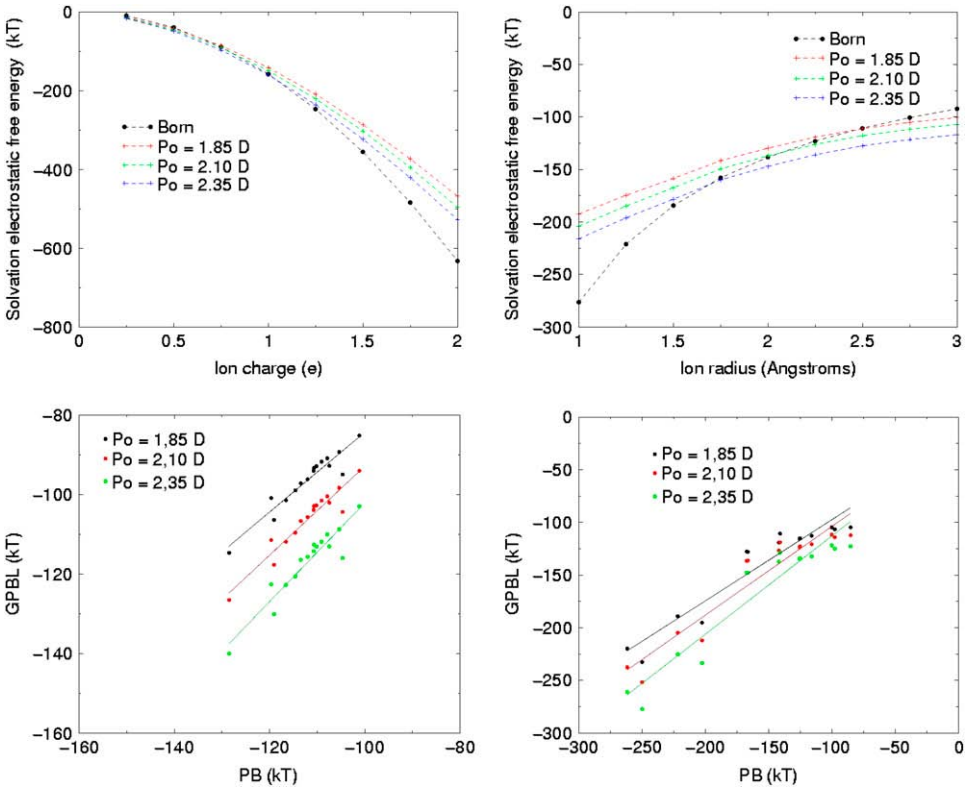

FIGURE 2 Calibration of the magnitude p_0 and size a of the solvent dipole using small molecules and ions. PB versus PBL energy for ions of different charges (*top left*) and radii (*top right*) for different values of p_0 with $a/2 = 1.4$ Å. Same thing for the 16 uncharged amino acids (*bottom left*), and the remaining four charged amino acids together with a set of charged small molecules (*bottom right*). This includes acetate, dimethylphosphate, ethoxyde, ethylthiolate, guanidinium, imidazolium, methylammonium, methylthiolate, methoxyde, and tetramethylammonium.

trend is observed in Fig. 2 *top right*, where again a value of $a/2 = 1.5$ Å would also agree with PB results, provided that p_0 is increased to 2.8–3.0D. For purely physical reasons, we will choose in the following $p_0 = 2.35D$ and $a/2 = 1.4$ Å unless otherwise stated.

For small solutes like uncharged or charged amino acids (Fig. 2, *bottom left* and *bottom right*), the PBL solvation free energies are always highly correlated with the PB energies, with correlation coefficients >0.95. To obtain a slope as close as possible to 1, however, the best value of p_0 may slightly vary depending on the set of small molecules chosen. For instance, for uncharged amino acids (Fig. 2 *bottom left*) $p_0 = 2.10D$ is best, while for charged small molecules shown in Fig. 2 *bottom right*, $p_0 = 2.35D$ is best. For another set of uncharged small molecules (not shown), the best value is $p_0 = 1.85D$, with $a/2 = 1.5$ Å. It is possible that by adjusting the set of charges and rescaling them one might arrive at a more universal value of p_0 for different sets of small molecules (44), but we have not tried this.

The program was also tested with simple geometries such as two oppositely charged plates and the solvent density of dipoles rightly shows accumulation of dipoles close to the plates (43).

Comparison with classical PBE

Here we describe the comparison of the electric potentials obtained with either PBL (this work) or PBE (APBS) with $\epsilon = 80$ in the solvent region and $\epsilon = 2$ inside the solute for a typical protein, aldose reductase (2ACS), using $p_0 = 2.35D$, 0.1 M NaCl and $a/2 = 1.4$ Å. The potentials are highly correlated inside the protein (Fig. 3 *a*) but are notably different outside the protein (Fig. 3 *b*). Interestingly, the magnitude of the electric field calculated with PBLE in the solvent region can be (especially in the first hydration shell) up to 50 higher than the one calculated with PBE (Fig. 3 *c*). As $|\vec{E}(\vec{r})|$ is the main driving force for building up a position-dependent solvent density (see Eq. 9), it can be seen that $\vec{E}_{\mathrm{PB}}$ will generate an almost flat solvent density, whereas $\vec{E}_{\mathrm{PBL}}$ will generate a position-dependent solvent density. These calculations also show that in general $u_{\mathrm{PBL}} > 1$, meaning that a linear version of PBLE would not be accurate. All these conclusions remain valid at higher values of p_0 up to 4.8D, as well as with different values of the dipole size. The question of whether or not this density is a physically reasonable one, depending on the nature of the atom exposed to the surface (see below), will be examined in detail in the next paragraphs.

It is also possible to derive an ion density from the PBL formalism (Eq. 8) which can be plotted in the following way: for each density value taken as a threshold for placing individual ions, we report the number of ions that have a density above this value and not overlapping with any other assigned ion. The results for a moderately negatively charged molecule (aldose reductase 2ACS, Fig. 4 *a*), compared to a highly and negatively charged one (halophilic alcohol dehydrogenase 2B5V, Fig. 4 *b*) indicate, as expected, that each of them condense more counterions than like-ions and that highly charged solutes condense more counterions than moderately charged solutes. This is also true with a B-DNA molecule as a solute (a dodecamer 9BNA, Fig. 4 *c*). Interestingly, all these curves (except for DNA) typically show two regimes: a

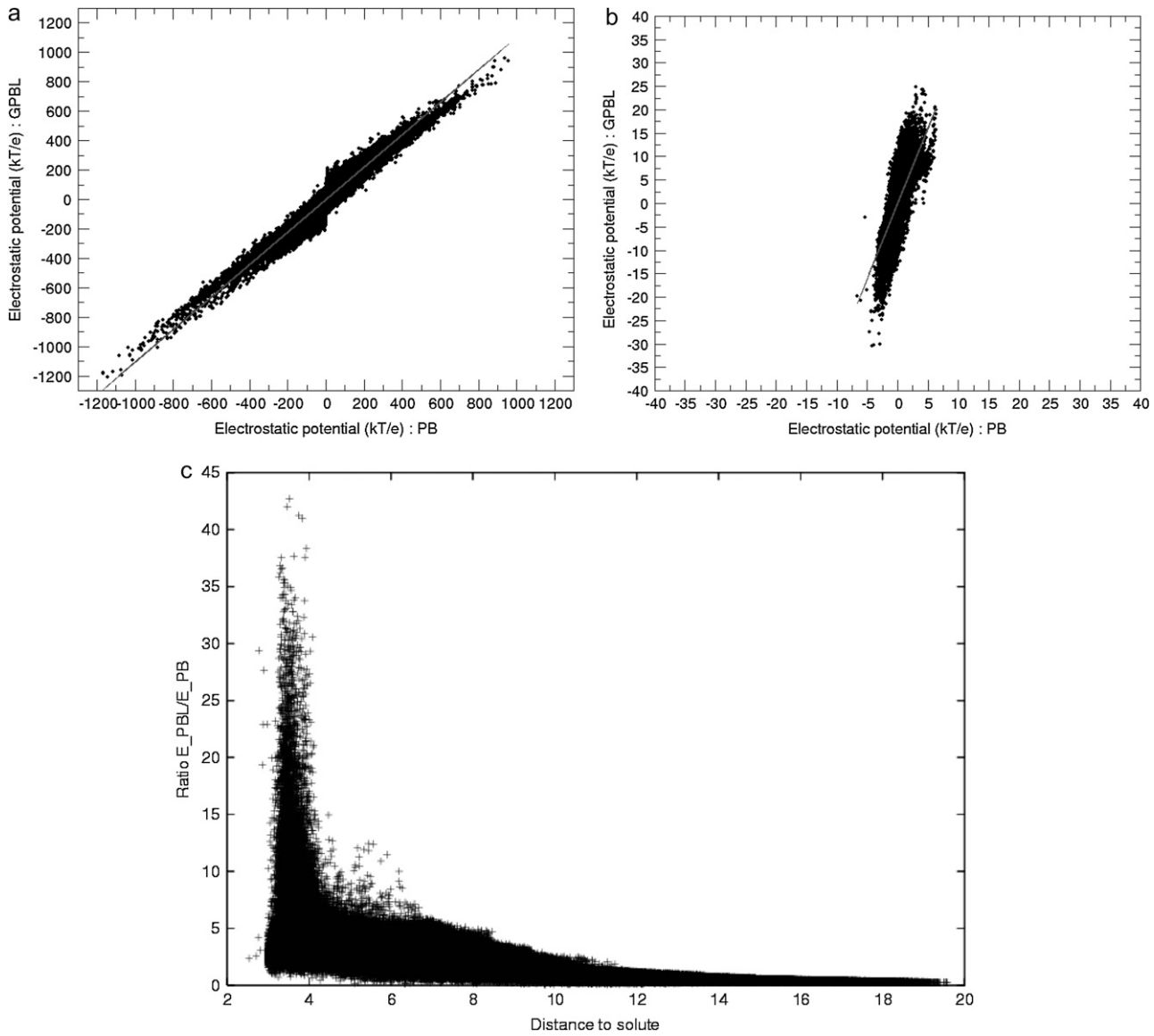

FIGURE 3 Comparison of electrostatic potentials and fields for the PB and PBL methods using 2ACS. Correlation between the PB and GBPL electrostatic potentials inside the protein solute (*a*) and outside, i.e., in the solvent region (*b*). (*c*) Ratio of the magnitudes of the electric field E found with either PBE or PBLE, as a function of the distance to the solute. This is especially important as it is the electric field that drives the solvent density (see Eq. 9). Here we have $a/2 = 1.2$ Å and $p_0 = 1.85D$; the same kind of curves was observed with increasing values of p_0 up to $4.8D$ and/or choosing $a/2 = 1.4$ Å.

dominant regime characterized by a slope at the midpoint transition that intercepts the x axis close to the net charge of the solute molecule, and a weaker regime that is common to both counterions and like-ions, at least for protein solutes.

