JCTC Journal of Chemical Theory and Computation

Integrated Continuum Dielectric Approaches To Treat Molecular Polarizability and the Condensed Phase: Refractive Index and Implicit Solvation

Jean-François Truchon,[†,‡] Anthony Nicholls,[§] Benoît Roux,[||] Radu I. Iftimie,[†] and Christopher I. Bayly*[,‡]

Département de chimie, Université de Montréal, C.P. 6128 Succursale centre-ville, Montréal, Québec, Canada H3C 3J7, Merck Frosst Canada Ltd., 16711 TransCanada Highway, Kirkland, Québec, Canada H9H 3L1, OpenEye Scientific Software, Inc., Santa Fe, New Mexico 87508, and Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Science, University of Chicago, 929 East 57th Street, Chicago, Illinois 60637

Received January 15, 2009

Abstract: The idea of using a dielectric continuum inside a molecule to accurately model molecular polarizability is extended to include a larger spectrum of bioorganic molecules and the condensed phase. Atomic polarization radii and an internal dielectric (ε_{in}) were fitted to reproduce ab initio B3LYP/aug-cc-pVTZ polarizability tensors taken from a data set of 707 molecules. The average unsigned error on the isotropic polarizability and anisotropy are 2.6% and 5.2%, respectively. It is shown that usual Poisson−Boltzmann contact radii and a low internal dielectric are not appropriate and require major revision. To account for the anisotropy of polarizability, the internal dielectric (ε_{in}) constant needs to be larger than 6.0. Reinterpreting the theoretical link between ε_{in} and the experimental refractive index (n), this study shows, with a set of 23 organic molecules spanning the entire range of n, that even with $\varepsilon_{in} = 24$ the obtained refractive indices can correlate well with experiment (slope of 1.00, intercept of 0.05, and R = 0.95). The novel methodology used here to calculate a macroscopic-like refractive index shows that the application of the EPIC parametrization to condensed phase leads to suitable behavior. Although the primary goal in developing EPIC was to include polarizability in explicit solvent calculations, we also extend the model to work with implicit solvent. This requires the use of a 3-zone smooth dielectric function to transition from the polarization dielectric inside the molecules to the dielectric continuum of the solvent. The parametrization and validation of this model are performed against 485 experimental free energies of hydration. Using 8 solvent cavity atomic radii and a single surface tension an average unsigned error of 1.1 kal/mol and a correlation coefficient of 0.9 are obtained, validating the use of the EPIC model in the condensed phase.

1. Introduction

The newly introduced treatment of electronic polarization by an internal continuum (EPIC) was shown to be accurate in reproducing experimental and density functional (DFT) molecular polarizability tensors with a remarkably small number of adjustable parameters.[1] Moreover, the high accuracy found when computing intermolecular interaction energies, in which the appropriate treatment of electronic polarization is crucial, opens up the possibility of using EPIC to include polarizability in force fields.[2] This led us to propose the use of EPIC to embed polarizability in all-atom explicit-solvent calculations. EPIC uses continuum dielectric

* Corresponding author phone: (514)428-3403; fax: (514)428-4930; e-mail: christopher_bayly@merck.com.
[†] Université de Montréal.
[‡] Merck Frosst Canada Ltd.
[§] OpenEye Scientific Software, Inc.
[||] University of Chicago.

10.1021/ct900029d CCC: $40.75
Published on Web 06/05/2009

electrostatic theory to account for the way electronic density polarizes under the presence of an external electric field that can come from either other molecules in explicit condensed phase calculations or the reaction field in an implicit solvent calculation. In contrast to the point inducible dipoles[3−5] or the Drude's oscillator models[6,7] that use the atomic nuclear positions as polarizable centers, EPIC employs a polarizability density that induces a dipole density, normally referred to as polarization, throughout the molecule volume as a response to the local electric field. In a recent study, Schropp and Tavan[8] proposed that the use of single centers in point inducible dipole polarizable calculations was responsible for the large difference between the best condensed phase atomic polarizability and the best vacuum phase atomic polarizabilities previously noticed.[9,10] Other studies, based on Quantum Mechanical (QM) assessment, suggest that the polarizability in condensed phase should only be slightly reduced.[11] The idea of using a continuum dielectric to account for electronic polarization was first formulated by Sharp et al.[12] but was not further pursued until Tan and Luo[13] optimized the internal dielectric of solutes to produce the electrostatic potential in the context of Poisson−Boltzmann calculations with different implicit solvents. In their two studies,[13,14] they do not attempt to give a detailed molecular polarizability description but rather focus on the shift in dipole moments when a solute is put in different solvent. It is difficult to decouple the solvent polarization from the solute polarization and the cooperative polarization when calculations in implicit solvent are done. The current study uses previously developed techniques[1,2] to separate the charge fitting from the polarizability fitting by optimizing, in the absence of atomic charges, an *electronic volume* to fit quantum mechanics (QM) polarizability tensors for molecules in vacuum, as was done originally with other polarizable models.[3,5,15] Curiously, we found that in order to accurately reproduce the polarizability tensors of even challenging molecules, the atomic radii needed to be much smaller than the van der Waals (vdW) contact radii usually used in implicit solvent calculations (e.g., Bondi radii[16]). At the same time, the internal dielectric needed to be surprisingly high in order to reproduce the anisotropy of the polarizabilities. While that work allowed for a systematic way of adjusting a dielectric function to account for electronic polarization, it raised two issues: the abnormally high internal dielectric of 14 seems questionable and the small radii made implicit solvent calculations impractical. Regarding the first issue, the dielectric inside the molecule is closely related to the refractive index squared ($\varepsilon_\infty = n^2$) of the pure liquid, which adopts values between 1.7 and 2.9 for organic liquids, far below our large values. Regarding the second issue, if such small atomic radii were used to define the molecular cavity in solvent, the free energy of charging would become unrealistically negative e.g. in Poisson−Boltzmann (PB) calculations. In this work, we specifically address both issues and demonstrate the physical soundness of the approach. An important change from our previous work is the use of a smooth dielectric boundary to represent both the solute and the solvent polarization. We present a newly designed dielectric function with 3 zones (3-zone dielectric) that permits the use of EPIC for implicit solvent calculations. We show that describing the dielectric function this way better reflects the underlying physical principles involved in solvation than the usual 2-zone dielectric (i.e., inside and outside the cavity).

Another question that we examine is the ability to optimize the EPIC parameters in a general and robust way with few parameters on a larger variety of chemical functionality than in earlier work. For this purpose, we have formed a large database of QM molecular polarizability tensors for 707 diverse bioorganic molecules (for a total of 4242 polarizability tensor elements) along with their optimized molecular geometries (cf. the Supporting Information). As will be outlined below, this data set contains a large variety of chemical functional groups representing a significant component of bioorganic chemistry. This substantially enlarged parametrization of the polarizable EPIC model is then used for the calculation of refractive indices and hydration free energies. The validity of both the internal dielectric function and the 3-zone dielectric function is assessed with the independent fit of the solvent cavity atomic radii (which define the third zone of the function) on 485 experimental free energies of hydration.

In the remainder of this article, section 2 presents the theoretical basis and methods employed, where we present the 3-zone dielectric function for implicit solvent calculations and we review the polarizability tensor calculation. This is followed by the theoretical background for the calculation of the refractive index. A theoretical layout for free energy of hydration calculations and computational details related to quantum calculations close this section. Section 3 describes the chemical data sets used in section 4 where the results and their analysis are presented. Section 4 closes with a 3-zone dielectric optimization on experimental hydration free energies, leading into the conclusons.

2. Theory and Methods

2.1. 3-Zone Dielectric in Implicit Solvents. The dielectric function in continuum approaches is fundamental as it is modulating all sources of polarization. In this work, we move away from our previous use of vdW envelope surfaces[17] toward a smooth functional form based on a sum of atomic Gaussians which has previously proven successful[18,19] in PB applications. Although useful, the hard dielectric boundary often leads to numerical problems: iterative convergence failure, slower convergence, strong dependency on orientation and translation, and unstable force evaluations.[18,20] The use of smooth solute/solvent dielectric boundary was shown to improve over the hard boundary on all these aspects. More specifically, the molecular dielectric function used in the present work is given by

$$\varepsilon(\vec{r}) = \varepsilon_{in} - (\varepsilon_{in} - \varepsilon_{ext})\exp(-A \cdot f_{in}(\vec{r})) \qquad (1)$$

where ε_{in} is the dielectric constant inside the molecular volume, and ε_{ext} the dielectric value outside. The dielectric here is expressed as a permittivity relative to the vacuum permittivity. The exponential behaves as a switching function that is turned on or off depending on the value of a molecular

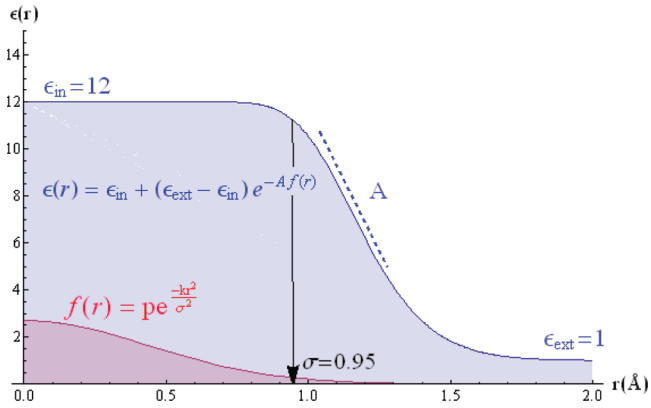

Figure 1. This figure shows the smooth dielectric function used in this work for a single atom with $\sigma = 0.95$ Å, $\varepsilon_{in} = 12$, $\varepsilon_{ext} = 1$, and $A = 10.0$ (Cl of the G1−12 set). Starting from the center of the atom ($r = 0$), the dielectric (blue curve) stays constant until the 'density', expressed with a sum of Gaussians (pink curve), reaches a certain small value that causes the dielectric to smoothly transition to the external dielectric value. The steepness and the position of the switching region depends on the value of the A parameter. The sum-of-Gaussians density expression is explained in eq 2 (see text).

'density' function $f_{in}(\vec{r})$. The A parameter modulates the steepness of the switching function. The details of the dielectric are then incorporated into the 'density' function

$$f_{in}(\vec{r}) = \sum_{i=1}^{Natoms} p \cdot \exp\left(-k\frac{|\vec{r}_i - \vec{r}|^2}{\sigma_i^2}\right) \qquad (2)$$

The summation runs over all atoms, and a 3-dimensional Gaussian defines the radial extent of the atomic volume; σ_i are atomic radii and $\vec{r}_i$ their positions. The σ_i will be the subject of an extensive parametrization in the next sections. The constant k is set to 2.3442 and p to 2.7 following the Grant et al. recommendation.[18] Equation 1 can be conceptually understood in terms of electronic density that would have a constant susceptibility (polarizability density) inside and drops rapidly as the density vanishes as shown in Figure 1.