New quantities given by PBL

We now discuss some of the new quantities given by PBL, as compared to PB, namely the direction and intensity of the solvent dipoles and examine their correlation through space. In Fig. 5 *a*, we have plotted the quantity $\langle \vec{P}(\vec{r}) \cdot \vec{P}(\vec{r}') \rangle$ as a function of the distance $\vec{r}$ (using spherical shells) away from the surface of the solute, for different values of $|\vec{r} - \vec{r}'|$. The $\vec{P}(\vec{r})$ values are calculated using Eq. 10. Starting at ~0.78 next to the solute surface, the correlation decreases to <0.5 at 5 Å away from the solute and 0.0 at 10 Å, for nearest neighbors. This is in-line with the mean value of the dipole intensity $\langle |\vec{P}| \rangle$, which also decreases to 0 after 3–4 hydration shells (data not shown).

Another quantity of interest is the $\epsilon(\vec{r})$ map (or, equivalently, the $\chi(\vec{r})$ map), which is given by Eq. 11 and is shown in Fig. 5 *b*. It can be seen to be smoothly varying away from the solute, reaching its plateau value after 5–7 Å. Indeed, for low values of $u = \beta p_0 |\vec{E}|$, we can expand Eq. 11 into $\epsilon(\vec{r}) = \epsilon_0 + c_{\text{dip}} \beta p_0^2 / 3$. However, the derived numerical value of ϵ_{bulk} (12.5 for $p_0 = 1.85D$ or 19.5 for $p_0 = 2.35D$) is substantially

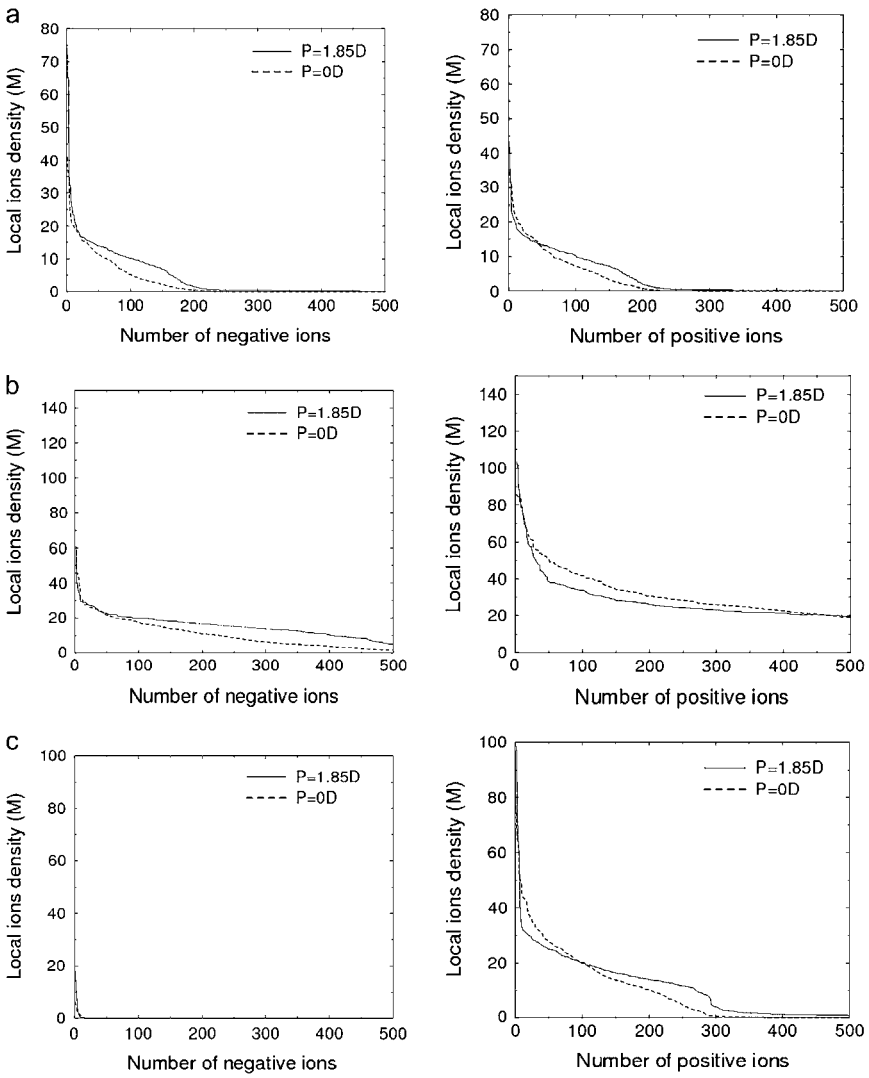

FIGURE 4 Ion condensation. Number of negative (*left panel*) and positive (*right panel*) ions placed as a function of threshold density ρ_{thres} for molecules of different net charges. The result of the same calculation with $p_0 = 0$ is indicated with a dotted line. (*a*) Aldose reductase (2ACS, net charge $= -2$); (*b*) halophilic alcohol dehydrogenase (2B5V, net charge $= -33$); and (*c*) B-DNA (9BNA, 12 bp, net charge $= -24$).

lower than the experimental value (i.e., 78); this will be addressed in the Discussion (see also (48,68,69)). For higher values of p_0 ($p_0 = 4.8D$), the peak (overshoot) around the first hydration shell is more pronounced and clearly demonstrates the nonlinear character of the underlying physical model (70).

Solvent density profiles and hydrophobicity scales

Solvent density profiles are calculated with Eq. 9 and typical ones are shown in Fig. 6: clearly, radial solvent density profiles close to polar atoms exposed to the surface (Oxygen or Nitrogen) present higher deviations from a flat profile, compared to nonpolar atoms (Carbon, Sulfur). This higher density of the solvent in the immediate vicinity of polar or charged atoms reflects both a polarization of the solvent and a first hydration layer whose strength depends on the nature of the exposed surface of the solute. In contrast to molecular dynamics (MD) simulations for example, the profiles show no second hydration layer.

At the amino acid level, we can derive a solvation score from the maximum value of this density profile, $\langle \rho_{\text{radial}}^{\max} \rangle$, averaged over all the constituent atoms of every amino acid, for each type of amino acid, thereby generating an hydrophobicity scale for the 20 amino acids. This was done and averaged for a set of over 56 different high-resolution x-ray structures of proteins and the result is reported in Fig. 7, *top left* and *top right*. It clearly shows, as expected, that charged amino acids (Asp, Glu, Lys, and Arg) have the highest values, while the hydrophobic ones (Leu, Ile, Val, Met, Trp, and Phe) have the lowest values. Calculating this density profile-derived hydrophobicity score with a slightly different definition using the integral of $|d\rho_{\text{radial}}/dr|^2$ (see below) does not change this ordering of amino acids, and merely rearranges the order of two pairs of intermediate amino acids (His, Ser) and (Asn, Gln). It is also possible to calculate a correlation coefficient between this set of hydrophobicity values with published ones, such as the one derived using crystallographic waters (71,72): a large value of 0.81 is obtained in this case. Similar high values of the correlation coefficients (0.82 and 0.80)

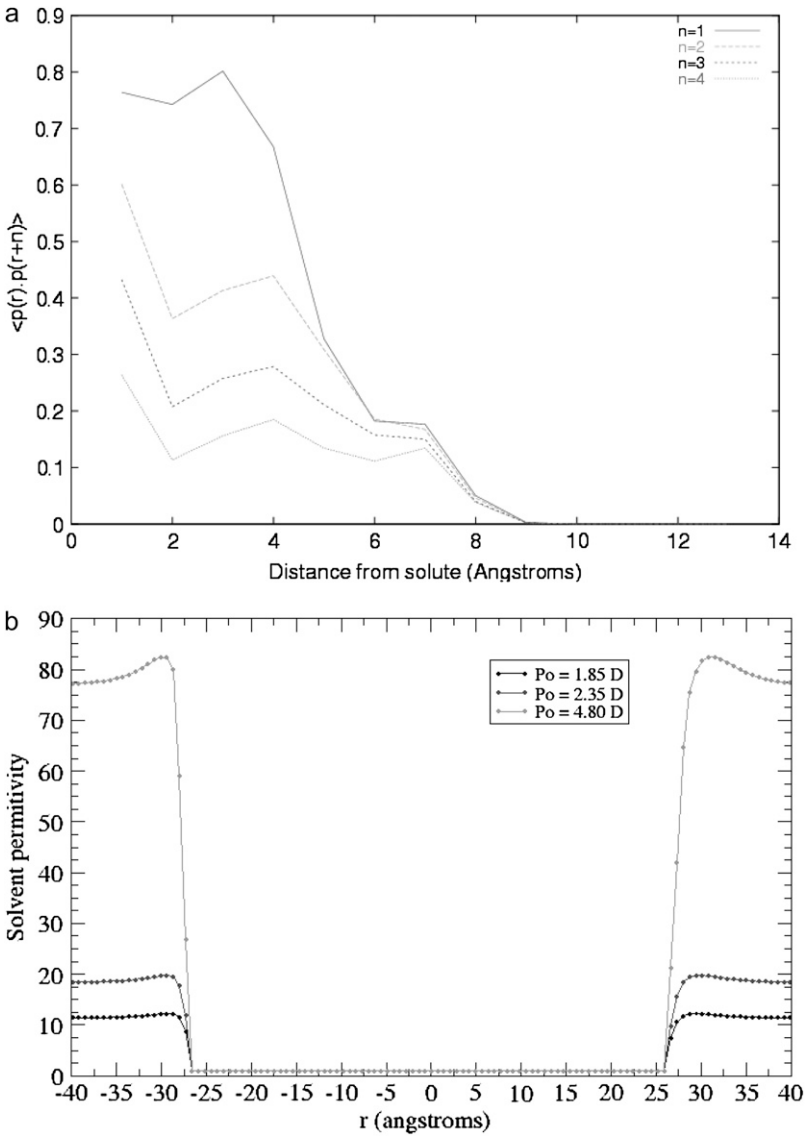

FIGURE 5 Dielectric properties of the solvent (2ACS). (*a*) Correlation through space of $\vec{P}(\vec{r})$ as a function of the distance away from the solute's surface, for lattice points separated by increasing distances ($|r - r'| = 1,2,3,4$ lattice points; in this case the grid spacing is 0.7 Å). (*b*) Profile of the dielectric function $\chi(\vec{r})$ in the solvent region as a function of the distance to the center of gravity of the solute for $p_0 = 1.85D$, $p_0 = 2.35D$, or $p_0 = 4.8D$.

were also obtained with other hydrophobicity scales derived from the experimental octanol/water transfer energies of a number of model compounds (73) or otherwise (74). In addition, there is very good agreement with the consensus qualitative ranking of Trinquier and Sanejouand (75).

At the atomic level, we can also define a solvation energy based on these solvation profiles, using the van der Waals theory of capillarity (37), which can be understood as a Landau expansion of the free energy in terms of analytical functions of the order parameter (solvent density) and its derivatives (32). This involves the integral of the square of the derivative of the density:

$$\mathcal{F}_1 = \frac{m}{2}\int d\vec{r}\,|\vec{\nabla}\rho(\vec{r})|^2, \tag{16}$$

where m is the second moment of the attractive potential between the particles, which is assumed to be independent of the density ρ.

In van der Waals capillarity theory, the surface tension reads, using the same m coefficient (37):

$$\sigma = m\int dr\left|\frac{d\rho_{\text{radial}}}{dr}\right|^2. \tag{17}$$

The mean value of this surface tension for different atoms can therefore be compared to the Eisenberg and McLachlan (61) atomic solvation parameters (ASP), for the same set of atoms. A correlation coefficient of 0.96 was obtained (Fig. 7, *bottom left* and *bottom right*). It is remarkable that the correlation coefficient should be so high, considering the fact that both sets of tension coefficients have been obtained by completely different methods. Indeed, Eisenberg and MacLachlan derived their ASP using experimental octanol/water transfer energies, which involve the difference between two solvation energies with two different solvents. We cannot calculate the octanol solvation energy with our model, because octanol is a

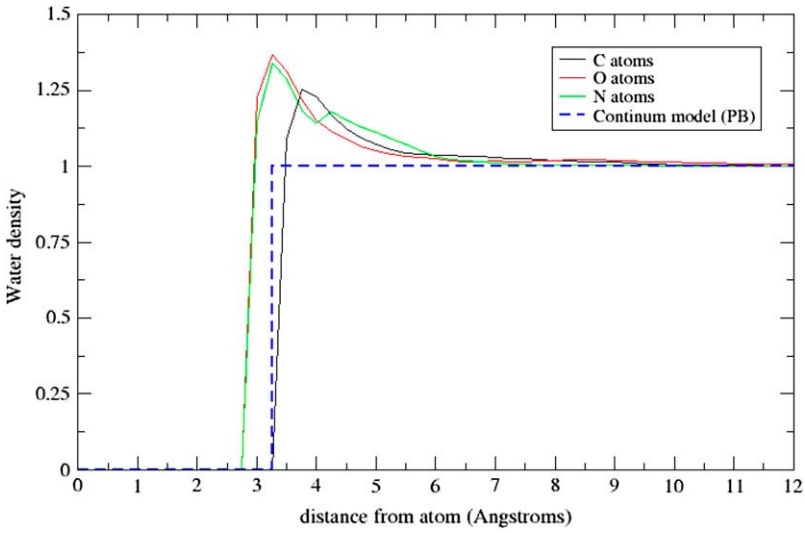

FIGURE 6 Radial solvent density profiles for different atom types at $p_0 = 1.85D$. This was obtained with a solute surface defined as the accessible surface area. Here the value $a/2 = 1.2$ Å was used; increasing this value to $a/2 = 1.4$ Å will lower the peaks of the solvent density, while increasing p_0 will increase them.

highly nonspherical molecule. We can only guess that this solvation energy is probably small compared to the water solvation energy. We repeated this calculation with different values of a and p_0 (Fig. 7 *bottom right*) and still get a very high correlation coefficient, but with a different slope and a different shift at the origin. We note that the $\mathcal{F}_1$ free energy can only be positive, given the way it is defined. This means that the tension coefficients derived from our methods are all negatives, whereas the Eisenberg and McLachlan's coefficients are positive for apolar atoms and negative for polar (or charged) atoms. We reconcile this data by pointing out that the reference energy of the two methods need not be the same.

A nonuniform hydration shell

The solvent density map can then be used to place a collection of water molecules around the solute macromolecule. The derived water molecule distribution will depart markedly from a uniform distribution around the solute, because the PBL solvent density depends strongly on the chemical nature of the exposed solute atoms nearby (Fig. 6). We expect that this set of assigned water molecules will be useful for a better fit of atomic models with experimental data (76).

To determine how many water molecules should be placed, several protocols are possible. The simplest one is to use a hydration level, h, that gives the number of grams (g) of H_2O per g of protein. Microwave dielectric studies of lysozyme indicate an h value of ~0.27 g/g for which the first hydration shell is complete (77). Several experimental techniques indicate that this h value should be taken at ~0.4 g/g (78,79), although there is considerable variation around this value in the literature, for different proteins. A more recent review summarizing results obtained with microwave dielectric studies indicates a range of 0.34–0.39 g/g (80). Working backward to see what dipole density ρ_{thres} should be used to place the number of water molecules dictated by $h =$ 0.4 g/g, one sees, on a group of 56 proteins, that there seems to be no universal value for ρ_{thres}. On closer inspection, there seems to be essentially two subgroups of proteins, with ρ_{thres} around either 1.2 ± 0.15 or 1.5 ± 0.15 (81).

To resolve this issue we performed first a preliminary survey on a single protein, counting the number of placed water molecules for all possible threshold values of the solvent density. The curve is sigmoidal and its derivative was calculated to find the mid-transition point (Fig. 8 *a*). Although somewhat noisy, this derivative curve can be fitted to a Gaussian curve. Its maximum was found to be located close to $\rho_{\text{max}} = 1.5$. Surprisingly, this was found to be always the case for all proteins examined, with very little variation. The number of placed water molecules by different methods for a set of proteins of increasing size is plotted in Fig. 8 *b*. A possible conclusion is that the threshold value for the density, set to 1.5, is a more universal way to estimate the number of water molecules around a protein, rather than the h value, which varies a lot in the literature (0.4 ± 0.2 g/g).

Yet another way to estimate the number of water molecules to be placed is simply to calculate the excess dipole density

$$N_{\text{w}} = \int d\vec{r}(\rho_{\text{dip}}(\vec{r}) - \rho_{\text{bulk}}). \tag{18}$$

The integral is over all grid points and $\rho_{\text{bulk}} = 0.033/\text{Å}^3$. Using $R_{\text{probe}} = 1.5$ Å to calculate the accessible surface, one recovers a number of water molecules very similar to what is obtained with other methods (Fig. 8 *b*).

This last method is probably the most objective one, allowing us to look at the influence of temperature on the number of condensed water molecules. We found a linear increase of 0.5 water molecule per K lowered, in the range 150–350 K. This is the expected trend but the model obviously lacks the level of detail, namely the directionality of hydrogen bonds, needed to fully address this problem. The linear fit of the average number of water molecules per residue in Fig. 8 *b* gives 0.33 ± 0.02 g of H_2O per g of protein, in excellent agreement with a recent review (80).

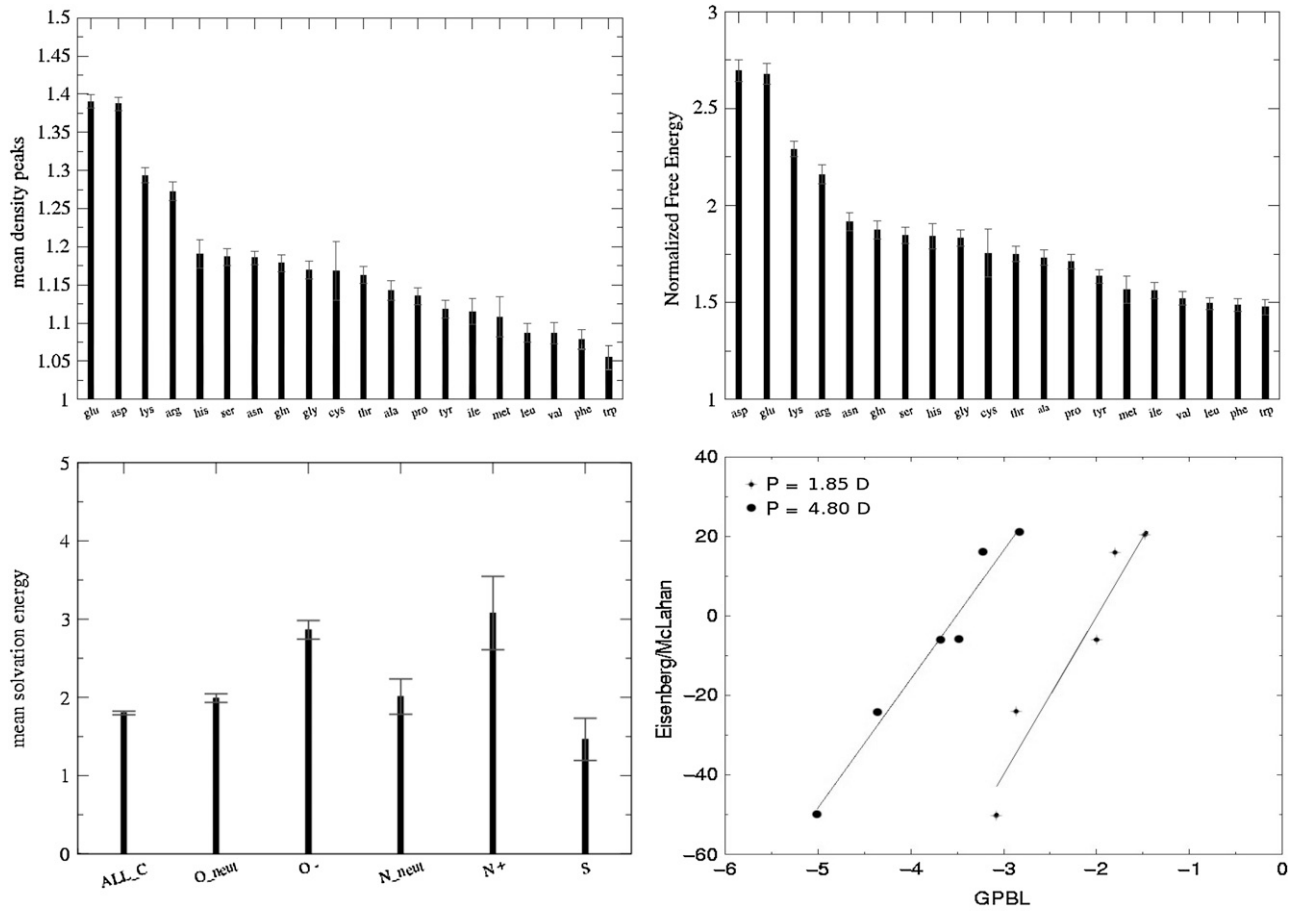

FIGURE 7 Hydrophobicity scales. (*Top*) Density profile-derived hydrophobicity scales for the 20 amino acids using either the maximum value of the curve (*left*) or the integral of its squared derivative (*right*). (*Bottom*) Mean values of the atomic surface tension coefficients with standard deviations at $p_0 = 1.85D$ (*left*) and comparison with the Eisenberg and MacLachlan (61) surface tension coefficients (*right*). In the latter case, both the $p_0 = 1.85D$ or $p_0 = 4.8D$ curves are shown.