The main methodological novelty proposed in this work is the 3-zone dielectric for the coupling of EPIC with implicit solvation. When atomic radii are optimized on QM-based molecular polarizability tensors, their resulting small size prohibits their use to define the cavity formed by the solute in implicit solvent calculations. Indeed, it presents a dilemma: on the one hand, accurate solute polarization requires atomic radii far smaller than accepted contact radii. On the other hand, the solvent boundary for implicit solvation requires atomic radii as large or larger than contact radii. The resolution to this dilemma is found in challenging the assumption that the atomic radii for solute polarization and for the solvent boundary should be the same. There is no underlying physical reason why the polarization response of an atom in a molecule would be uniform all the way out to its contact radius; on the contrary our QM model for molecules tells us the electron density (the source of electronic polarization) drops exponentially in moving from an atomic nucleus toward the contact surface of the molecule. We believe that it is more reasonable to think that the radial extent of the electronic polarization can be different from the vdW radius used for the solvent cavity. The idea presented here is that both kinds of smooth surfaces could be simultaneously used: one for solute polarization, formed with the smaller atomic polarization radii, and one for solvent polarization, defined with the solvent cavity atomic radii. In between the two surfaces is a transition region of low dielectric since it describes where the solvent and the solute electrons are both at a minimum. This leads to a 3-zone dielectric function to which we give the form

$$\varepsilon(\vec{r}) = \varepsilon_{in} + (\varepsilon_{trans} - \varepsilon_{in}) \exp[-A \cdot f_{in}(\vec{r})] + (\varepsilon_{solv} - \varepsilon_{trans}) \exp[-B \cdot f_{solv}(\vec{r})] \qquad (3)$$

where ε_{in} is the dielectric constant inside the molecular cavity, ε_{solv} is the bulk solvent dielectric constant (80 for water), and ε_{trans} is the dielectric constant in the zone of transition between the solute and the solvent. For the smooth inner dielectric boundary, A has the same meaning as in eq 1, and $f_{in}(\vec{r})$ is given by eq 2. The additional exponential term, for the outer dielectric boundary (with solvent), is a switching function that turns on when a second Gaussian sum ($f_{solv}(\vec{r})$) becomes sufficiently small. The $f_{solv}(\vec{r})$ term is also given by eq 2 with the difference that the atomic radii are larger as they define the solvent cavity. The B parameter is responsible for the steepness of the cavity boundary, but with a sufficiently large value it has the effect of moving the position of the boundary as if the radii were scaled. The radial behavior of the 3-zone dielectric is illustrated in Figure 2 for a single atom and for the 4-pyridone molecule, both with typical parameters. In eq 3, it is important to set $\varepsilon_{trans} = 1$ when the first zone of the dielectric function is fitted on molecular polarizability since the shape of the dielectric function needs to drops to one in order to present the same ability to polarize. Also, if the atomic partial charges are fitted with DRESP, a change in the first zone boundary would also change the ability of the dielectric to form the full internal polarization taken into account during the charge fitting process.

2.2. Molecular Polarizability Tensor. In this section, we review the methodology previously developed to calculate molecular polarizability tensor with a finite difference Poisson solver,[1] and we summarize how the parameters involved are optimized in this work.

2.2.1. Method. Our formulation of electronic polarization based on continuum electrostatics allows the calculation of induced multipolar moments by considering the bound charge density, which results from the polarizability density of the media (from the *bound* the electrons in our case). A formula to calculate the bound charge density is[21]

$$\frac{\rho^b(\vec{r})}{\varepsilon_0} = -\vec{\nabla}([\varepsilon(\vec{r}) - 1]\vec{E}(\vec{r})) = -\vec{\nabla} \cdot \vec{P}(\vec{r}) \qquad (4)$$

where ρ^b is the bound charge density, and $\vec{E}(\vec{r})$ the total electric field. Physically, ρ^b is a consequence of the formation of dipoles at each point in space due to the electric field (the polarization $\vec{P}(\vec{r})$ or dipole density). The bound charge density can be thought of as an induced charge density from

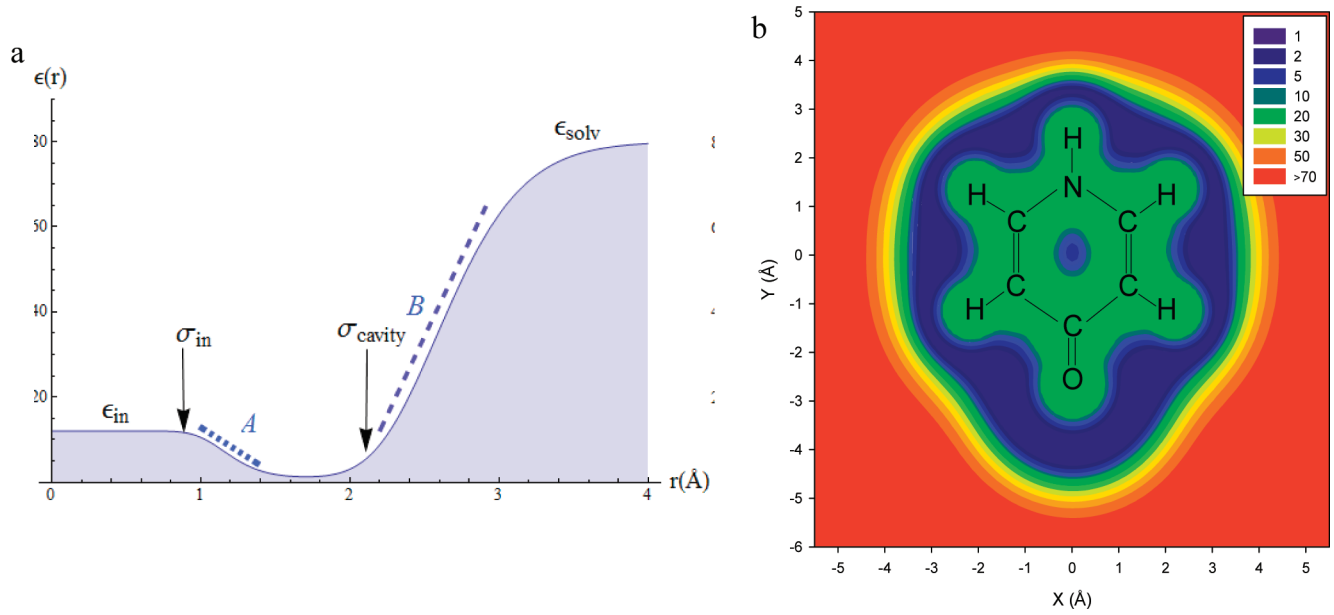

Figure 2. The 3-zone dielectric function allows an accurate description of both the solute polarization and the solvent polarization within the EPIC approach. (a) The radial component of the dielectric for a single atom (G1−12 aromatic carbon) is shown together with the polarization (σ_{in}) and the solvent cavity (σ_{cavity}) atomic radii. Each plateau of the dielectric function defines a zone. The middle intermediate zone corresponds to the solute/solvent contact distance. (b) The resulting dielectric function is also shown in the ring plane of 4-pyridone (b) when applying the G2−12 parameters.

the dielectric polarization that appears where the dielectric varies, as an excess of charge builds due to the head or tail of the dipole density. Although the polarization occurs everywhere the dielectric is greater than one, the bound charge density appears in regions of spaces where $\varepsilon(\vec{r})$ varies, such as the dielectric boundary of a molecule. Equation 4 is useful since it transforms the locally induced dipoles into a scalar value, the bound charge density, which can be used more easily as done below. In eq 4, $\varepsilon(\vec{r}) - 1$ plays the role of a local polarizability density, also called the electric susceptibility, and $\vec{P}(\vec{r}) = (\varepsilon(\vec{r}) - 1)\vec{E}(\vec{r})$ corresponds to the induced dipole density (polarization). The analogy with the point inducible dipole model, a different polarizable model, is obvious since, in that case, the atomic induced dipole is given by $\vec{\mu}(\vec{r}_i) = \alpha_i\vec{E}(\vec{r}_i)$ where $\vec{\mu}(\vec{r}_i)$, α_i, and $\vec{E}(\vec{r}_i)$ are the dipole induced at the atomic position $\vec{r}_i$, the atomic polarizability, and the electric field at $\vec{r}_i$. Here, the polarization is more smoothly distributed over the molecular volume. Equation 4 is intrinsic to the definition of Poisson's equation.

A classical example, for which an analytical solution exists, is the dielectric sphere in vacuum experiencing an external electric field. In this case the mathematics show that bound charges appear on the surface of the sphere with opposite charge sign on both hemispheres, resulting in an induced potential equivalent to an ideal induced dipole moment aligned with the external field located at the center of the sphere. The induced dipole moment is proportional to the external electric field, and the sphere polarizability α_{sphere} is given by the Clausius-Mossoti equation

$$\alpha_{sphere} = \left(\frac{\varepsilon_{sphere} - 1}{\varepsilon_{sphere} + 2}\right) R_{sphere}^3 \tag{5}$$

where R_{sphere} is the sphere radius. For a molecular system, the analytical solution is unknown, and we use a finite difference algorithm to solve Poisson's equation numerically with a uniform electric field in the form of a voltage clamp applied by means of the boundary conditions. More precisely, a uniform electric field in the z direction can be produced with a null potential on one side of the grid boundary and the value $-E_{ext} \times L_z$ on the opposite side, where L_z is the box size in the z direction, and E_{ext} is the magnitude of the applied field. On the four other sides, parallel to the field, the grid boundary potential is simply calculated as a linear interpolation along the z direction: $\varphi(z\text{-}z_0) = -(z\text{-}z_0) \times E_{ext}$. As with the dielectric sphere in vacuum, a molecular dielectric volume responds linearly to the applied field (given an isotropic dielectric function), and the proportionality constant is the molecular polarizability tensor. The field is applied in three orthogonal directions to build the polarizability tensor, which depends on the orientation of the molecule

$$\bar{\alpha} = \begin{bmatrix} \frac{\mu_{x,x}}{E_{ext}} & \frac{\mu_{x,y} + \mu_{y,x}}{2E_{ext}} & \frac{\mu_{x,z} + \mu_{z,x}}{2E_{ext}} \\ & \frac{\mu_{y,y}}{E_{ext}} & \frac{\mu_{y,z} + \mu_{z,y}}{2E_{ext}} \\ & & \frac{\mu_{z,z}}{E_{ext}} \end{bmatrix} \tag{6}$$

where $\mu_{x,y}$ is the x component of the induced dipole moment when an external electric field of magnitude E_{ext} is applied in the y direction. Some experimental values are available for the eigenvalues of this tensor in vacuum ($\varepsilon_{ext} = 1$); also, the polarizability tensor can be calculated using approaches based on quantum mechanics (QM) methods such as density functional theory.

The induced dipole moment is calculated analogously to the sphere dielectric system, integrating the bound charge

density over space. From eq 4 (or simply from Gauss's law), one can show that

$$\rho^b(\vec{r}) = -\rho^f(\vec{r}) + \varepsilon_0 \vec{\nabla} \cdot \vec{E}(\vec{r}) \quad (7)$$

In the present context, there is no free charge density $\rho^f(\vec{r})$ (from atomic partial charges, for instance), and as such the bound charge density, induced only by the external uniform electric field, is given by the divergence of the field. With a finite difference solver, the total charge (bound and free charges) can be calculated by integrating over each differential volume element (grid cube) which leads to bound charges on grid points. This can be done simply by calculating

$$\frac{q_{ijk}}{\varepsilon_0} = \frac{q^b_{ijk} + q^f_{ijk}}{\varepsilon_0} = -\left(\frac{h_y h_z}{h_x}\right)(\varphi_{i+1jk} + \varphi_{i-1jk} - 2\varphi_{ijk}) - \left(\frac{h_x h_z}{h_y}\right)(\varphi_{ij+1k} + \varphi_{ij-1k} - 2\varphi_{ijk}) - \left(\frac{h_x h_y}{h_z}\right)(\varphi_{ijk+1} + \varphi_{ijk-1} - 2\varphi_{ijk}) \quad (8)$$

where q_{ijk}, q^b_{ijk}, and q^f_{ijk} are the total charge, the bound charge, and the free charge inside the volume element associated with the *ijk* grid point, and φ_{ijk} and $\varphi_{ijk\text{-}1}$ are the electrostatic potential at the (*x,y,z*) and (*x,y,z-dz*) grid points, respectively. The grid spacing in *x*, *y*, and *z* are given by h_x, h_x, and h_z. The grid free charge q^f_{ijk} are zero for this calculation, and, in general, it is given by the atomic partial charges as distributed on the grid. Finally, the total dipole moment is given by

$$\vec{\mu} = \sum_{i,j,k}^{Grid} \vec{r}_{ijk} q_{ijk} \quad (9)$$

With the free charges equal to zero (no atomic partial charge), the dipole calculated is then the induced dipole, and the only contributor is the bound charge density. More generally, any molecular electric moment can be calculated with analogs to eq 9. The overall procedure to calculate the polarizability tensor requires three solutions from the numerical solver. The calculation does not involve atomic partial charges (free charges) which allows them to be fit independently (although this must still be done in the context of the molecular dielectric).