Having determined the solvent density threshold value, it is possible to color the molecular (or accessible) surface of the solute according to the maximum value of the solvent density next to each surface point using a color code: red for a value above the threshold and blue for a value below the threshold. This makes the visualization of poorly solvated regions easy as large blue patches on the surface of the (otherwise red) solute. This works well not only for membrane proteins but also for homodimeric proteins, where the buried area upon dimerization closely matches patches of poorly solvated accessible surface (SAS) in >80% of the cases studied (42,81).

Solvent mixtures

Three different calculations were undertaken with mixtures of dipolar solvents, which can be studied within the same theoretical framework (see Appendix B). One with $p_1 \ll p_2$, one with $p_1 = p_2$, and one with $p_1 \gg p_2$, keeping their relative concentrations constant, namely $c_1 = 55$ M and $c_2 = c_1/100$. Qualitatively, the results can be summarized as follows: strong dipoles will cluster near the solute, in regions of high electric field. Weak dipoles will be distributed randomly both in the bulk and near the solute. If $p = 0$, there is no spatial preference for the dipoles which are distributed randomly in the solute-excluded volume. This is very much in agreement with MD studies of proteins in organic polar solvent(s)/water mixtures (82,83). A corollary to this observation is that it does not seem possible to explain the segregation of hydrophobic molecules such as lipids out of water with this model; rather, we believe that an explicit additional interaction energy based on Flory's theory between apolar solvent molecules is necessary to reproduce this effect (84). Also, the model is not good for nonspherical (elongated) solvents, although it does allow for a different size of the second (diluted) solvent (see Appendix B).

Discriminating the right structure in an ensemble of decoys

We have used the partial solvation energy $\mathcal{F}_1$ given by Eq. 16 to evaluate a number of decoys of a given protein.

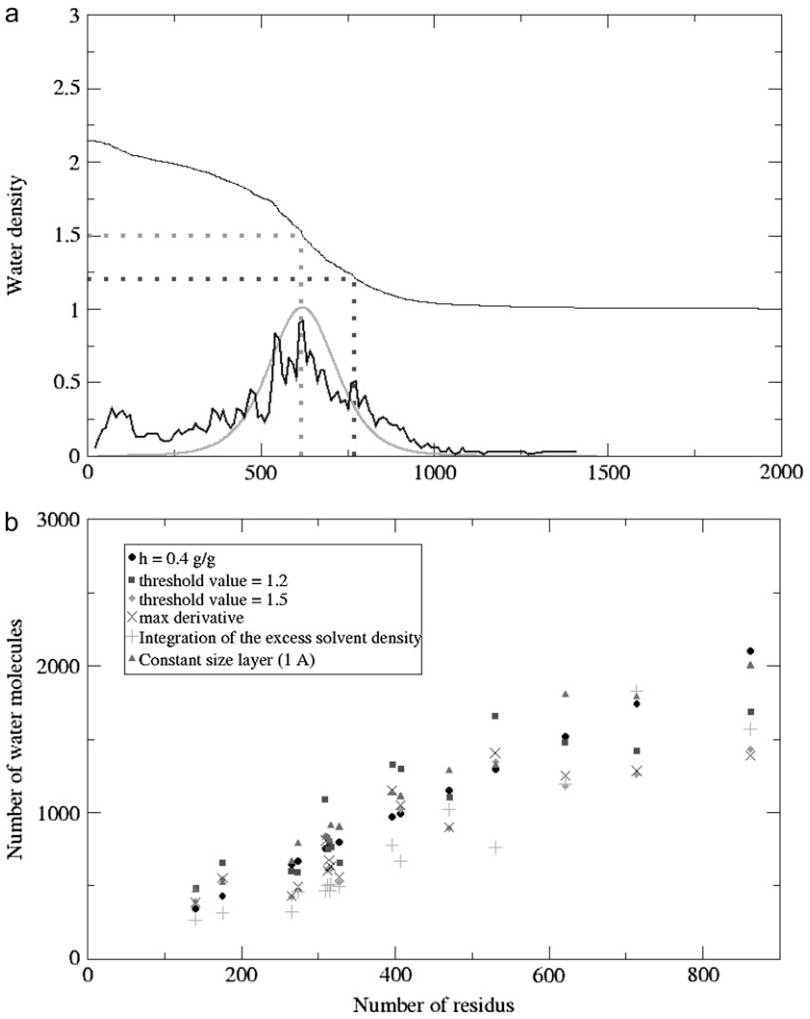

FIGURE 8 Condensed water molecules in the first hydration shell. (*a*) Number of water molecules placed, as a function of the density threshold (*upper curve*). The lower curve is the derivative of the upper curve together with a tentative fit of the lower curve with a Gaussian function. (*b*) Number of water molecules placed using different methods: a constant hydration value ($h = 0.4$ g/g), a density threshold criterion (1.2 or 1.5 or the density set to the maximum derivative of the $N_w = f(\rho_{thres})$ curve in (*a*)), integration of the excess density and a layer of constant size (1 Å) around the solute (null hypothesis). The dependence on the size (number of amino acids) of the solute proteins is shown.

The first set of decoys was generated for lysozyme (1LZR) by decorating the wild-type backbone with all 129 possible circular permutations of the sequence. The side-chain configurations was optimized using the self-consistent mean-field algorithm (50). As a control, the Environ energy was evaluated (63), which clearly identifies the wild-type structure as the one with the lowest score. Two types of solvation energy were considered. The first one is derived from the radial density profile and its derivative, as in the van der Waals capillarity theory (see Eq. 16). The other one is based on the fraction of SAS that is poorly solvated, namely with a surrounding solvent density less than a given value set to 1.2. The results are presented in Fig. 9, *a* and *b*, and show that the wild-type structure does indeed possess the best score in both cases. In a subsequent step (Fig. 9 *c*), we defined a generalized Z-score as the sum, for every accessible atom, of the surface tension coefficient γ_i, compared to the mean value $\langle\gamma_i\rangle$ for this atom type, in units of standard deviation σ (γ_i). This performs slightly better than the two other scores for lysozyme decoys (Z-score = 4).

Considering the fact that calculating these energies did not require any parameter fitting, the results look indeed promising.

In the second experiment, a published set of 630 decoys was used (62) for the C-terminal domain of L7/L12 ribosomal protein (PDB code 1CTF). The same kind of results were obtained, but in this case the true structure only shows a score 1.6σ above the mean (Fig. 10 *a*). There is a better discrimination for the score derived from $\mathcal{F}_1$ (Eq. 16), compared to the score derived from the poorly solvated SAS (Z-score = 1.1; and other data not shown).

Electrostatics free energy also performs reasonably well in this case (Z-score = 1.6), but this is also true for the null hypothesis: PB free energy with $\epsilon_{in} = 2$ and $\epsilon_{out} = 78$ lead to a Z-score of 1.8 for the same decoy set.

However, there is a clear tendency of higher scores in $\mathcal{F}_1$ for decoys with decreasing root mean-square deviations (RMSDs) with the native structure (Fig. 10 *b*), with a correlation coefficient of -0.34; this is not the case with the electrostatic free energy. Altogether, the results for the 1CTF decoys show that our solvation energies can enrich a large

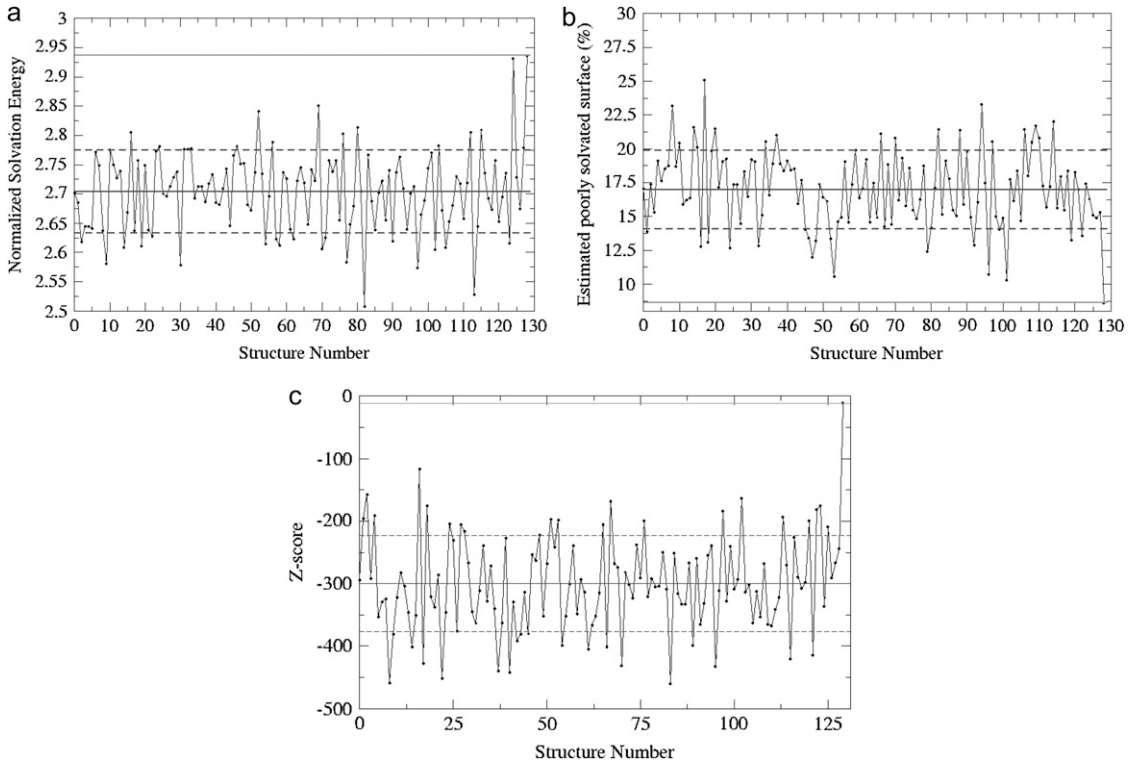

FIGURE 9 Lysozyme decoys derived from circular permutations of the sequence. (*a*) Gradient-based partial solvent energy $\mathcal{F}_1$. (*b*) Solvent energy based on the proportion of poorly solvated SAS. (*c*) Z-scores derived from local atomic surface tension coefficients. In all cases the correct structure is the last one and its energy is indicated by a continuous line. The mean value of the energy is materialized by a continuous line bracketed by two dotted lines (± 1 SD).

starting population of decoys into a smaller population containing the true structure but do not always identify the native structure as the one with the lowest energy (highest Z-score). Similar results were obtained with other decoys from Park and Levitt (62) (data not shown).

Comparison with MD simulations

A number of different molecular dynamics simulations were performed on the protein 1TIM, using a box of 73.5 Å on each side on the SPC/E water model. The solute was restrained to an almost fixed conformation by a very large elastic constant. The simulation box was divided into a grid and the water density was calculated for each 8-ns-long trajectory. The protocol was similar to the one described in Lounnas et al. (28), Lounnas and Pettitt (85,86), Higo and Nakasako (66), and Higo et al. (67). Convergence of the MD simulations was assessed in the following way: a correlation coefficient was calculated between the density maps derived from three different independent simulations, starting with different random velocities of the atoms in the box. Therefore, it is legitimate, under the ergodicity hypothesis, to compare the resulting solvent densities to the PBL solvent densities, which are thermodynamic averages taken over all dipoles and ions configurations.

For grids with grid steps of either 1.4 or 2.1 Å, the simulations had clearly converged, with correlation coefficients between the three simulations of 0.95 or 0.99, respectively. The average density map for each individual starting conformation of the solute was compared to the one obtained using PBL by calculating a direct correlation coefficient. The latter was found to be 0.63 for the whole map and 0.32 when restricted to the solvent region, when the grid size was set to 2.1 Å. The solvent correlation coefficient is only 0.26 for a grid size of 1.4 Å. It is quite obvious by visual inspection that the two types of maps (MD and PBL) are quite different: the MD map contains many well-defined peaks, whereas the PBL map is very smooth, without too many bumps. This means that our solvent model is in fact a rather coarse-grained water model, similar to other models studied by Makarov et al (87) and Pettitt et al. (88), with which it actually compares well.

Clearly the PBL solvent map is compatible with, but lacks the level of detail of, MD maps. In particular it is unable to pick up networks of water molecules sharing at least one hydrogen bond. This is not really surprising as it contains no information on the directionality of the hydrogen bonds that water molecules can make. Indeed the potential energy underlying MD simulations is much more detailed than PBL.

It should be noted that if we were to compute a solvent density map from the solution of the standard PB equation, we would not observe any correlation at all with the solvent maps derived from MD simulations. This is due to the fact that the electric field obtained by PBE in the solvent region is

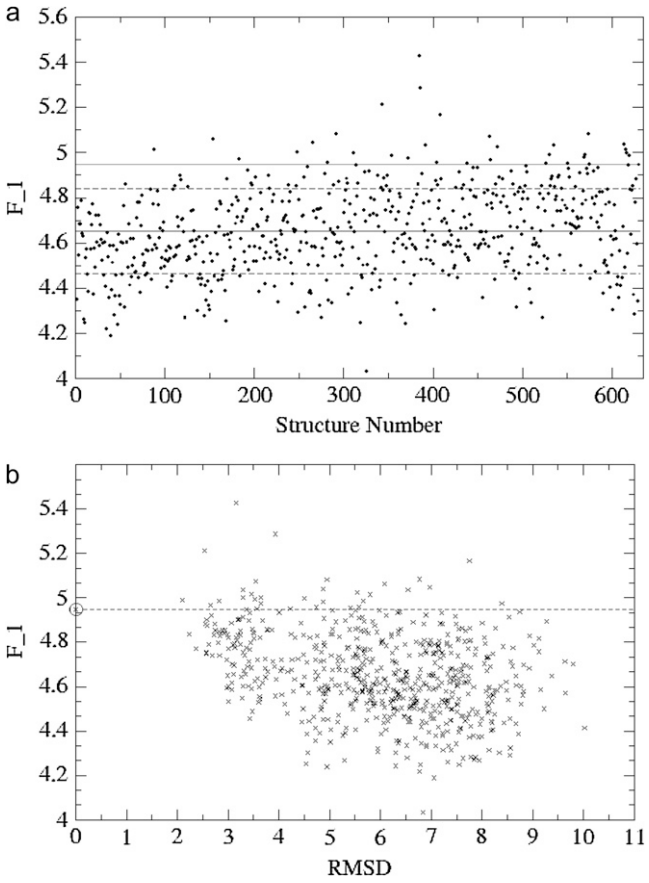

FIGURE 10 1CTF decoys. (*a*) Gradient-based partial solvent energy $\mathcal{F}_1$ as a function of the decoy structure number. The correct structure is the last one and its energy is indicated by a continuous line. The mean value of the energy is materialized by a continuous line bracketed by two dotted lines (± 1 SD). (*b*) Partial solvent energy $\mathcal{F}_1$ as a function of the RMSD with the correct model. The native structure energy (RMSD = 0, *circle*) is shown with a dotted line.

very weak, as described in Fig. 3 *c*, leading to $u_{\mathrm{PB}} \ll 1$ and $\rho(\vec{r}) = \rho_{\mathrm{bulk}}$.

Comparison with crystallographic waters

In this experiment, the PBL solvent density for all the crystallographic water molecules of a very high-resolution structure (aldose reductase, 1USO, Resolution 0.67 Å) was calculated by interpolating the PBL solvent density map. A global score, taken as the sum of these interpolated densities at the water positions, was then compared to the score of a set of 10,000 random distributions of the same number of water molecules, randomly generated within the solvent-allowed space and with a maximum distance from the original position of 5 Å. This involves 613 water molecules. A Z-score of 2.4σ above the average was obtained for the crystallographic waters, with $p_0 = 1.85D$, $a/2 = 1.5$ Å, and $R_{\mathrm{probe}} = 1.5$ Å. The null hypothesis with electric potential derived from the GB model (APBS) leads to a Z-score close to 0, as the solvent density derived from $\Phi_{\mathrm{PB}}(\vec{r})$ is flat.

However, the comparison of a thermodynamic average of the solvent density with crystallographic water molecules obtained in a frozen solvent at 100 K may be somewhat artificial; indeed Halle (89) has argued that freezing crystals may trap conformations that are not the thermodynamic free energy minimum. In addition, we note that even MD simulations do not agree in detail with the experimental cryogenic crystallographic studies either (66,67,90).

DISCUSSION

Limitations of the model

A generalized Poisson-Boltzmann equation has been implemented, accounting for solvent effects as explicit spherical Langevin dipoles placed on a grid around the solute. It generates in a self-consistent way a continuous and position-dependent dielectric value in the solvent region, a feature that has long been recognized to be important for solvation energies and the calculation of electrostatic forces. However, it is clear that the dielectric constant in the bulk using the physical value of the dipole moment of water ($p_0 = 2.35D$) is far from the expected experimental value (see Fig. 5 *b*).

This is actually a known problem in mean field theories of dipolar solvent such as mean spherical approximation (69), along with the absence of a second hydration shell in the solvent density profiles. Only the first hydration shell can be reproduced with this model, in the absence of a correct treatment of the correlations among solvent molecules. A possible cure to this last problem might be envisioned by systematic expansion of the partition function (91) but its implementation for the moment seems quite impracticable for biological macromolecules. An obvious way to get a better value for ϵ_{bulk} would be to increase the value of the dipolar moment of the solvent, as ϵ_{bulk} varies like $\epsilon_0 + c_{\mathrm{dip}}\beta p_0^2/3$. From this, it can be calculated that a value of $p_0 = 4.86D$ would be needed to reproduce $\epsilon_{\mathrm{bulk}} = 78$.

However, extensive calibration tests made on the solvation energy of simple ions and other simple organic and small biological solutes have convinced us to stick with the physical value of the dipolar moment of water 1.85–2.35D (81). Also, increasing p_0 would not qualitatively change the nature of the solvent density maps. The route to get a more realistic value of ϵ_{bulk} is well known and involves taking into account the geometry and directionality of the two hydrogen bonds that a water molecule can make, as worked out originally by Fröhlich (68). Indeed a more sophisticated model of the water molecule would be a superposition of a dipole and a quadrupole, the latter being positioned perpendicular to the dipole with its end points in coincidence with the two hydrogen atoms. In this formalism, this would involve adding an anisotropic tensor for quadrupoles $Q_{\alpha,\beta}$ in the free energy (70). It is also conceivable to even include octupoles as in some recent water models (92), but this rapidly results in rather involved algebra.

Another route is possible. Indeed, recent spherical models of water appear to be able to capture most of the waterlike properties of a solvent, including the correct value of ϵ_{bulk}, provided that an attractive short-range solvent-solvent potential is taken into account (93).

Related work and formalism

The formalism itself bears strong resemblance to the one developed by Coalson and Duncan (70), as well as the more recent work by Ramirez and Borgis (46) and Ramirez et al. (94). Although we were unaware of these articles when this work was started, our results turn out to be fairly similar to theirs (70). However, our treatment of steric exclusion of the different species in solution is different and its effects on the analytical expressions can clearly be seen in the $D(\Phi(\vec{r}))$ denominator term (Eq. 3), which prevents solvent density values to go beyond the physical limit of close packing. In contrast, Coalson and Duncan (70) introduce an extra potential for steric exclusion, rendering the numerical solution of the problem more complicated. Clearly, the main drive of our work is to show that the model is valid for large biological macromolecules of any size and shape in a reasonable CPU time and make it available to the biological community (42).