2.2.2. Computational Details. The finite difference Poisson calculations were performed with a modified version of the OpenEye Inc. ZapTK.[22] The distance between two grid points was set to 0.35 Å, and the grid boundary was at least 5 Å away from the surface defined by the polarization radii. Atomic charges of $\pm 0.001e$ were assigned randomly on the atoms as the grid energy was used to determine the convergence of the algorithm set to 0.000001 k_BT. The results were not sensitive to these small charges. Atom typing was assigned via SMARTS[23−25] with the OpenEye Inc. OEchem toolkit.[26]

2.2.3. Optimization of the Polarizabilities. The atomic radii were optimized in order to minimize a chi-square function using a Levenberg−Marquardt algorithm as implemented in scipy,[27] a scientific Python library. The error was defined as the difference between the 6 components of the polarizability tensor obtained with B3LYP and EPIC

$$\chi^2 = \sum_i^{molecules} \sum_{k=xx,xy,xz,yy,yz,zz} (\alpha^{EPIC}_{k,i} - \alpha^{QM}_{k,i})^2 \quad (10)$$

where $\alpha_{xy,i}$ is one of the six independent polarizability tensor elements of molecule i either under optimization (EPIC) or from the QM target values. By using the six independent tensor elements, we included both the magnitude and the direction of the polarizability in a natural way.[28] We optimized the cube of the polarization radii because their contribution to the polarizability grows with the atomic volume (cf. eq 5). For analysis purposes, we also defined the average polarizability (eq 11) and the anisotropy of the polarizability tensor (eq 12) below

$$\alpha_{avg} = (\alpha_1 + \alpha_2 + \alpha_3)/3 \quad (11)$$

$$\Delta\alpha = \sqrt{\frac{(\alpha_1 - \alpha_2)^2 + (\alpha_1 - \alpha_3)^2 + (\alpha_2 - \alpha_3)^2}{2}} \quad (12)$$

where $\alpha_1 \leq \alpha_2 \leq \alpha_3$ are the eigenvalues of the polarizability tensor. The polarizability anisotropy is significantly harder to fit than the average polarizability. We defined the error in the average polarizability (eq 13) and anisotropy (eq 14) for a set of molecules as

$$\delta_{avg} = \frac{1}{N}\sum_i^N \frac{|\alpha^{QM}_{i,avg} - \alpha_{i,avg}|}{\alpha^{QM}_{i,avg}} \quad (13)$$

$$\delta_{aniso} = \frac{1}{N}\sum_i^N \frac{|\Delta\alpha^{QM}_i - \Delta\alpha_i|}{\alpha^{QM}_{i,avg}} \quad (14)$$

where *N* is the total number of molecules considered and *QM* corresponds to the target value. Finally, the relative root-mean-square deviation (RRMS) of the tensor was defined as

$$RRMS = \frac{\sum_i^{molecules} \sum_{k=xx,xy,xz,yy,yz,zz} (\alpha^{EPIC}_{k,i} - \alpha^{QM}_{k,i})^2}{\sum_i^{molecules} \sum_{k=xx,xy,xz,yy,yz,zz} (\alpha^{QM}_{k,i})^2} \quad (15)$$

and constituted a single metric for the overall fitness of the optimized polarizability tensors. If the RRMS was calculated for a single molecule, the summations on the molecules in the numerator and the denominator were simply omitted.

2.3. Refractive Index Calculations. *2.3.1. Theory.* The dielectric constant of an isotropic material at the high frequency limit (ε_∞) is related to the material refractive index[29] *n* by

$$n^2 = \varepsilon_\infty \quad (16)$$

where *n* is usually measured with the D line of the sodium spectrum at 589 nm (n_D). The ε_∞ corresponds to the material's

dielectric constant solely due to the electronic polarization since the frequency of the visible light is too high for nuclear relaxation to contribute. Typically, a pure liquid of an organic compound will have a refractive index between 1.3 and 1.7 leading to a ε_∞ between 1.7 and 2.9. Since the work of Debye and Onsager,[12,13,30,31] it has become a dogma that the interior dielectric (ε_{in}) of a solute cavity in implicit solvent models should be close to the experimental ε_∞ in order to capture the dipole moment change due to the cooperative solute−solvent polarization. It is when we seek for accuracy in solute polarization that we found the generally accepted relation $\varepsilon_\infty = \varepsilon_{in}$ to badly fail.[1] A way to reconcile this puzzling finding is by computing the macroscopic refractive index that corresponds to what is measured instead of assuming it is the same as the *internal refractive index* (quoting Onsager[30]). The Clausius-Mossoti equation relates the polarizability of a sphere to its interior dielectric. Since ε_∞ and n are macroscopic intensive quantities, their measurement should not depend on the size of the studied sample, given that it is large enough to exhibit a macroscopic behavior, the worst case being the use of a single molecule. It is not to say that Onsager's uses of the Clausius-Mossoti equation with the radius of a single molecule were not justified. In fact, he was primarily interested in the molecular polarizability (α_{mol}) and used the formula

$$\alpha_{mol} = \left(\frac{n_D^2 - 1}{n_D^2 + 2}\right)\frac{3V}{4\pi N} \tag{17}$$

where V is the volume of the liquid sphere considered, and N the number of molecules it contains. In eq 17, the rightmost factor corresponds to the cube of an effective single spherical molecule radius. It is however understood that the same molecular polarizability is obtained as long as the V/N factor is preserved and is therefore size independent with the key assumption that ε_∞ is filling the space uniformly, i.e. that it is a spatially averaged value. In order to calculate the refractive indices for the general case where the internal dielectric is not uniformly distributed in the liquid, we generated pure liquid configurations from molecular dynamics (MD) simulations at room temperature and cut out spherical clusters (or droplets) from individual snapshots. We maintained the V/N ratio by fixing the density to experiment and calculating the droplet effective ε_{in} with the formula

$$n^2 = \frac{R_{droplet}^3 + 2\alpha_{droplet}}{R_{droplet}^3 - \alpha_{droplet}} \tag{18}$$

where $R_{droplet}$ and $\alpha_{droplet}$ are the droplet radius and polarizability. We assigned the dielectric function on all molecules and applied the procedure outlined above to calculate the droplet polarizability and thereby access the droplet refractive index. Averaging the droplet refractive index over many droplets yields an approximation of the bulk refractive index.

2.3.2. Computational Details. To obtain the liquid phase droplets, molecular dynamic simulations, using the AMBER 8.0 package, were performed on 3375 molecules (15 × 15 × 15) in a cubic box. The NVT ensemble and periodic boundary conditions allowed the density to be fixed to the experimental value. The temperature was set to 20 °C to match the experimental conditions used to report refractive indices and maintained constant with the Berendsen's weak coupling algorithm[32] with the kinetic energy adjusted every 1 ps. The short-range nonbonded interaction cutoff was set to 8.0 Å, and long-range interactions computed with particle mesh Ewald[14,33] using the default Amber 8.0 setup. The molecules were charged with AM1-BCC,[34,35] and the Generalized Amber Force Field (GAFF)[36] was used. The SHAKE procedure[37] was used to fix all bond lengths to hydrogen.

The initial liquid box was generated by positioning the molecules on a cubic lattice, randomly oriented with the Marsaglia[38] quaternions method. The system was first minimized until the root-mean-square (rms) of the gradient is less than 0.1 kcal/mol/Å. This was followed by a 8 ps annealing phase integrated by steps of 1 fs, during which the nonbonding interactions were gradually turned on and the temperature increased from 0 K to 40 K and decreased to 0 K. The system was then heated over 20 ps up to 293.15 K with a 2 fs integration time step. Following a 1 ns equilibration, 50 evenly spaced snapshots were written over a 2 ns production run. Each of the liquid boxes for a given molecule was then wrapped in the primary cell. A sphere with a diameter set to 85% of the box length formed a liquid droplet when picking all molecules with an atom lying inside the sphere. The droplet radius was then determined by considering the position of the outermost non-hydrogen atoms. The precise definition of the radius is not unique, and we have verified, for example, that using the experimental density to calculate the radius of the corresponding ideal sphere gives refractive indices within ±0.01 of those obtained by the chosen algorithm. Also, this model assumes a perfectly spherical object, ignoring the dimples formed because of the finite size of the spheres. The relatively large size of the droplet and the averaging over 50 independent configurations reduced the effect of this approximation.

The solution to Poisson's equation in the presence of the voltage clamp boundary conditions was obtained on a rectangular grid sized to encompass the full droplet plus half its radius on each side of the droplet. The target grid spacing was set to be 0.5 Å. The smooth dielectric functions (eq 1), fitted on the molecular polarizability tensors only, were assigned together with the matching atomic polarization radii, internal dielectric ε_{in}, and A parameter. The external dielectric was always set to the vacuum value $\varepsilon_{ext} = 1$. The convergence criteria for the ZapTk solver was based on the grid energy and set to 0.0001 kT. This convergence criteria required the assignment of atomic charges that we choose to be $\pm 0.001e$ were randomly placed on half the atoms, keeping an overall neutral system. Given the strength of the external field applied, this was not perceptibly affecting the answer.

2.4. Free Energy of Hydration. *2.4.1. Theory.* Implicit solvent models are commonly used to incorporate the effects of solvation in molecular models as a mean field.[39−43] These models considerably reduce the computational burden needed to sample the solvent configurational space when each atom of the solvent are explicitly simulated. An important valida-

Table 1. Reported Optimal Polarization Radii (σ_{in}) and Atom Typing for the Four G1 Sets Defining the Internal Dielectric (Eq *1*)

SMARTS[23−25]	typical functional groups	radius (Å)			
		G1−4[a]	G1−9[a]	G1−12[a]	G1−24[a]
ε_{in}[b]		4	9	12	24
A[b]		10	5	10	4.19
	H				
[H]	all H	0.83	0.65	0.55	0.52
	C				
[CX4]	alkanes	0.78	0.79	0.67	0.62
[c,CX2,CX3]	aromatic, sp, sp^2	1.25	1.02	0.87	0.78
	N				
[n,NX1,NX3,$(Nc),$(NN)]	aromatic, nitriles, sp^3, aniline, hydrazine,	1.09	0.89	0.76	0.69 0.74[c]
[$(N[C,S]=*)]	amides, amidines, sulfonamides	0.89	0.77	0.64	0.58
[$(N=C)]	imine, amidine	1.07	0.93	0.81	0.76
[$([#7]∼[OX1])]	N-oxides, nitro	0.00	0.79	0.68	0.59
	O				
[$([OX2]([H])[#6,#7]),o,$([OD2]([CX4,c])[CX4]), $(O=[c,C,S])]	alcohols, furan, hydroxamic acids, ethers, ketones, aldehydes, amides, sulfones	0.88	0.73	0.63	0.60
[$(OC=[O,N])]	esters, carboxylic acids	0.68	0.55	0.46	0.46
[$([OX1]∼[#7])]	N-oxide, nitro	1.08	0.89	0.77	0.74
	Others				
[S,s]	all sulfur atoms	1.44	1.22	1.06	1.01
[F]		0.77	0.62	0.53	0.51
[Cl]		1.30	1.09	0.95	0.91
[Br]		1.47	1.24	1.07	1.03
	Water				
	special fit				
[$([OX2]([H])[H]]		0.93	0.86	0.76	0.75
[$([H][OX2][H])]		0.64	0.45	0.36	0.31
	Charged Atoms				
[$([#1][#7+]),$([#1][#7][#6]=[#7+][#1]), $([#1][#7][#6]=[#7+]),$([#1][n+]∼c∼n), $([#1]n∼c∼[n+])]	proton in guanidiniums, amidiniums, ammoniums, pyridiniums	0.44	0.43	0.37	0.01
[$([O-]C=O),$(O=C[O-])]	O in carboxylates	1.20	1.02	0.88	0.85
[$([NX4+]),$([#7+]=C−N),$(N−C=[N+]))]	N in ammoniums, guanidiniums, amidiniums,	0.00	0.34	0.39	0.52
[$([n+]∼c∼n),$([n]∼c∼[n+])]	N in imidazoliums	0.00	0.00	0.00	0.42

[a] Model name. [b] Parameter kept fixed during optimization. [c] Nitrile nitrogen radius made different for G1-24.

tion for solvation models comes from the experimental free energy of hydration (ΔG_{hyd}) that consists in the chemical potential difference for the transfer of a solute from vacuum to bulk solvent. The computational evaluation of ΔG_{hyd} is separated into two processes. First, the nonpolar free energy of hydration (ΔG_{np}) comes from the formation of the solute-shaped cavity in the bulk solvent that causes a reorganization of the solvent molecules and nonpolar interactions between the solute and the solvent. Second, the electrostatic free energy of hydration (ΔG_{elec}) results from the electrostatic work necessary to place the solute charge density in the solute cavity, involving interactions between solute and solvent charge densities and their response to one another. This results in the equation

$$\Delta G_{hyd} = \Delta G_{elec} + \Delta G_{np} \tag{19}$$

The long-standing use of implicit solvent to evaluate ΔG_{elec} is based on a high continuum dielectric solvent region that gets polarized by a static solute electric field. While the solute cavity is traditionally formed with a molecular surface with a discrete transition of the dielectric function at the solute−solvent boundary, we chose a smooth boundary transition as explained earlier. The solute cavity volume and shape is determined by atomic radii. For a given set of charges, atomic radii that are too small exaggerate the affinity of the solute for water, while radii that are too large will have the opposite effect. The calculation of ΔG_{elec} is normally done with a nonpolarizable solute, or, if the cavity is assigned a $\varepsilon_{in} > 1$, the very significant screening of the atomic partial charges requires a special treatment that was not done until recently.[2] For nonpolarizable solutes, knowing that water increases the dipole moment of solvated molecules often by as much as 15%, the atomic charges should not be fit on a gas phase QM ESP. For this reason, the charges are often generated from RESP[44] or AM1-BCC[34,35] that are known to be sufficiently overpolarized compared to the gas phase.