The work of Ramirez and Borgis (46) is also very similar in spirit to ours but relies on a different formalism based on density functionals (see also (95)). They do treat the correlations among solvent molecules but do not include the free ions. They need to run a number of preliminary MD simulations to derive the solvent-solvent correlation function, which is then taken into account within their density functional formalism. Similar in spirit is the earlier work of the same group on a semiimplicit model of the solvent made of polarizable pseudo-particles (PPP model) (96). Here the pseudo-particles can move and are no longer fixed on a lattice. It was shown recently that it is possible to make it work with a polarizable force field of the solute and to run stable 20-ns-long MD simulations (44). The authors point out that the reaction field is underestimated by at least a factor of two-thirds and solve the problem by carefully calibrating hydration energy of small solutes through rescaling the set of partial charges. We also plan to apply our method to the emerging polarizable force fields for protein solutes in the next future.

Beglov and Roux examined a model closely related to ours, including effects due to solvent-solvent interactions through the HNC equation (19). We note that their equations of $\rho_{dip}(\vec{r})$ and $\vec{P}(\vec{r})$ are the same as ours if correlations are ignored and if we set $D(\Phi(\vec{r})) = 1$ (20).

Imai et al. (97,98) have recently developed a 3D-RISM algorithm that performs very well for predicting the hydration properties of macromolecules. Indeed, they found a way to tackle the problematic application of HNC approximation in the presence of an impenetrable solute. However, although benefiting from powerful FFT techniques, the method still needs hundreds of cycles for convergence to be achieved. Although there are attempts to couple this method to MD simulations (99), we are not aware of any large-scale application of this method to biological macromolecules.

pH effects

It is possible to take into account the pH effect in very much the same way as the one described in recent publications (100,101). In this case the charge density of the solute ρ_f is modified and is now a function of both pH and $\Phi(\vec{r})$. It is then possible to introduce this new ρ_f into Eq. 1 to get a new Eq. 4 that is now pH-dependent. A set of calculations at different values of pH has to be run to derive the local pK_a of the titratable amino acids, the pH at which the partial charge is 0.5. It is anticipated that such calculations would lead to better estimates of intrinsic pK_as of buried ionizable groups (102,103), which can be compared to experimental values measured by NMR and are often part of catalytic sites; this has implications for the prediction of active sites from the structure alone in structural genomics projects (104).

The question arises as to how pH effects are coupled to side-chain flexibility (102) and in principle discrete movements of charged side chains (rotamers) responding to pH changes could be explored by Monte Carlo methods (103). Because the expected changes in the electric potential are small, a good guess can be made about this function and less iterations will be necessary to find the solution at each Monte Carlo step.

PERSPECTIVES AND CONCLUSION

There are actually two classical problems with PB theory and its derivatives: 1), the solute is fixed; and 2), correlation effects are difficult to include, as it is essentially a mean-field theory. Attempts to include these effects can be made if the interacting species are rather dilute (21,105) but the problem becomes more difficult if this approximation is not valid. In our methodology, there is a way to include molecular flexibility of the solute that is indirect and involves normal modes for representing collective and large-amplitude movements (106,107). We have actually tested this procedure using normal modes derived from the elastic network model (108) and gathered preliminary results showing that the results displayed in Fig. 3 are qualitatively conserved.

To summarize the results presented here, we find that the PBL Model provides a physically sound solvent density that is coherent with what is expected from purely physicochemical grounds. Indeed, it allows us to derive an hydrophobicity scale for amino acids that is highly correlated with known ones derived from completely different techniques, both on the amino-acid level and on the atomic level. The number of condensed water molecules is also within the range of what is expected and close to what other methods predict (76,109). It gives a first hydration layer that is con-

sistent with what is known from experimental SAXS and SANS studies (29,110) and solvation energies derived from it give excellent discrimination of the native structure against decoys obtained by circular permutation of the sequence. The results on a 1CTF decoy set, however, are not as high as expected and obtained, for instance, in knowledge-based methods. In view of these results, we conclude that there is a need to take into account the solvent-solvent correlations in our model.

In this respect, we have recently started to use a Yukawa attractive effective potential between solvent molecules to reproduce the second hydration shell, similar in spirit to Coalson et al. (111) but still projected onto a lattice. The full results will be reported elsewhere.

APPENDIX A: PHENOMENOLOGICAL DERIVATION OF THE FREE ENERGY

The partition function Z of the system described in Fig. 1 can be evaluated analytically using field-theoretic methods, from which the free energy can be derived as $\mathcal{F} = -k_B T \log(Z)$ (43). All equilibrium thermodynamic quantities can subsequently be derived using either Z or $\mathcal{F}$. Here we derive the free energy using a phenomenological approach, following the formalism of Borukhov et al. (45).

The free energy is the difference of two terms, the electrostatic energy on one hand and the product of entropy with temperature, on the other hand:

$$\mathcal{F} = U_{el} - TS. \tag{19}$$

The electrostatic energy is given by

$$\begin{aligned} U_{el} = \int d\vec{r} \Big\{ & -\frac{\epsilon_0}{2}|\vec{\nabla}\Phi(\vec{r})|^2 + \rho_f(\vec{r})\Phi(\vec{r}) + zeC_+(\vec{r})\Phi(\vec{r}) \\ & - zeC_-(\vec{r})\Phi(\vec{r}) + C_{dip}(\vec{r}) \int d\vec{p}(\Omega) A(\vec{r},\Omega)(\vec{p}(\Omega).\vec{\nabla}\Phi(\vec{r})) \\ & - \mu_+ C_+(\vec{r}) - \mu_- C_-(\vec{r}) - \mu_{dip} C_{dip}(\vec{r}) \Big\}, \end{aligned} \tag{20}$$

where $\Omega = (\theta, \phi)$ denotes the direction of the dipole $\vec{p}$ with respect to the electric field $\vec{E}(\vec{r})$, and $A(\vec{r}, \Omega)$ is the normalized Boltzmann distribution of a dipole with electrostatic energy $-\vec{p} \cdot \vec{E}$ (94),

$$A(\vec{r},\Omega) = \frac{e^{\beta p_o |\vec{\nabla}\Phi(\vec{r})|\cos(\theta)}}{\mathrm{sinhc}(\beta p_o |\vec{\nabla}\Phi(\vec{r})|)}, \tag{21}$$

where $p_0 = |\vec{p}|$, $\mathrm{sinhc}(u) = \sinh(u)/u$, $\beta = 1/k_B T$, and k_B is the Boltzmann constant.

The first term in U_{el} is the self-energy of the electric field, the next four terms are all of the form $\rho^{elec}\Phi$ (for a dipole $\vec{p}$, at first order, $\rho^{elec}_{dip} = \vec{p}.\vec{\nabla}$) and the last three terms couple the free ions and dipoles to a reservoir through their chemical potentials, namely μ_+, μ_-, and μ_{dip}. C_+, C_-, and C_{dip} are the concentrations of the free ions and dipoles, respectively.

The entropy includes both a translational and a rotational contribution,

$$\begin{aligned} S_{trans} = -\frac{k_B}{a^3} \int d\vec{r} \Big\{ & C_+(\vec{r})a^3 \ln(C_+(\vec{r})a^3) + C_-(\vec{r})a^3 \ln(C_-(\vec{r})a^3) + C_{dip}(\vec{r})a^3 \ln(C_{dip}(\vec{r})a^3) + (1 - C_+(\vec{r})a^3 - C_-(\vec{r})a^3 \\ & - C_{dip}(\vec{r})a^3) \ln(1 - C_+(\vec{r})a^3 - C_-(\vec{r})a^3 - C_{dip}(\vec{r})a^3) \Big\} \end{aligned} \tag{22}$$

where a is the lattice size,

$$S_{rot} = -\frac{k_B}{a^3} \int d\vec{r} C_{dip}(\vec{r}) a^3 \int d\Omega A(\vec{r},\Omega) ln(A(\vec{r},\Omega)). \tag{23}$$

The four terms in S_{trans} represent the entropy of positively charged free ions, negatively charged free ions, dipoles, and vacant sites, respectively (45) while S_{rot} represents the orientational entropy of the dipoles and can be further transformed using Eq. 21.

Setting $\lambda_{ion} = \exp(\beta\mu_{ion})$ and $\lambda_{dip} = \exp(\beta\mu_{dip})$ and proceeding as in Borukhov et al. (45) by imposing

$$\frac{\partial \mathcal{F}}{\partial C_+(\vec{r})} = \frac{\partial \mathcal{F}}{\partial C_-(\vec{r})} = \frac{\partial \mathcal{F}}{\partial C_{dip}(\vec{r})} = 0, \tag{24}$$

we get (42)

$$\begin{aligned} \beta\mathcal{F} = & -\frac{\beta}{2} \int d\vec{r} \epsilon_0 |\vec{\nabla}\Phi(\vec{r})|^2 + \beta \int d\vec{r} \rho_f(\vec{r}) \Phi(\vec{r}) \\ & - \frac{1}{a^3} \int d\vec{r} \ln \Bigg(1 + 2\lambda_{ion} \cosh(\beta e z \Phi(\vec{r})) \\ & + \lambda_{dip} \frac{\sinh(\beta p_o |\vec{\nabla}\Phi(\vec{r})|)}{\beta p_o |\vec{\nabla}\Phi(\vec{r})|} \Bigg). \end{aligned} \tag{25}$$

If $\lambda_{dip} = 0$, we recover the free energy leading to the classical PBE with site exclusion effects (17,18,45).

It is possible to take into account explicitly the van der Waals interaction $W(\vec{r})$ between the solute and the free ions and dipoles by simply adding a term $\int d\vec{r}(C_+(\vec{r}) + C_-(\vec{r}) + C_{dip}(\vec{r}))W(\vec{r})$ to the free energy. This amounts to having position-dependent chemical potentials, namely $\mu_{ion} - W(\vec{r})$ and $\mu_{dip} - W(\vec{r})$, respectively. $W(\vec{r})$ is due to the fixed solute and needs to be calculated only once.

APPENDIX B: POSSIBLE EXTENSIONS OF THE PBL MODEL

The model of Fig. 1 can be extended in a number of ways:

1. The model can of course accommodate a mixture of free ions of different valency, such as $MgCl_2$, for instance. Only the term $\cosh(\beta e z \Phi)$ in the free energy is affected (see, for example, (45)),

$$\begin{aligned} \beta\mathcal{F} = & -\frac{\beta}{2} \int d\vec{r} \epsilon_0 |\vec{\nabla}\Phi(\vec{r})|^2 + \beta \int d\vec{r} \rho_f(\vec{r}) \Phi(\vec{r}) \\ & - \frac{1}{a^3} \int d\vec{r} \ln \Bigg(1 + \lambda_{dip} \frac{\sinh(\beta p_0 |\vec{\nabla}\Phi(\vec{r})|)}{\beta p_0 |\vec{\nabla}\Phi(\vec{r})|} \\ & + \sum_{i=1,2} \lambda_i^{ion} \exp(-\beta e z_i \Phi(\vec{r})) \Bigg). \end{aligned}$$

The fugacities λ_1^{ion} and λ_2^{ion} must be also recalculated accordingly. This has been implemented in our website (42).

2. Mixtures of solvents of the same size but different dipole moments p_1 and p_2 can also be treated. The free energy then reads

$$\beta\mathcal{F} = -\frac{\beta}{2}\int d\vec{r}\,\epsilon_0|\vec{\nabla}\Phi(\vec{r})|^2 + \beta\int d\vec{r}\rho_f(\vec{r})\Phi(\vec{r}) - \frac{1}{a^3}\int d\vec{r}\,\ln\left(1 + 2\lambda_{ion}\cosh(\beta ez\Phi(\vec{r})) + \sum_{i=1,2}\lambda_{dip_i}\frac{\sinh(\beta p_i|\vec{\nabla}\Phi(\vec{r})|)}{\beta p_i|\vec{\nabla}\Phi(\vec{r})|}\right).$$

3. To treat the case of a mixture of dipolar solvents with different sizes, we resort to an approximation of S_{trans}, which is valid if the second solvent p_2 is diluted compared to the first one (p_1, water, 55M). Setting k such that ka^3 is the volume of p_2 ($k < 1$) and a^3 the volume of p_1, we find (42)

$$\beta\mathcal{F} = -\frac{\beta}{2}\int d\vec{r}\epsilon_0|\vec{\nabla}\Phi(\vec{r})|^2 + \beta\int d\vec{r}\rho_f(\vec{r})\Phi(\vec{r}) - \frac{1}{a^3}\int d\vec{r}\,\ln\left\{\left(1 + \lambda_{dip2}\frac{\sinh(\beta p_2|\vec{\nabla}\Phi(\vec{r})|)}{\beta p_2|\vec{\nabla}\Phi(\vec{r})|}\right)^{\frac{1}{k}} + \lambda_{dip1}\frac{\sinh(\beta p_1|\vec{\nabla}\Phi(\vec{r})|)}{\beta p_1|\vec{\nabla}\Phi(\vec{r})|}\right\}.$$

4. We also give the free energy for a system containing free ions of smaller size than dipoles. Setting k such that ka^3 is the volume of ions ($k < 1$) and a^3 the volume of p_0, we find (18,42)

$$\beta\mathcal{F} = -\frac{\beta}{2}\int d\vec{r}\epsilon_0|\vec{\nabla}\Phi(\vec{r})|^2 + \beta\int d\vec{r}\rho_f(\vec{r})\Phi(\vec{r}) - \frac{1}{a^3}\int d\vec{r}\,\ln\left\{(1 + 2\lambda_{ion}\cosh(\beta ez\Phi(\vec{r})))^{\frac{1}{k}} + \lambda_{dip}\frac{\sinh(\beta p_0|\vec{\nabla}\Phi(\vec{r})|)}{\beta p_0|\vec{\nabla}\Phi(\vec{r})|}\right\}.$$

5. Finally, we give the free energy for the case where the dipoles have an intrinsic polarizability α_0. In this case, the degrees of freedom of not only the direction (as before) but also the magnitude of the dipole vectors should be integrated out, which can be assumed to be constrained to fluctuate around their mean position in a harmonic potential (70). Here we set $\alpha_s = \beta p_0^2/3$, with $\alpha_s/\alpha_0 \sim 20$ numerically (48),

$$\beta\mathcal{F} = -\frac{\beta}{2}\int d\vec{r}\epsilon_0|\vec{\nabla}\Phi(\vec{r})|^2 + \beta\int d\vec{r}\rho_f(\vec{r})\Phi(\vec{r}) - \frac{1}{a^3}\int d\vec{r}\ln\left\{1 + 2\lambda_{ion}\cosh(\beta ez\Phi(\vec{r})) + \lambda_{dip}\left(\frac{\sinh(u)}{u} + \frac{\alpha_0}{3\alpha_s}\cosh(u)\right)e^{\frac{\alpha_0 u^2}{6\alpha_s}}\right\}.$$

In most cases we have $1 \ll \lambda_{dip}\sinh(u)/u$, where $u = \beta p_0 E$, and $\lambda_{ion} \ll 1$. A Taylor expansion of the last term $\ln(1+\ldots)$ can be done at low field.

Marc Delarue and Henri Orland thank the hospitality of KIPT at University of California at Santa Barbara, California, where this work was initiated. We thank Marc Baudoin and his team at Institute Pasteur for allowing us to use CPU-intensive common resources there, R. Navaza for help with maintaining the computers, and Daniel Borgis for comments on the manuscript.

We acknowledge support from Apple Inc., through their ARTS program (2007), and P.K. acknowledges support from the National Institutes of Health through grant No. GM080399.

REFERENCES

1. Orozco, M., and J. Luque. 2000. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* 100:4187–4226.
2. Perutz, M. 1976. Electrostatic effects in proteins. *Science*. 201:1187–1191.
3. Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science*. 268:1144–1149.
4. Marcus, R. 2006. Enzymatic catalysis and transfers in solution. *J. Chem. Phys*. 125:194504–194524.
5. Simonson, T., and B. Roux. 1999. Implicit solvent models. *Biophys. Chem*. 78:1–20.
6. Warwicker, J., and H. Watson. 1982. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol*. 157:671–679.
7. Gilson, M., A. Rashin, R. Fine, and B. Honig. 1985. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol*. 184:503–516.
8. Gilson, M., K. Sharp, and B. Honig. 1987. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem*. 9:327–335.
9. Nicholls, A., and B. Honig. 1991. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem*. 12:435–445.
10. Rocchia, W., E. Alexov, and B. Honig. 2001. Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B*. 105: 6507–6514.
11. Baker, N. A. 2005. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol*. 15:137–143.
12. Koehl, P. 2006. Electrostatics calculations: latest methodological advances. *Curr. Opin. Struct. Biol*. 16:142–151.
13. Holst, M. 1993. Multilevel methods for the Poisson-Boltzmann equation. PhD thesis, University of Illinois at Urbana-Champaign, IL.
14. Holst, M., R. Kozack, F. Saied, and S. Subramanian. 1994. Treatment of electrostatic effects in proteins: multigrid-based Newton iterative method for solution of the full nonlinear Poisson-Boltzmann equation. *Proteins Struct. Funct. Genet*. 18:231–245.
15. Holst, M. J., and F. Saied. 1995. Numerical solution of the nonlinear Poisson-Boltzmann equation: developing more robust and efficient methods. *J. Comput. Chem*. 16:337–364.
16. Holst, M., N. A. Baker, and F. Wang. 2000. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation. I. Algorithms and examples. *J. Comput. Chem*. 21:1319–1342.
17. Borukhov, I., D. Andelman, and H. Orland. 1997. Steric effects in electrolytes: a modified Poisson-Boltzmann equation. *Phys. Rev. Lett*. 79:435–438.
18. Chu, V., Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach. 2007. Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory. *Biophys. J*. 93:3202–3209.
19. Beglov, D., and B. Roux. 1995. Numerical solution of the HNC equation for solutes of arbitrary geometry in three dimensions. *J. Chem. Phys*. 103:360–364.
20. Beglov, D., and B. Roux. 1996. Solvation of complex molecules in a polar liquid: an integral equation theory. *J. Chem. Phys*. 104:8678–8689.
21. Burak, Y., and D. Andelman. 2000. Hydration interactions: aqueous solvent effects in electric double layer. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*. 62:5296–5310.
22. Burak, Y., and H. Orland. 2006. Manning condensation in two dimensions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 73:010501–010504.