In the 3-zone dielectric model that we propose in this article (cf. Figure 2a and eq 3), the first zone should accurately account for the solute polarizability, which allows for the use of vacuum phase atomic charges obtained taking into account the internal dielectric function. The second zone located between the internal dielectric and the solvent is set to vacuum, and the transition to the full implicit solvent model of the third zone needs to be parametrized. Following the suggestion of Grant et al.[18] for their nonpolarizable 2-zone dielectric function, we fixed the B parameter in eq 3

to 11.8, which leaves the solvent cavity atomic radii to be fitted on the experimental free energy of hydration. However, in order to compare the calculated ΔG_{hyd} to experiment, we needed to use existing values or methods for ΔG_{np}. Fortunately, converged molecular dynamics free energy[45,46] calculations based on free energy perturbation (FEP) calculations are available for each compound from our hydration free energy data set. We feel this is the best achievable theoretical estimation of ΔG_{np}, so this is our preferred estimation in current study. However, since this is not very useful for prospective evaluations of ΔG_{hyd}, due to the heavy computational demands for such FEP calculations, we also tested a surface area based model that calculates ΔG_{np} as

$$\Delta G_{np} = \gamma \times S \tag{20}$$

where γ is a surface tension, and S is the surface area of the molecule as defined by a solvent accessible surface[17] created with a 1.4 Å rolling probe and the Bondi radii.[16] This crude approximation has proven useful, and it can be improved upon by atom typing the γ[47] or by using some treatment of the dispersion energy[48−51] instead; however, in this work a single value of γ was fitted for each model.

2.4.2. Computational Details. The atomic partial charges responsible for the permanent electrostatic potential (ESP) were determined by a least-squares-fit on the QM ESP calculated on a face-centered-cubic grid of points. Following Jakalian et al.,[34] the grid spacing was set to 0.5 Å, and the grid points were positioned around the molecule in a volume formed by two vdW surfaces, each built with Bondi radii scaled by a factor of 1.4 and 2.0. The dielectric scales down by a factor of $1/\varepsilon_{in}$ the effect of the charges; this is partly compensated by the bound charges appearing from the internal polarization. Hence, the least-squares-fit requires a Poisson solver in order to capture the overall effect, which depends on the shape of the dielectric boundary. It is noteworthy that the EPIC polarizability model is independent of the charge fitting process; as a result, charges are fitted after the solute dielectric parameters are optimized. The details of the procedure, called DRESP, can be found elsewhere.[2]

A finite difference Poisson solver was written to allow the implementation of the 3-zone dielectric model. Here is a brief description of the algorithms implemented. We use successive over-relaxation (SOR) and a Gauss-Seidel iterative scheme[52,53] where the over-relaxation parameter w is estimated by

$$w = \frac{2}{1 + \sqrt{1 - \lambda_{\max}^2}} \qquad \lambda_{\max} = 1 - \frac{\pi^2}{2(N-1)^2} \tag{21}$$

where N is the number of grid points in one of the dimension of the grid.[52] This crude estimate of the spectral radius of the A matrix in the finite difference form of the Poisson's equation used (see the Appendix of ref 18) was sufficient to reduce by a factor of approximately 30 the number of Gauss-Seidel steps necessary.

The free charges of the system were assigned on the grid with a quadratic inverse interpolation scheme[18] that has the advantage of conserving the dipole moment, has a continuous first derivative, and is more robust to the effects of rotation and translation. The same interpolation rule is used to calculate the potential in between grid points. In our calculations, we use a convergence criteria base on grid energy defined as the sum of the electrostatic potential times the distributed free charges on the grid. This convenient criterion is directly related to the energy in an absolute way and thus ensures that relative energies are also converged. The boundary conditions, in energy calculations, were determined with a Coulomb potential.

The ΔG_{elec} was computed by taking the grid charge energy difference between a solution obtained in vacuum ($\varepsilon_{ext} = 1$) and another solution in water ($\varepsilon_{ext} = 80$) from the resulting Poisson's equation and calculated with

$$\Delta G_{elec} = \frac{1}{2} \sum_{i}^{Atoms} q_i(\varphi(\vec{r}_i)^{water} - \varphi(\vec{r}_i)^{vacuum}) \tag{22}$$

where q_i is the atomic partial charge of atom i, and $\varphi(\vec{r}_i)^{\text{vacuum}}$ is the interpolated electrostatic potential at atom i position $\vec{r}_i$. The grid spacing for the solver was set to 0.35 Å, and the minimum distance between the solute internal radii and the grid boundary was set to 7 Å. In those cases where the solute was nonpolarizable, ε_{in} was set to one. Finally, the parameters (solvent cavity atomic radii and surface tension) were adjusted with the same Levenberg−Marquardt algorithm used for the fit to the polarizability tensor. All parameters were simultaneously optimized.

2.5. Quantum Calculations. The B3LYP exchange-correlation functional[54,55] was used for all DFT quantum calculations of this work within the Gaussian 03 software.[56] All molecular structures of this work were initially relaxed with B3LYP and the 6-31++G(d,p) basis set.[57−59] Property calculations required larger basis sets for accuracy. The electrostatic potential values were obtained with B3LYP and the 6-311++G(3df,3pd) extended triple-ζ basis set.[57−59] The molecular polarizability tensor computations used the aug-cc-pVTZ basis set,[60] as it was shown to lead to accurate results.[61] The implemented method in Gaussian 03 to calculate the molecular polarizability tensor is the Coupled Perturbed Hartree−Fock (CPHF) method.[62] The Hartree−Fock calculations performed to fit water-adapted atomic partial charges were also performed with the Gaussian 03 software with the 6-31G(d,p) basis set.

3. Data Sets

In this work, we made extensive use of three kinds of data: B3LYP/aug-cc-pVTZ polarizability tensors, free energies of hydration, and refractive indices. A total of five data sets were created.

3.1. Polarizability Training Data Set (PTD). A training data set was used to optimize the internal radius in order to match B3LYP polarizability tensors. To this end, we made use of the previously published training data sets[1] and added new molecules for a total of 265 polarizability tensors. In this data set, many neutral functional groups are represented:

alkanes, alkenes, alkynes, halogens (bromo, fluoro, chloro), alcohols, thiols, amines, ethers, thioethers, nitriles, aldehydes, ketones, esters, thioesters, amides, acids, ureas, imines, amidines, sulfones, sulfoxides, sulfonamides, heteroaromatics, hydrazines, hydroxamic acids, N-oxides, pyridones, and peptides. In addition, charged functional groups were also included with the sole purpose of examining charged side chains in amino acids. They were carboxylates, guanidiniums, imidazoliums, and ammoniums. The strength of the polarizability training data set is in the wide coverage of functional groups, but its weakness is the lack of polyfunctional molecules. To get this level of coverage would require calculations on a great many more larger molecules and consequently an enormous amount of computational power. The intention in this paper is to assess whether a small and reasonably general first set of parameters can adequately treat a wide variety of bioorganic small molecules in addition to most biomolecules.

3.2. Polarizability Validation Data Set. The polarization validation data set is composed of the previously published validation sets[1] and 401 molecules from the hydration free energy data set (below) not included in the polarizability training data set. In addition, a few special molecules such as neutral and charged peptides, melamine, sugars, etc. were added, giving a total of 442 molecules.

3.3. Polarizability Data Set. The polarizability data set is the combination of the validation and training data sets, making available all 707 polarizability tensors together with the molecule coordinates (see the Supporting Information).

3.4. Hydration Free Energy Data Set. This data set is built from a compilation of 504 experimental free energies of hydration of neutral molecules recently published with the corresponding ΔG_{np} and ΔG_{chg} from Molecular Dynamics based absolute free energy calculations.[45] We took the published data set, eliminated the iodine- and phosphorus-containing compounds, and formed a data set of 485 molecules on which we could fit the solvent part of the dielectric function (eq 3) and the surface tension coefficient (γ).

3.5. Refractive Indices Data Set. The refractive indices data set contains 23 small organic molecules (cf. Figure 5) that are liquids at 20 °C, for which the density and the refractive indices are taken from the CRC Handbook of Chemistry and Physics.[63] They span a variety of functional groups, and most of the entire spectrum of refractive indices measured for bioorganic molecules.

4. Results and Discussion

4.1. Polarizability Tensor. This work follows the precedent of ref 1 in fitting atomic polarization radii and a single inner dielectric constant to QM molecular polarizability tensors to produce an accurate EPIC model of electronic polarization. This is done independently of the permanent electrostatic potential; all atomic partial charges are therefore set to zero. In this section, we generalize the parametrization to account for most of the biomolecules and a significantly wider spectrum of bioorganic functional groups. In contrast

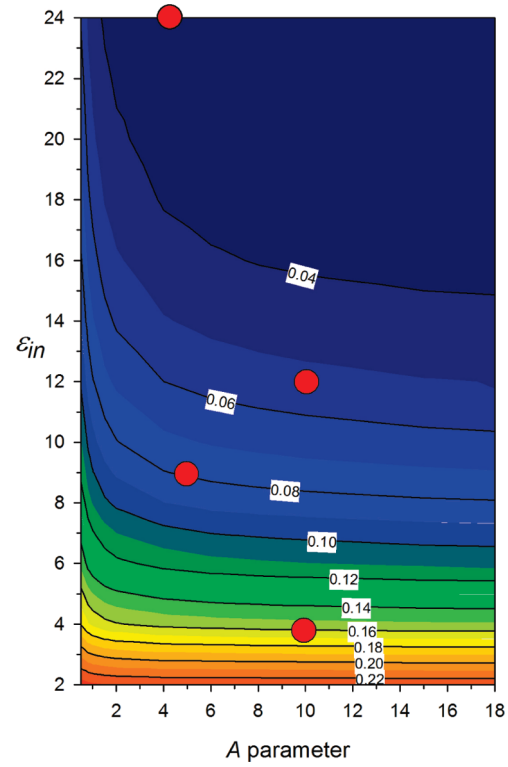

Figure 3. The iso-contour plot of the RRMS error between B3LYP/aug-cc-pVTZ and EPIC polarizability tensors are shown as a function of the ε_{in} and *A* parameters of eq 1. This RRMS surface was generated from a simultaneous fit of the H, alkyl C, aromatic C, and aromatic N atomic polarization radii on training set of 11 aromatic and 14 alkane molecules against their B3LYP polarizabilities. It shows that in order for a single dielectric model to fit the polarizabilities of these two chemical classes to within 10% error, the ε_{in} needs to be sufficiently large (>6). Deviations in the anisotropy of the polarizability are the main source of error for lower values of ε_{in}.

to our previous work, we use a smooth dielectric function as described earlier and a single internal dielectric (ε_{in}) value.

4.1.1. Choice of ε_{in} and A Parameters. It was previously shown that a more accurate polarizable model was obtained when different ε_{in} were fitted for alkanes and aromatics. However, the single-ε_{in} model performed as well as the multi-ε_{in} model and DFT against experimental directional polarizabilities. Furthermore, in another study[2] that examined the local electronic polarization, the same single-ε_{in} model was only slightly worse than the multi-ε_{in} model. In this work we pursue the single-ε_{in} model because it greatly simplified the Poisson solver implementation and the robust parametrization for a wide spectrum of bioorganic chemistry.

Before the global parametrization of polarizability atomic radii, a range-finding study was performed with a smaller training set examining which combination of ε_{in} and *A* (cf. eq 1) is best to use for extending the EPIC parametrization previously initiated.[1] We used a set of 13 alkanes (set g in ref 1) including methane, propane, cyclopropane, butane (cis, trans), hexane (cis, trans), and neopentane, together with a set of 10 heteroaromatic molecules (set a in ref 1). We formed the two-dimensional grid of ε_{in} and *A* pairs and optimized four radii (hydrogen, alkane carbon, aromatic

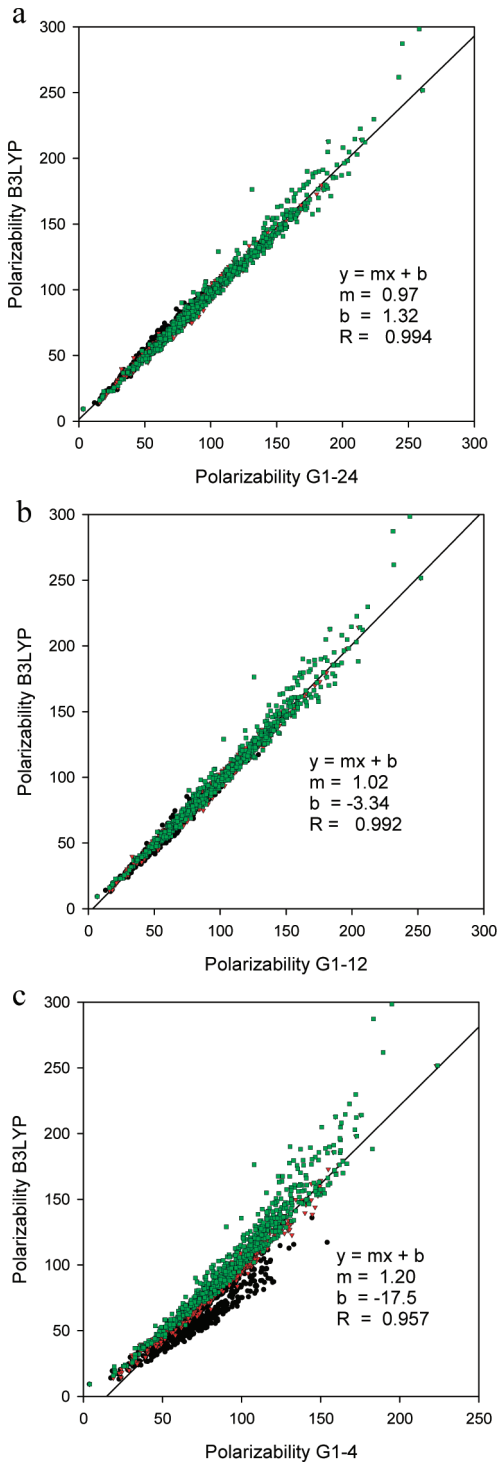

Figure 4. Correlation graph between the B3LYP/aug-cc-pVTZ directional polarizabilities (α_1 black circles, α_2 red triangles, and α_3 green squares in au) for three G1 dielectric parameter sets (cf. Table 1). Each figure shows the data for 707 molecules for a total of 2121 points. From these figures, it is clear that a small number of parameters (optimized on 265 molecules) can generalize well. A large $\varepsilon_{in} = 24$ (a) produces the best fit, a medium range $\varepsilon_{in} = 12$ produces slightly larger discrepancies, and a small $\varepsilon_{in} = 4$ produces significantly larger deviations, in keeping with the results of the range-finding study on the small data set.

carbon, and aromatic nitrogen) for each point of the grid. The polarizability tensor RRMS deviation from QM for this

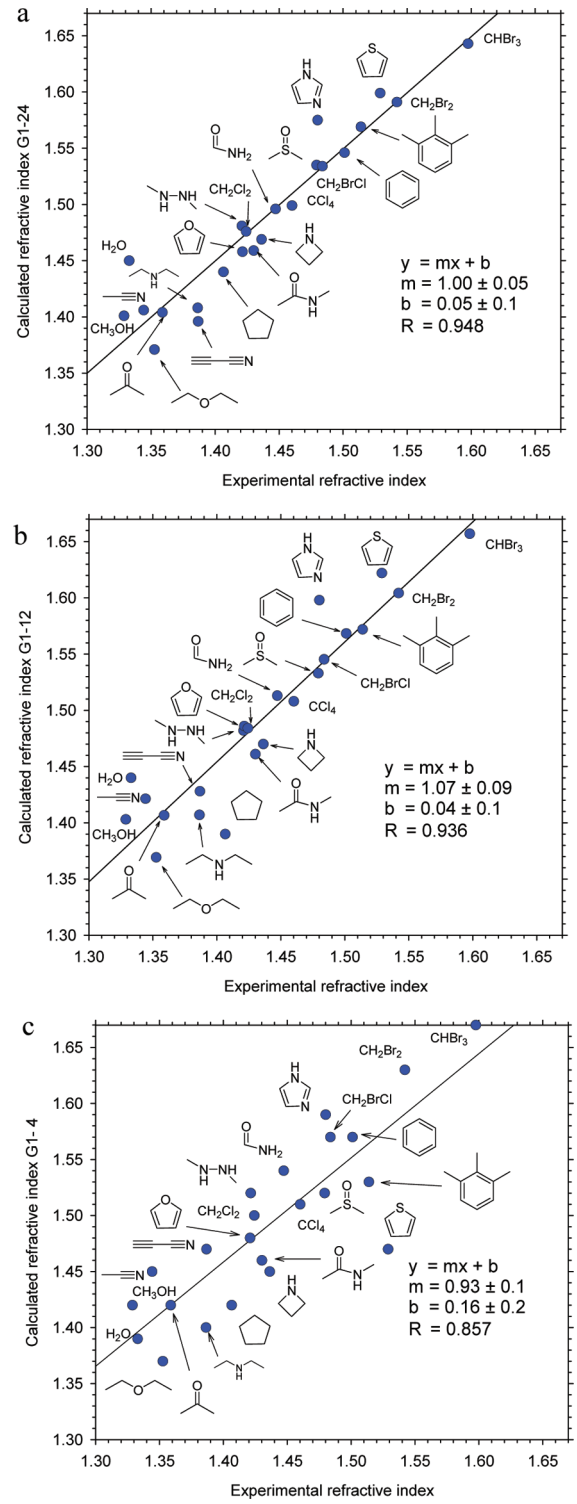

Figure 5. The calculated refractive indices (n) of 23 organic molecules are compared to experiment. Three dielectric parameter sets are used a) G1−24, b) G1−12, and c) G1−4 (Table 1). For each set, the preoptimized radii can be found in Table 1. The reported refractive indices (n) were obtained by polarizing a liquid droplet formed by carving spheres from periodic MD liquid simulation snapshots. The Clausius-Mossoti equation leads to $n^2 = \varepsilon_\infty$ close to experiment, in spite of the large ε_{in}. The predicted values are systematically higher than experiment, which can be explained by potential artifacts or a polarizability shift when passing from vacuum to condensed phase (see text). As with the polarizabilities, the predictions deteriorate with decreasing ε_{in}, in keeping with the results of the range-finding study on the small data set.

data set at each (ε_{in}, A) pair is shown as an iso-contour plot in Figure 3. It is clear that in order to fit a general dielectric function, a sufficiently large ε_{in} is needed. Also, the flatness of the error surface allows for multiple equivalent choices, a potential advantage if other criteria become more stringent in the development of the polarizable model. As shown by red circles in Figure 3, four starting points were selected for further examination: G1−24 ($\varepsilon_{in} = 24$, $A = 4.188$), G1−12 ($\varepsilon_{in} = 12$, $A = 10$), G1−9 ($\varepsilon_{in} = 9$, $A = 5$), and G1−4 ($\varepsilon_{in} = 4$, $A = 10$). In the case of G1−24 only, the A parameter was relaxed to a value of 4.18. The G1−12 seems slightly superior to the G1−9. Finally, while the G1−4 parameter set showed the worst RRMS, it was still a good case for having a small value of ε_{in}, identified by Tan and Luo[13] as being optimal. Each of the G1 ε_{in} and A choices was fixed in the global parametrization of atomic polarization radii described below. Finally, Figure 3 shows that making a poor selection of (ε_{in}, A), in particular having $\varepsilon_{in} < 6$, cannot be redeemed by adjusting either the radii or the A parameter.

4.1.2. The Optimized Polarization Radii. The parametrization of the four G1 sets on the 265 molecules of the polarizability training data set proceeded as described in the Method section. The ε_{in} and A values were fixed, and the atomic polarization radii σ_i were adjusted to optimize the fit to the B3LYP polarizability tensors. The atom typing of the radii was a primary concern, and we aimed at minimizing the number of radii fitted to reduce the fitting complexity, ensuring a better generalization of the chemistry. Each nonsymmetric molecule produced 6 data points from their polarizability tensor; structurally symmetric molecules produced fewer data points. The number of fitted parameters was kept small compared to the number of associated data points to prevent overfitting. The determination of the atom typing was done iteratively by hand: first, the polarizability training data set was designed in terms of chemical functional-group classes. Adjustable parameters were added gradually with new molecules having unmet chemical functionalities. Often, the addition of a new chemical functionality class led to one or two additional parameters. We also merged atom types when the radii values were similar and the fitness metrics (χ^2, δ_{avg}, and δ_{aniso}) were not significantly affected. For example, the alkane H and C radii were the first to be fitted. This was followed by aromatic C, H, and N. It was determined early that a single aromatic and alkyl atom type for C and H could be utilized. Then the alcohol oxygen radius, halogen radii, alkene carbon, and alkyne carbon radii were individually fitted. The final stage was a global simultaneous fit of all radii with the entire polarizability training data set. Because of its special importance as a solvent, water was treated separately with its own special O and H radii.

The resulting polarization radii are given in Table 1 for the four G1 parameter sets. The ordering of atom types in Table 1 is important since the atom typing was done in sequence from top to bottom to deal with the issue of a particular atom falling in more than one category (H for instance). The first observation is that all polarization radii are significantly smaller than vdW contact radii such as Bondi,[16] Pauling,[64] or Parse[40] often used in Poisson−Boltzmann approaches. This finding unveils the two different natures of the physical phenomena described. On the one hand, the polarization radii aim at calibrating how the electrons polarize in reaction to an external field created, for example, by an interacting molecule. On the other hand the vdW radii determine the position of the repulsive molecular wall toward other molecules. It is also expected that the larger the ε_{in}, the smaller will be the radii: to maintain the overall polarization the dielectric must increase as the radii decrease. This is illustrating a general feature of the model that produces larger polarizabilities when either the 'electronic volume', decided by the radii, or the internal dielectric increase. The sort of relationship involved is given in eq 5 above for a hard sphere and elsewhere for a diatomic.[1]

It is also interesting to compare polarization radii between elements and between the different chemical environments. First, it is remarkable that the diverse carbon atom contexts can be covered by only two atom types: sp^3 and others. The smaller sp^3 carbon radius implies that carbon makes a much smaller contribution to the overall polarizability when sp^3 hybridized than when pi electrons are involved, i.e. in the sp or sp^2 hybridization states. This can be rationalized by the presence of π^* molecular orbitals, the different number of connected H atoms, and the difference in the molecule shape and the related anisotropy.