23. Grant, J., B. Pickup, and A. Nicholls. 2001. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.* 22:608–640.
24. Schutz, C., and A. Warshel. 2001. What are the dielectric ''constants'' of proteins and how to validate electrostatic models? *Proteins Struct. Funct. Genet.* 44:400–417.
25. Ehrenson, S. 1989. Continuum radial dielectric functions for ion and dipole solution systems. *J. Comput. Chem.* 10:77–93.
26. Hassan, S., F. Guarneri, and E. Mehler. 2000. A general treatment of solvent effects based on screened Coulomb potentials. *J. Phys. Chem. B.* 104:6478–6489.
27. Li, X., S. Hassan, and E. Mehler. 2005. Long dynamics simulations of proteins using atomistic force fields and a continuum representation of solvent effects. *Proteins Struct. Funct. Bioinform.* 60:464–484.
28. Lounnas, V., B. Pettitt, and G. Phillips. 1994. A global model of the protein-solvent interface. *Biophys. J.* 66:601–614.
29. Svergun, D., S. Richard, M. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95:2267–2272.
30. Zacharias, M. 2003. Continuum solvent modeling of non-polar solvation: improvement by separating surface area dependent cavity and dispersion contribution. *J. Phys. Chem. A.* 107:3000–3004.
31. Wagoner, J., and N. Baker. 2006. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. USA.* 103:8331–8336.
32. Lum, K., D. Chandler, and J. Weeks. 1999. Hydrophobicity at small and large length scales. *J. Phys. Chem. B.* 103:4570–4577.
33. Huang, D., and D. Chandler. 2002. The hydrophobic effect and the influence of solute-solvent interactions. *J. Phys. Chem. B.* 106:2047–2053.
34. Head-Gordon, T. 1995. Is water structure around hydrophobic groups clathrate-like? *Proc. Natl. Acad. Sci. USA.* 92:8308–8312.
35. Head-Gordon, T., J. Sorensen, A. Pertsemlidis, and R. Glaeser. 1997. Differences in hydration structure near hydrophobic and hydrophilic amino acids. *Biophys. J.* 73:2106–2121.
36. Pertsemlidis, A., A. Saxena, A. Soper, T. Head-Gordon, and R. Glaeser. 1998. Direct evidence for modified solvent structure within the hydration shell of a hydrophobic amino acid. *Proc. Natl. Acad. Sci. USA.* 93:10769–10774.
37. Rowlinson, J., and B. Widom. 1982. Molecular Theory of Capillarity. Oxford University Press, Oxford, UK.
38. Warshel, A., and M. Levitt. 1976. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103:227–249.
39. Warshel, A., and S. Russell. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* 17:283–422.
40. Russell, S., and A. Warshel. 1985. Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* 185:389–404.
41. Warshel, A., and A. Papazyan. 1998. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* 8:211–217.
42. Azuara, C., E. Lindahl, P. Koehl, H. Orland, and M. Delarue. 2006. Incorporating dipolar solvents with variable density in the Poisson-Boltzmann treatment of macromolecular electrostatics. *Nucleic Acids Res.* 34:W34–W42.
43. Abrashkin, A., D. Andelman, and H. Orland. 2007. Dipolar Poisson-Boltzmann equation: ions and dipoles close to charge interfaces. *Phys. Rev. Lett.* 99:77801.
44. Masella, M., D. Borgis, and P. Cuniasse. 2008. Combining a polarizable force-field and a coarse-grained polarizable solvent model. *J. Comput. Chem.* 29:111–118.
45. Borukhov, I., D. Andelman, and H. Orland. 2000. Adsorption of large ions from an electrolyte solution: a modified Poisson-Boltzmann equation. *Electr. Acta.* 46:221–227.
46. Ramirez, R., and D. Borgis. 2005. Density functional theory of solvation and its relation to implicit solvent models. *J. Phys. Chem. B.* 109:6954–6963.
47. Jackson, J. 1975. Classical Electrodynamics. John Wiley and Sons, New York.
48. Israelachvili, J. 1992. Intermolecular and Surface Forces. Academic Press, New York.
49. Koehl, P., and M. Delarue. 1995. A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modeling. *Nat. Struct. Biol.* 2:163–170.
50. Koehl, P., and M. Delarue. 1994. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* 239:249–275.
51. Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
52. Dolinsky, T., J. Nielsen, J. M. Cammon, and N. Baker. 2004. Pdb2pqr: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32:W665–W667.
53. Richards, F. M. 1977. Areas, volumes, packing, and protein-structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.
54. Jiang, J., and A. Brünger. 1994. Protein hydration observed by x-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* 243:100–115.
55. Press, W., B. Flannery, S. Teukolsky, and W. Vetterling. 1992. Numerical Recipes in Fortran: The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.
56. Baker, N. A., D. Sept, J. Simpson, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.
57. Brünger, A. T., P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewiski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54:905–921.
58. Jones, T., J. Zou, S. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A.* 47:110–118.
59. Emsley, P., and K. Cowtan. 2004. COOT, a model-building program for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60: 2126–2132.
60. Shrake, A., and J. A. Rupley. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79:351–371.
61. Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.
62. Park, B., and M. Levitt. 1996. Energy functions that discriminate x-ray and near native folds from well-constructed decoys. *J. Mol. Biol.* 258: 367–392.
63. Koehl, P., and M. Delarue. 1994. Polar and nonpolar atomic environments in the protein core: implications for folding and binding. *Proteins Struct. Funct. Genet.* 2:264–278.
64. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.
65. Simone, A. D., G. Dodson, C. Verma, A. Zagari, and F. Fraternali. 2005. Prion and water: tight and dynamical hydration sites have a key role in structural stability. *Proc. Natl. Acad. Sci. USA.* 102:7535–7540.
66. Higo, J., and M. Nakasako. 2002. Hydration structure of human lysozyme investigated by molecular dynamics simulation and cryogenic x-ray crystal structure analyses: on the correlation between

crystal water sites, solvent density, and solvent dipole. *J. Comput. Chem.* 23:1323–1336.

67. Higo, J., H. Kono, H. Nakamura, and A. Sarai. 2000. Solvent density and long-ranged dipole field around a DNA-binding protein studied by molecular dynamics. *Proteins Struct. Funct. Genet.* 40:193–206.
68. Fröhlich, H. 1958. Theory of Dielectrics. Oxford University Press, Oxford, UK.
69. Hansen, J., and I. McDonald. 1986. Theory of Simple Liquids. Academic Press, New York.
70. Coalson, R., and A. Duncan. 1996. Statistical mechanics of a multipolar gas: a lattice field theory approach. *J. Phys. Chem.* 100:2612–2620.
71. Kuhn, L., M. Siani, M. Pique, C. Fisher, E. Getzoff, and J. Tainer. 1992. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J. Mol. Biol.* 228:13–22.
72. Kuhn, L., C. Swanson, M. Pique, J. Tainer, and E. Getzoff. 1995. Atomic and residue hydrophilicity in the context of folded protein structures. *Proteins Struct. Funct. Genet.* 23:536–547.
73. Fauchere, J.-L., and V. Pliska. 1983. Hydrophobic parameters π of amino-acid side chains from the partitioning of *n*-acetyl-amino-acid amides. *Eur. J. Med. Chem. Chim. Ther.* 18:369–375.
74. Rose, G., and R. Wolfenden. 1993. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 22:381–415.
75. Trinquier, G., and Y. Sanejouand. 1998. Which effective property of amino acids is best preserved by the genetic code? *Protein Eng.* 11:153–169.
76. Steinbach, P., and B. Brooks. 1993. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* 90:9135–9140.
77. Knab, J., J. Chen, and A. Markelz. 2006. Hydration dependence of conformational dielectric relaxation in lysozyme. *Biophys. J.* 90: 2576–2581.
78. Bone, S., and R. Pethig. 1982. Dielectric studies of the binding of water to lysozyme. *J. Mol. Biol.* 157:571–575.
79. Bone, S., and R. Pethig. 1985. Dielectric studies of protein hydration and hydration-induced flexibility. *J. Mol. Biol.* 181:323–326.
80. Suzuki, M., J. Shigematsu, and T. Kodama. 1996. Hydration studies of proteins in solution by microwave dielectric analysis. *J. Phys. Chem.* 100:7279–7282.
81. Azuara, C. 2006. In silico study of the solvation of proteins. PhD thesis, Universite Paris VII, Paris, France.
82. Yang, L., J. Dordick, and S. Garde. 2004. Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity. *Biophys. J.* 87:812–821.
83. Micaelo, N., and C. Soares. 2007. Modeling hydration mechanisms of enzymes in nonpolar and polar organic solvents. *FEBS J.* 274:2424–2436.
84. Netz, R., D. Andelman, and H. Orland. 1996. Protein adsorption on lipid monolayers at their coexistence region. *J. Phys.* 6:1023–1047.
85. Lounnas, V., and B. Pettitt. 1994. Distribution function implied dynamics versus residence times and correlations: solvation shells in myoglobin. *Proteins Struct. Funct. Genet.* 18:133–147.
86. Lounnas, V., and B. Pettitt. 1994. A connected-cluster of hydration around myoglobin: correlation between molecular dynamics simulations and experiments solvation shells in myoglobin. *Proteins Struct. Funct. Genet.* 18:148–160.
87. Makarov, V., B. Andrews, and B. Pettitt. 1998. Reconstructing the protein-water interface. *Biopolymers.* 45:469–478.
88. Pettitt, B., V. Makarov, and B. Andrews. 1998. Protein hydration density: theory, simulations and crystallography. *Curr. Opin. Struct. Biol.* 8:218–221.
89. Halle, B. 2004. Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl. Acad. Sci. USA.* 101:4793–4798.
90. Nakasako, M. 2004. Water-protein interactions from high-resolution crystallography. *Phil. Trans. Roy. Soc. Lond.* 359:1191–1206.
91. Netz, R., and H. Orland. 1999. Beyond Poisson-Boltzmann: fluctuations and correlations. *Eur. Phys. J. E.* 3:203–214.
92. Chowdury, S., M. Tan, and T. Ichiye. 2006. Dynamical properties of the soft sticky dipole-quadrupole-octupole water model: a molecular dynamics study. *J. Chem. Phys.* 125:144513–144518.
93. Buldyrev, S., P. Kumar, P. Debenedetti, P. Rossky, and H. Stanley. 2007. Water-like solvation thermodynamics in a spherically symmetric solvent model with two characteristic lengths. *Proc. Natl. Acad. Sci. USA.* 104:20177–20182.
94. Ramirez, R., R. Gebauer, M. Mareschal, and D. Borgis. 2002. Density functional theory of solvation in a polar solvent: extracting the functional from homogeneous solvent simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 66:031206.
95. Biben, T., J. Hansen, and Y. Rosenfeld. 1998. Generic density functional for electric double layer in a molecular solvent. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 57:R3727.
96. Basdevant, N., T. Haduong, and D. Borgis. 2004. A semi-implicit solvent model for the simulation of peptides and proteins. *J. Comput. Chem.* 25:1015–1029.
97. Imai, T., Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata. 2006. A theoretical analysis on hydration thermodynamics of proteins. *J. Chem. Phys.* 125:24911–24919.
98. Imai, T., Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata. 2007. Theoretical analysis on changes in thermodynamic quantities upon protein folding: essential role of hydration. *J. Chem. Phys.* 126: 225102–225117.
99. Miyata, T., and F. Hirata. 2008. Combination of molecular dynamics method and 3D-RISM theory for conformational sampling of large flexible molecules in solution. *J. Comput. Chem.* 29:871–882.
100. Borukhov, I., D. Andelman, R. Borrega, M. Cloitre, L. Leibler, and H. Orland. 2000. Polyelectrolyte titration: theory and experiment. *J. Phys. Chem. B.* 104:11027–11034.
101. Fleck, C., R. Netz, and H. von Grunberg. 2002. Poisson-Boltzmann theory for membranes with mobile charged lipids and the pH-dependent interaction of a DNA molecule with a membrane. *Biophys. J.* 82:76–92.
102. Georgescu, R., E. Alexov, and M. Gunner. 2002. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys. J.* 83:1731–1748.
103. Bashford, D. 2004. Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci.* 9:1082–1099.
104. Ondrechen, M., J. Clifton, and D. Ringe. 2001. THEMATICS: a simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci. USA.* 98:12473–12478.
105. Burak, Y., and D. Andelman. 2001. Discrete aqueous solvent effects and possible attractive forces. *J. Chem. Phys.* 114:3271–3279.
106. Nakamura, H., T. Sakamoto, and A. Wada. 1988. A theoretical study of the dielectric constant of proteins. *Protein Eng.* 2:177–183.
107. Simonson, T., G. Bricogne, and D. Perahia. 1991. Intramolecular dielectric screening in proteins. *J. Mol. Biol.* 218:859–886.
108. Lindahl, E., C. Azuara, P. Koehl, and M. Delarue. 2006. NOMAD-REF: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* 34:W52–W56.
109. Phillips, G., and B. Pettitt. 1995. Structure and dynamics of the water around myoglobin. *Protein Sci.* 4:149–158.
110. Merzel, F., and J. Smith. 2002. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci. USA.* 99: 5378–5383.
111. Coalson, R., A. Walsh, A. Duncan, and N. Bien-Tal. 1995. Statistical mechanics of a Coulomb gas with finite size particles: a lattice field theory approach. *J. Chem. Phys.* 102:4584–4594.