The nitrogen atoms were subdivided into four atom types among which two encompass almost all instances in the data sets. The first of these is a general nitrogen type assigned to amines, nitriles, hydrazines, or anilines for example. The smaller second major nitrogen radius makes amide, amidine, or sulfonamide nitrogen less polarizable. Surprisingly, the more specific nitro and N-oxide nitrogen radius, in the G1−4 set, has a radius of zero. The dielectric on this nitrogen atom is only slightly smaller than ε_{in} because of the large bound oxygen radii and the short N−O bond, typically 1.2 Å, which allows dielectric from the oxygen to spread over onto the nitrogen. It is also interesting to note that in the G1−24 set, there was a gain in accuracy when the nitrile nitrogen had its own radius.

The oxygen atom behavior can mainly be accounted for by two adjustable radii types, which was a significant advantage in the fitting process—the N-oxide and nitro functional groups still being an exception. Another interesting result is the large radius of the sulfur atom that is comparable to the bromine radius. However, it is not to say that the polarizability contribution of sulfur is equivalent. In fact, the bromine bonds are longer and hence offer a larger polarizable volume. This argument is also useful to explain why the fluorine radius is smaller than the hydrogen radius. For example, the model predicts a polarizability for tetrafluoromethane of 18 au compared to 17 au for methane and a polarizability of 76 au for hexafluorobenzene compared to 70 for benzene, all in close agreement with B3LYP. Because of water's special importance as a solvent, both the oxygen radius and the hydrogen radius were optimized to exactly match the B3LYP polarizability tensor. These various behaviors of the atomic polarization radii underscore their

Table 2. Error Obtained with the Optimized Polarization Radii of the G1 Sets When EPIC Molecular Polarizability Tensors Are Compared to B3LYP for Different Molecule Data Sets

model[a]	δ_{avg}[b] (%)	δ_{aniso}[b] (%)	RRMS[b] (%)
Polarizability Training Data Set: 265 Molecules			
G1−4	5.0	20.9	12.7
G1−9	3.2	9.1	6.7
G1−12	2.9	5.3	5.0
G1−24	2.3	5.2	4.4
Polarizability Validation Data Set: 442 Molecules			
G1−4	4.0	18.2	12.3
G1−9	2.7	7.6	6.7
G1−12	2.6	5.1	5.3
G1−24	2.1	5.4	4.6
Polarizability Data Set: 707 Molecules			
G1−4	4.4	19.2	12.4
G1−9	2.9	8.2	6.7
G1−12	2.7	5.2	5.2
G1−24	2.2	5.4	4.6

[a] Model using the parameters given in Table 1. [b] Cf. section 2.2.3.

difference in nature and purpose from vdW contact radii, which is why they must be treated differently.

Finally, charged species pose a special challenge that we decided to address specifically for charged side chains in proteins: Arg, Lys, Asp, Glu, and His. Further generalization of the radii for charged species while retaining the same level of accuracy in the polarization tensor would require a more extensive parametrization. One reason for this is the expected reduction in polarizability on the neighbor atoms through the strong induction caused by the charged site. On the other hand, the electrostatic interactions around charged centers will be dominated by the monopole (i.e., the distribution of the charge itself), so high accuracy in the effects of polarization may become less important than with neutral species.

4.1.3. Polarizability Tensors. The G1 parametrizations clearly showed the capacity of EPIC to produce accurate polarizabilities with a minimum of atom types. The choice of ε_{in} and A combinations made based on the very small range-finding subset showed the same behavior in the polarizability training data set, the polarizability validation data set, and their combination (polarizability data set), made of 265, 442, and 707 molecules, respectively. Table 2, which summarizes the errors, shows the accuracy of the obtained models. The G1−24 data set has an unsigned average error of 2% on the average polarizability (eq 13) and a 5% error on the anisotropy of the tensor (eq 14). With the point inducible dipole polarizable models, such a low level of error was obtained only when anisotropic atomic polarizabilities were fitted,[5,15,65−67] making their generalization very challenging. The other G1 models are worse, and as predicted from the range-finding study results shown in Figure 3, the G1−4 set is inadequate to reproduce the directional difference in the polarizability (the large δ_{aniso} values in Table 2). That the error obtained on both the training data set and the validation data set was similar indicates that our radii are not overfit. Finally, the three directional polarizabilities (eigenvalues of the tensor) obtained for the 707 molecules (2121 data points) are compared to the corresponding B3LYP values in Figure 4 for three representative G1 sets. The excellent correlation is obvious for the G1−24 and G1−12 and deteriorates in the G1−4 EPIC model (the Pearson correlation coefficients are 0.99, 0.99, 0.96 and the slopes 0.97, 1.02, and 1.20, respectively). An apparent outlier is the α_3 (longitudinal polarizability) of (3E)-hexa-1,3,5-triene for which B3LYP gives a value of 176 au compared to the EPIC value of 125 au. For this specific molecule, Sekino et al.[68] showed that B3LYP greatly overestimates the α_3 value of acetylene chains. Their better estimate, based on very accurate CCSD and MP2 QM results, predicts a value of ∼135 au, close to the EPIC value. Another remarkable discrepancy between EPIC and B3LYP is observed in Figure 4 for the α_3 of 1,4-dioxidopyrazine (doubly oxidized nitrogen on pyrazine) that is predicted to be 103 au by the G1−12 model versus 129 au by B3LYP. A similar observation can be made for 4-nitroaniline. Although we have not found better estimates for these molecules, they most certainly constitute a challenge both for classical and *ab initio* polarizability calculations.

4.2. Refractive Indices. In the previous subsection, we have developed dielectric functions that predict remarkably well, relative to QM, the polarizabilities of a single molecule in the gas phase. In this section we present the *macroscopic* refractive index calculations and the corresponding effective high frequency limit dielectric (ε_∞). In a previous publication,[1] we proposed that the vacuum of the intermolecular spacing may be sufficient to reduce the effective macroscopic ε_∞ resulting from the high intramolecular ε_{in} obtained in the optimization to polarizability tensors. Here we use a theoretical approach to verify this hypothesis. Another important point addressed by the refractive index calculation is the transferability of the dielectric function from the gas phase to the condensed phase.

As explained in further detail in the Theory and Method sections, we form liquid droplets containing thousands of molecules from snapshots obtained by MD simulations and calculate the effective ε_∞ by the use of the Clausius-Mossoti equation. The small range spanned by experimental refractive indices makes this test somewhat stringent. Figure 5 shows the correlation between the results obtained with three representative EPIC parametrizations and experiment; G1−9 is omitted here and for the remainder of the article because the results are so similar to those of G1−12. The first observation is the close agreement between the magnitudes of the ε_∞ values. This clearly demonstrates that the effective ε_∞ of the liquid droplets have the appropriate value in spite of the high ε_{in}, confirming our hypothesis. Figure 6 provides a visual explanation for the apparent mismatch between the low effective ε_∞ compared to the high ε_{in}. This figure shows the molecular dielectric inside a CCl_4 droplet when it is sliced through its center. The G1−24, G1−12, and G1−4 models have a quite variable low-dielectric intermolecular space. The coloring scheme of the dielectric function (eq 1) assigns red when $\varepsilon(r) = \varepsilon_{in}$ and dark blue when $\varepsilon(r) = 1$. The low dielectric intermolecular space increases with ε_{in} as the atomic radii decrease. It is striking that these three parametrizations produce the same refractive index and the same molecular polarizability in spite of the very different ε_{in}. If

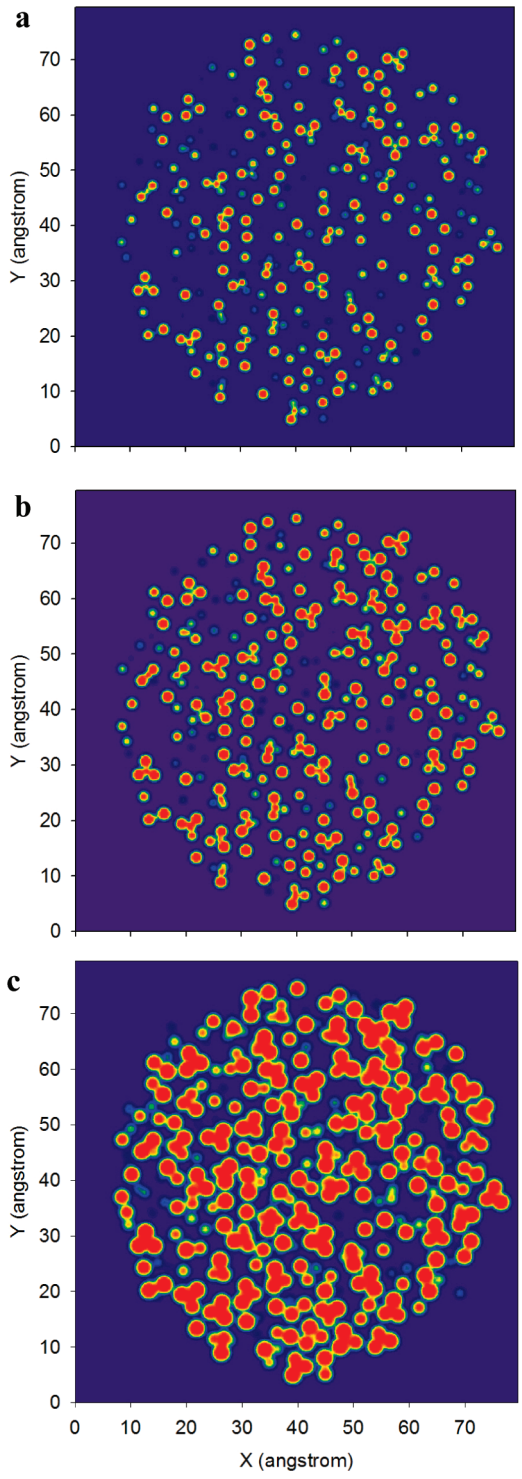

Figure 6. One of the 50 CCl_4 droplets is cut in its center, and three dielectric functions (eq 1) are plotted: a) G1−24 b) G1−12, and c) G1−4. The red color is associated with $\varepsilon(r) = \varepsilon_{in}$ and blue with $\varepsilon(r) = 1$.

ε_{in} is further reduced below 4, the whole droplet will become filled with a uniform dielectric (as the atomic radii increase and start to overlap), and the simultaneous prediction of the molecular polarizability and the refractive index will become compromised.

Also noticeable in Figure 5 is that the correlation with experiment follows the previous assessment of the models based on molecular polarizabilities: the G1−24 parametrization (Figure 5a) has a $R = 0.95$, slightly better than the G1−12 (Figure 5b) with $R = 0.94$, which is in turn significantly better than the G1−4 correlation with $R = 0.86$ (Figure 5c). However, Figure 5 shows a 0.05 systematic overestimation of the refractive indices in all cases which could correspond to a small overpolarization, a result not reflected in the gas-phase polarization tensors. The source for this deviation is not clear, but we have several hypotheses. First, the Clausius-Mossoti equation is valid for a perfect sphere, whereas we are dealing with an imperfect surface created by nanoscopic droplets. Second, we have verified that an underestimation of the droplet radius by only 3−4% (1 Å in a range of 25−35 Å) could systematically shift the calculated refractive indices by 0.05. Third, it is also possible that the liquid phase polarizability may be truly smaller than the predicted gas phase polarizability since a drop of 11% of the polarizability could explain the 0.05 shift. This would be in agreement with other studies that found similar phenomena[8,11,69] and based their reasoning on the increased Pauli exchange repulsion from the closer contact of the molecules in condensed phase. However, the magnitude of this effect differs considerably from study to study.

4.3. Hydration Free Energies. The calculation of hydration free energies is aimed at assessing whether the dielectric polarization model capable of accurately reproducing gas-phase polarizability tensors can be used "as is" in implicit solvent calculations. Because of the difference in nature and behavior between the atomic polarization radii and the atomic cavity radii used for the solute−solvent boundary, the 3-zone dielectric model is required. The hydration free energies were calculated with the G1−24, G1−12, and G1−4 polarizability models found in Table 3. For each polarizability model, charges were fitted to the vacuum-phase QM ESP as described in section 2.4.2; thus neither the various G1 molecular polarizability parameters nor the atomic charges used for the hydration calculations have been influenced by any effects of solvent. As discussed in ref 2, the high internal dielectric screening of the G1 polarizability causes the DRESP fitted charges to be of markedly higher magnitudes than if they were fitted using an internal dielectric of 1. By the same token, the differences between the various G1 models also result in different charges for each model. A comparison of the resulting G1−24, G1−12, and G1−4 charges for several example molecules and the G1−12 charges for the entire hydration data set are given in the Supporting Information. With the atomic charges in hand, each of the solute models was then used to optimize the solvent cavity atomic radii (referred to as cavity radii and noted G2 in what follows) referred to in the second Gaussian summation in eq 3. Each set of cavity radii is thus associated with a given charge set and molecular polarizability set, for example G2−12 is the set of cavity radii associated with the G1−12 molecular polarizability parameters and the charge set fitted using the G1−12 parameters. We decided to set $B = 11.8$ in all calculations, following the Grant et al.[18] suggestion as it was found to make the Bondi radii[16] optimally reproduce the hard dielectric boundary results with the same smooth boundary as used in this work. The results reported in Table 3 are split into two main categories based

Table 3. Solvent Cavity Atomic Radii (σ_{cavity}) and γ for the 3-Zone Dielectric Model Optimized on 485 Experimental Free Energy of Hydration with Different G1-n Solute Models and ΔG_{np} Sources

	model[a]					
	G2-HF	G2−4	G2−12	G2−24	G2−12SA	
	solute					
charges[b]	HF[c]	B3LYP[d]	B3LYP[d]	B3LYP[d]	B3LYP[d]	
ε_{in}[e]	1	4	12	24	12	
A[e]		10	10	4.19	10	
ref. Table 1		G1−4	G1−12	G1−24	G1−12	
ΔG_{np}	FEP[f]	FEP	FEP	FEP	SA[g]	
B	11.8	11.8	11.8	11.8	11.8	
	optimized implicit solvent parameters					Bondi
H[h]	0.98	0.95	0.97	1.02	0.98	1.20
C	1.95	2.03	2.02	1.95	2.01	1.70
N	1.74	1.74	1.74	1.68	1.69	1.55
O	1.81	1.79	1.78	1.75	1.75	1.52
S	2.60	2.27	2.29	2.33	2.41	1.80
F	2.09	2.09	2.08	2.05	2.49	1.47
Cl	2.38	2.36	2.47	2.41	2.46	1.75
Br	2.18	2.23	2.46	2.45	2.63	1.85
γ[i]					6.8	
AUE[j]	1.06	0.99	1.04	1.08	1.13	
Stdev[k]	1.00	0.96	0.99	1.00	0.90	
rms[l]	1.45	1.38	1.44	1.47	1.45	
R[m]	0.89	0.90	0.90	0.89	0.88	
RRMS[n]	0.34	0.33	0.34	0.35	0.34	
AE[o]	−0.18	−0.17	−0.17	−0.17	0.02	

[a] Tag names for each of the optimized solvent cavity radii. [b] Atomic partial charges from an ESP-fit or a DRESP fit on the given quantum method. [c] Prepolarized charges from HF/6-31G(d,p). [d] Vacuum charges from B3LYP/6-311++G-(3df,3pd). [e] A and ε_{in} of eq 3 for the solute internal dielectric. The atomic radii used in the internal dielectric are given in Table 1. [f] ΔG_{np} from free energy perturbation.[45] [g] ΔG_{np} calculated using the surface area (eq 20) with the γ term optimized. [h] Cavity atomic radii are given in angstrom. [i] Nonpolar surface tension from eq 20 in cal/Å^2. [j] Average unsigned error in kcal/mol. [k] Standard deviation of the unsigned error. [l] Root-mean-square deviation in kcal/mol. [m] Pearson correlation coefficient. [n] Relative root-mean-square deviation. [o] Average signed error in kcal/mol: experiment - calculated.

on the method used to approximate ΔG_{np}. The surface area (SA) based method follows eq 20 and required the optimization of the surface tension parameter (γ). The main effort here is however concentrated with the use of ΔG_{np} by converged free energy perturbation (FEP) calculations.[45,46] This is the first category of results that we examine below.

4.3.1. Results with FEP-Based Nonpolar Term. Two main classes of solute are studied as reported in Table 3. First, we set $\varepsilon_{in} = \varepsilon_{trans} = 1$ in eq 3, effectively turning eq 3 into a conventional 2-zone dielectric function with a nonpolarizable solute as defined previously by Grant et al.[18] For the nonpolarizable solute model, we used static atomic charges as given by ESP-fitting to the conventional overpolarized HF-6-31G(d,p) wave function (G2-HF). These charge sets are positive controls, following the traditional approaches for nonpolarizable force fields and which have been shown to produce the right degree of static polarization of the solute in water.[70] The G2-ε_{in}, with $\varepsilon_{in} = 4$, 12, and 24 (cf. Table 1), has polarizable solutes assigned charges fitted to the B3LYP/6-311++G(3df,3pd) ESP known to reproduce the gas-phase dipole moment of the molecules, being usually between 10% and 20% smaller than what is normally expected in water.

Table 4. Effects of Using a Different Solvent Cavity Radii Set (Table *3*) with the G1-12 Solute Model (Table *1*) on ΔG_{hyd}

	G2−12	G2−24	G2−4	G2-HF
AUE[a]	1.04	1.08	1.17	1.10
Stdev[b]	0.99	1.03	1.19	1.01
R[c]	0.90	0.90	0.86	0.89
RRMS[d]	0.34	0.35	0.39	0.35
AE[e]	−0.17	0.16	−0.23	0.15

[a] Average unsigned error in kcal/mol. [b] Standard deviation on the AUE in kcal/mol. [c] Pearson correlation coefficient. [d] Relative root-mean-square deviation. [e] Average signed error in kcal/mol: experiment - calculated.

We look for the polarizability to compensate for the use of gas-phase charges.

It is quite interesting to observe in Table 3 that by allowing the cavity radii to optimize in each model, the same level of error over the 485 experimental free energies of hydration is obtained for the G2-HF, G2−4, G2−12, and G2−24 solute models. The average unsigned error (AUE) compared to experiment is 1 kcal/mol with a standard deviation of 1 kcal/mol. The Pearson correlation coefficient (R) is around 0.89 in all these G2 models. The relative root-mean-square deviation (RRMS) obtained is 0.35, and the average signed error (AE) is found to be between −0.15 kcal/mol and −0.18 kcal/mol. These errors can be compared to the Rizzo et al.[47] results, on almost the same data set (460 neutral molecules included in the 485 that we use), that produce an AUE of 1.47 kcal/mol with RESP charges and R = 0.88. The reported numbers of Rizzo et al. were obtained with a SA evaluation of ΔG_{np} that allow them to subsequently optimize 14 atom typed surface tensions (γ), which improved the AUE to 1 kcal/mol while R = 0.89. For comparison, in the current study, we fit 8 atomic radii. In addition, the recent work of Mobley et al.[46] using Bondi radii and the single γ fitted by Rizzo et al. on an almost identical data set to ours obtained a root-mean-square deviation of 2.05 kcal/mol. Finally, in a different article, Mobley et al.[45] obtained a rms of 1.26 kcal/mol and R = 0.89 with explicit-solvent converged FEP calculations. The FEP based ΔG_{np} used in this work comes from this latter study. Our results are comparable or better to most other studies. We attribute the small errors to the optimization of the radii, not necessarily to the quality of the solute model. We can however examine the fitted cavity radii with the different solute models to understand the effects of the electrostatic model on the solute cavity size.

The level of solute polarization brought by the polarizable solute models (G2−4 to G2−24) seems similar to what is obtained with the G2-HF solute model. This can be assessed by comparing the atomic radii and the cross-validation error showed in Table 4 where the G1−12 solute model is used with the different G2 radii sets. The level of error produced when $\varepsilon_{in} = 4$, 12, and 24 or with the G2-HF cavity radii is similar, the G2−4 being the worst. The cross-validation results of Table 4 also show the transferability of the third zone dielectric parameters given that the solute has the physically appropriate electrostatic behavior. A possible advantage of the polarizable solute model is when the solvation free energies are computed relative to a solvent much less polar than water (e.g., a nonpolar solvent or a

nonpolar binding site in a protein). In this case, the HF based charges may not be appropriate.[71]

The fitted radii of Table 3 are significantly different from the contact Bondi radii reported in the last column. First, the H radius is a little smaller than the usual 1.1 Å contact radius in all cases (the Bondi radius of 1.2 Å for H was recognized to be a little too large and was revised to be 1.1 Å[72]). The carbon radius obtained here is much larger than the Bondi radius and makes the C−H bonds behave like a united atom model. In this perspective the carbon radius size obtained here is similar to the Nina et al.[73] carbon radius they calculated by looking at MD water charge density in explicit solvent simulations. For the other elements, we also find larger radii than Bondi, in agreement with a recent study by Nicholls et al.[31]

The larger cavity radii can be rationalized by considering the difference between contact radii (Bondi) and the cavity radii needed in implicit solvent calculations. The former is defined by crystal contacts between neighboring molecules, and the latter is defined by where the mean solvent charge density begins. In terms of the 3-zone dielectric model, the contact radii would be located in the middle of the second zone where the electronic density should be minimal given that the dielectric goes to one (no electrons to be polarized on the solute side). This is supported by the fact that Fermi repulsion between the solvent molecule electrons and the solute electrons reduces the total electronic density to its minimum exactly in the contact zone. In Figure 2a, the contact radius of an aromatic carbon atom would become 1.7 Å, exactly the Bondi radius value. Similarly the middle of the blue area in Figure 2b defines the contact line between solvent molecules and the solute.

Although we claim here that having cavity radii larger than Bondi radii may be physically motivated, it is not possible at this stage to know if this effect should be as large as we find. In particular, the fluorine radii in Table 3 are surprisingly large. This was also found by Nicholls et al.[31] where their optimal fluorine radius was 2.4 Å. Knowing that fluorine is particularly hydrophobic, this may be just another peculiar behavior of this atom. The Cl and Br radii difference in the G2−4, G2−12, and G2−24 sets uncover a drawback of using a small ε_{in}. Because the polarization radius of Cl and Br are larger in the G1−4 than in the other EPIC parametrizations, the transition zone shown in Figure 2 cannot reach $\varepsilon(r) = 1$ (in the case of Br, it only decreases to $\varepsilon(r) = 3$), and as a result the full polarizability coming from the halogen atom is not reached as the solvent cuts into the first zone dielectric function. This prevents enough solute bound charge density from building up.

A last point to mention in regard to the 3-zone dielectric function is its potential advantage in reducing the occurrence of the reentrant surface problem that often brings a lot of fluctuation in energy or force computations in proteins, for instance. The problem is the artificial formation of a cavity with high dielectric inside a protein due to the irregularity of molecular surfaces. The large size of the atomic cavity radii in the G2 sets and the use of a smooth dielectric function should contribute to create a sufficiently deep buffer of low dielectric and make implicit solvent models more stable. Indeed, the smoothness of the surface around 4-pyridone observed in Figure 2b looks like a solvent accessible surface.[17] This entire question is however left for future research.

4.3.2. Results with the Surface Area-Based Nonpolar Term. Although the use of the very computationally intense FEP-based ΔG_{np} may be better physically grounded, the obtained models cannot practically be used in a prospective manner due to the heavy computational demands for such FEP calculations. For this reason, we also optimized the cavity radii and the surface tension with the G1−12 solute models. In these calculations the solvent accessible surface area was calculated with the Bondi radii and kept constant. The results are reported in Table 3. The error levels reported are comparable to those obtained with ΔG_{np} from FEP calculation. The G2−12/SA model gives error levels a little larger than the G2−12: AUE = 1.13 kcal/mol with a standard deviation of 0.90 kcal/mol, R = 0.88, RRMS = 0.34, rms = 1.45 kcal/mol, and AE = 0.02 kcal/mol. The radii obtained for the G2−12/SA fit are similar to the G2−12 fit except for S, F, Cl, and Br. It is possible that the hydrophobicity of these atoms is overestimated by the single surface tension term used with the Bondi radii to determined ΔG_{np}.

5. Conclusions

The EPIC approach to molecular polarizability has been parametrized to include many more chemical functional groups than the previous effort.[1] This required generating a data set of 707 B3LYP/aug-cc-pVTZ molecular polarizability tensors. The ability of EPIC to account for both the average polarizability and the anisotropy of the tensor was remarkable given that the optimization of only 14 parameters (excluding water and charged species) led to a relative unsigned error in the average polarizability and anistropy of 2.6% and 5.2%, respectively (G1−12). An example of the parsimony of the atom typing is that a single radius parameter was sufficient for aromatic, nitrile, amine, aniline, or hydrazine types of nitrogen. Obtaining the same level of error with both the validation and training data sets suggests that overfitting is not an issue. With previous polarizable models, such as point-inducible dipoles, this level of accuracy could only be attained with added complexity such as anisotropic polarizable centers or molecule-specific Thole screening parameters.[5,15,28,65−67]

We found that the anisotropy could only be reproduced accurately if the interior dielectric constant was higher than 9.0. Above this value, almost any interior dielectric can also work well as long as the atomic polarization radii are appropriately adjusted. The need for a high interior dielectric raised the question of the physical soundness of the model as $\varepsilon_{in} = \varepsilon_{\infty} = n^2$ has become a dogma in the implicit solvent literature[13,30,31,39,40,74−76] whenever the dielectric constant is used to replace electronic response. The conceptual flaw of this equality comes from the fact that the interior dielectric (ε_{in}) is not uniformly distributed, whereas the refractive index (n) comes from a macroscopic measurement assuming a uniform ε_{∞} in space. To verify that the optimized models agree with experimental refractive indices, we devised a new protocol to calculate a liquid refractive index from micro-

scopic simulations. To this end, 23 organic molecules spanning the entire range of bioorganic-molecule-like refractive indices state were simulated by molecular dynamics in the liquid state and ε_∞ was calculated with the Clausius-Mossoti relation. The obtained refractive indices now come from the effective polarization of liquid configurations having intramolecular high dielectric with low-dielectric interstices. The results show a good correlation for all three G1-ε_{in} parameter sets. The highest interior dielectric ($\varepsilon_{in} = 24$) gave the best correlation with a slope of 1.00, an intercept of 0.05, and a correlation coefficient of 0.95. It is interesting to note that the polarizability anisotropy may play a role since the G1−4 parameters ($\varepsilon_{in} = 4$) gave the poorest correlation and was also the worst model for polarizability anisotropy. These results indicate that, when coupled with the appropriate radii, many choices of high ε_{in} can give results in good agreement with the experimental refractive indices.

To use the EPIC polarizable electrostatic model with implicit solvent, we have developed a smoothed-boundary 3-zone dielectric function that works with the internal dielectric continuum model. The three zones are the internal dielectric constant ε_{in}, a transition zone that tends to the vacuum dielectric ($\varepsilon_{trans} = 1$), and a third zone defined by the molecular cavity boundary, where the dielectric function reaches the bulk liquid dielectric. With this function, keeping the first zone fixed at the atomic polarization radii and ε_{in} determined for the gas phase polarizabilities, only the molecular cavity boundary needs to be parametrized. A data set of 485 experimental free energy of hydration was used to optimize the solvent cavity radii, one per element, with different charge models. The resulting level of error was smaller than found in previous implicit solvent studies with a typical average unsigned error of 1 kcal/mol, a standard deviation of about 1 kcal/mol, and a Pearson correlation coefficient of 0.9. Atomic charge sets fitted from the unpolarized gas phase B3LYP QM ESP coupled with EPIC polarization led to cavity radii comparable to those obtained with the polar-condensed-phase-like HF/6-31G(d,p) charges. The low sensitivity of the optimal cavity radii resulting from the fit with different polarizable solutes (the different G1-ε_{in}) further supports the generality of the approach. These results clearly show that EPIC can lead to accurate description of solute polarization in implicit solvent. The anisotropy of the molecular polarizability does not seem to play an important role in fitting experimental hydration free energies. However, when considering intermolecular interactions, such as in an enzyme active site, the heterogeneity of the environment and of the interactions may require an accurate directional polarizability. An important example of this is in cation-π interactions[2,77]

The proposed global optimization scheme involves several independent layers. The polarizability part is fitted on uncharged QM molecular polarizability tensors. The charges are added with the DRESP fit on *ab initio* electrostatic potentials calculated on a grid, as usual, except that here the QM method can be systematically improved since gas phase properties are needed. For implicit solvation, solvent-related radii are obtained from a fit to experimental hydration free energies. Flexibility and transferability have been demonstrated for each stage. The ease with which we could fit the polarizability of so many functional groups leads us to believe that the further extension of the parametrization and atom typing should be straightforward given more data. Moreover, the decoupling of the fitted polarization from the fitted charges as well as the physical soundness of each step makes the above parametrization scheme even more robust and general than is possible for two-body additive force fields.

This work partly addresses the question of applicability of EPIC in polar condensed phase. The calculations of the refractive indices were well behaved and show the high values for ε_{in} are not unphysical. It is not clear if the slightly larger calculated polarizabilities of the droplets were due to a change in polarizability when going from gas phase to condensed phase. Also, the level of electronic induction seen in the 3-zone implicit solvent calculations suggests that both solute and solvent polarization in polar media is well modeled. To confirm those findings, it would be interesting to perform explicit atoms simulations with the EPIC model and Poisson's equation.

The EPIC approach to polarizability has shown unprecedented accuracy and flexibility on many accounts for such a simple model. Although the optimized parameters are unconventional compared to traditional Poisson−Boltzmann applications, it is for sound physical reasons that even clarify aspects of the implicit solvent approaches. In this paper and the two previous ones,[1,2] EPIC was shown to be a powerful tool to include the effects of electronic polarization in molecular mechanics type calculations, especially appropriate to biomolecular force fields.

Acknowledgment. This work was made possible by the computational resources of the réseau québécois de calcul haute performance (RQCHP). The authors are grateful to OpenEye Inc. for free academic licenses. R.I.I. acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC). J.-F.T. is supported by NSERC through a Canada graduate scholarship (CGS D) and by Merck & Co. through the MRL Doctoral Program I. B.R. is supported by NIH grant GM072558.

Supporting Information Available: DFT and EPIC polarizabilities for all 707 molecules examined together with the optimized coordinates of the molecules used to calculate the polarizability tensor, table of the G1−24, G1−12, and G1−4 fitted charges for several example molecules, G1−12 charges together with B3LYP/6-31++G(d,p)-optimized geometries for the 485 molecules examined for hydration free energies, calculated and experimental refractive indices, and detailed free energies of hydration. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Truchon, J.-F.; Nicholls, A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. *J. Chem. Theory Comput.* **2008**, *4*, 1480–1493.

(2) Truchon, J.-F.; Nicholls, A.; Grant, J. A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. *J. Comput. Chem.* in press.

(3) Applequist, J.; Carl, J. R.; Fung, K. K. *J. Am. Chem. Soc.* **1972**, *94*, 2952–2960.

(4) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(5) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8543–8551.

(6) Lamoureux, G.; Allouche, D.; Souaille, M.; Roux, B. *Biophys. J.* **2000**, *78*, 330A–330A.

(7) Lamoureux, G.; Roux, B. *Biophys. J.* **2001**, *80*, 328A–328A.

(8) Schropp, B.; Tavan, P. *J. Phys. Chem. B* **2008**, *112*, 6233–6240.

(9) Lamoureux, G.; MacKerell, A. D.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(10) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

(11) Morita, A. *J. Comput. Chem.* **2002**, *23*, 1466–1471.

(12) Sharp, K.; Jean-Charles, A.; Honig, B. *J. Phys. Chem.* **1992**, *96*, 3822–3828.

(13) Tan, Y. H.; Luo, R. *J. Chem. Phys.* **2007**, *126*, 094103.

(14) Tan, Y. H.; Tan, C. H.; Wang, J.; Luo, R. *J. Phys. Chem. B* **2008**, *112*, 7675–7688.

(15) Birge, R. R. *J. Chem. Phys.* **1980**, *72*, 5312–5319.

(16) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.

(17) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.

(18) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608–640.

(19) Prabhu, N. V.; Zhu, P. J.; Sharp, K. A. *J. Comput. Chem.* **2004**, *25*, 2049–2064.

(20) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59–75.

(21) Griffiths, D. J. *Introduction to Electrodynamics*; 3rd ed.; Prentice-Hall Inc.: Upper Saddle River, NJ, 1999.

(22) *Zap Toolkit, pre-release version*; Santa Fe, NM, USA, 2007.

(23) Weininger, D. *J. Chem. Inf. Model.* **1990**, *30*, 237–243.

(24) Weininger, D.; Weininger, A.; Weininger, J. L. *J. Chem. Inf. Model.* **1989**, *29*, 97–101.

(25) Weininger, D. *J. Chem. Inf. Model.* **1988**, *28*, 31–36.

(26) *OEChem Toolkit, version 2.2.1*; Santa Fe, NM, USA, 2007.

(27) SciPy: Open Source Scientific Tools for Python, 2001.

(28) Elking, D.; Darden, T.; Woods, R. J. *J. Comput. Chem.* **2007**, *28*, 1261–1274.

(29) Bottcher, C. J. F. *Theory of Electric Polarization*; 2nd ed.; Elsevier Scientific Publishing Company: New York, 1973; Vol. 1.

(30) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(31) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769–779.

(32) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(33) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(34) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(35) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(36) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(37) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(38) Marsaglia, G. *Annu. Math. Stat.* **1972**, *43*, 645–646.

(39) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671–679.

(40) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(41) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(42) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1–20.

(43) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J. Phys. Chem. B* **2008**, *112*, 938–946.

(44) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(45) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*.

(46) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

(47) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(48) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–499.

(49) Pitera, J. W.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 3163–3164.

(50) Choudhury, N.; Pettitt, B. M. *Mol. Simul.* **2005**, *31*, 457–463.

(51) Wagoner, J. A.; Baker, N. A. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8331–8336.

(52) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435–445.

(53) Varga, R. S. *Matrix Iterative Analysis*; 2nd ed.; Springer: New York, 2000.

(54) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(55) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Menucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.;

Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, revision C.02*; Gaussian Inc.: Wallingford, CT, USA, 2004.

(57) Kendall, R. A.; Dunning, J.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(58) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(59) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294–301.

(60) Woon, D. E.; Dunning, J. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(61) Hammond, J. R.; Kowalski, K.; deJong, W. A. *J. Chem. Phys.* **2007**, *127*, 144105.

(62) Dykstra, C. E.; Jasien, P. G. *Chem. Phys. Lett.* **1984**, *109*, 388–393.

(63) *Handbook of Chemistry and Physics*, 83 ed.; CRC Press: New York, 2002.

(64) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: Ithaca: New York, 1960.

(65) Shanker, B.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 3879–3881.

(66) Bode, K. A.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 17820–17824.

(67) Harder, E.; Anisimov, V. M.; Whitrield, T.; MacKerell, A. D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3509–3521.

(68) Sekino, H.; Maeda, Y.; Kamiya, M.; Hirao, K. *J. Chem. Phys.* **2007**, 126, 014107.

(69) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621–627.

(70) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.

(71) Eksterowicz, J. E.; Miller, J. L.; Kollman, P. A. *J. Phys. Chem. B* **1997**, *101*, 10971–10975.

(72) Rowland, R. S.; Taylor, R. *J. Phys. Chem.* **1996**, *100*, 7384–7391.

(73) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.

(74) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465–1471.

(75) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *BBA -Proteins Proteom.* **2006**, *1764*, 1647–1676.

(76) Green, D. F.; Tidor, B. *J. Phys. Chem. B* **2003**, *107*, 10261–10273.

(77) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6290–6295.

CT900029D