Dielectric properties of proteins play important roles in their structural and functional characteristics.\(^1\)\(^-\)\(^3\) A simple approach is to approximate the protein interior region by a dielectric continuum with a single dielectric constant, which is embedded in a cavity surrounded by another dielectric continuum representing the solvent. The inhomogeneous nature of the protein interior, important for example in the dielectric response on the active site or embedded chromophore,\(^4\) is neglected in this approach. Although atomic point charges have been introduced inside the cavity to refine the model,\(^3\) an appropriate value of the dielectric constant for a protein is not well defined in general.

Molecular dynamics (MD) simulations allow for a detailed prediction of the protein dielectric properties in solution.\(^5\)\(^,\)\(^6\) Despite considerable progress made recently, the computational cost involved in simulating large proteins in solution is still demanding.

In this report, a physically well-defined middle ground between the dielectric continuum approach and detailed atomistic models such as MD methods is presented. A protein molecule in solution is represented by a set of polarizable dipoles embedded in a dielectric medium of solvent molecules. The interface separating the cavity containing the dipoles from the solvent region can be approximated as a solvent-accessible molecular surface generated by a suitable algorithm from the native structure of the protein. Dielectric properties of the protein are approximated by replacing each residue of the protein by a polarizable dipole. The positions of the dipoles can be consistently assigned based on the native structure of the protein. A set of polarizability values appropriate for each of the residues would completely determine model parameters, allowing for calculations of the protein dielectric response function.

An intrinsic set of polarizabilities is introduced and obtained for each of the 20 naturally occurring amino acids. This set is furthermore assumed to be universal, such that once a set of reasonable values has been determined, they can be used for predicting dielectric properties of any large proteins. The intrinsic polarizability of a dipole would be modified by the interactions with other dipoles and the effect of solvent polarizations, depending on the particular environments produced by the protein of interest. We note that Vöges and Karshikoff\(^7\) have introduced a similar model for the electronic polarizabilities of residues employing the idea of local dielectric constants for proteins. A distinguishing feature of our model is the way by which the nuclear intrinsic polarizabilities are determined. Therefore a combination of our model and the model by Vöges and Karshikoff may offer a realistic representation for the dielectric inhomogeneity of proteins.

Our strategy of obtaining a reasonable set of intrinsic polarizabilities is to perform MD simulations of small proteins, and to compare the dipole moment fluctuations generated from the dynamical trajectories with a predicted expression of the model. A critical test of the concept of intrinsic polarizabilities would be the consistency of values determined for one residue with those for different residues of the same amino acid type, both within a single protein and in different proteins. By simulating three small proteins, and analyzing the data in terms of the model, we have found a reasonable set of intrinsic polarizabilities of amino acids, which is consistent within a protein and between different proteins.

Before providing a detailed presentation of our model and calculations we would like to put our work in a proper context. The basic idea of using an atomistic force field to simulate properties of proteins has been well established by extensive studies.\(^5\)\(^,\)\(^6\) In this case a consistent set of atomic parameters are obtained from quantum chemistry calculations or by fitting experimental data, and this set of parameters is assumed to be universal. Indeed many studies have shown the universality of these parameters.\(^8\) Due to the long-range nature of the electrostatic interaction and the well-defined protein native structure it is reasonable to assume that the dielectric properties of a protein can be described by a collection of normal modes originated from the collective atomic motions around its native structure. Since a protein consists of 20 structurally universal units it seems natural to assume that the polarizabilities of the amino acids in proteins are universal. For the electronic part of the polarizability this
The Hamiltonian of a collection of polarizable dipoles inside a dielectric cavity can be written as

$$ H = H_p + H_s + H_{ps} . $$

In Eq. (1), $H_p$ is the Hamiltonian for the interacting polarizable dipoles inside the cavity,

$$ H_p = \frac{1}{2} \sum_i (\mathbf{m}_i - \bar{\mathbf{m}}_i)^2 / \alpha_i + \frac{1}{2} \sum_{i \neq j} \mathbf{m}_i \cdot \mathbf{T}(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{m}_j , $$

where $\mathbf{m}_i$ is the dipole moment vector of the $i$th residue centered at the position $\mathbf{r}_i$ with polarizability $\alpha_i$, and $\mathbf{T}(\mathbf{r}_i - \mathbf{r}_j) = \nabla_i \nabla_j (1/|\mathbf{r}_i - \mathbf{r}_j|)$ is the dipole–dipole interaction tensor, where $\nabla_i$ and $\nabla_j$ are the gradients with respect to $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively. In Eq. (2), the polarizability $\alpha_i$ of a particular residue is defined in terms of dipole moment fluctuations with respect to $\bar{\mathbf{m}}_i$, which is also assumed to be an intrinsic property of the corresponding amino acid. The specific values of these intrinsic dipole moments, however, will not affect predictions of the protein dielectric response, as will be shown in the following.

As has been shown previously, the solvent Hamiltonian $H_s$ in Eq. (1) can be written in a quadratic form in terms of electric field of the solvent $\mathbf{E}(\mathbf{r})$:

$$ H_s = \frac{1}{2} \int \int \left[ \int_{\text{out}} d\mathbf{r} d\mathbf{r}' E(\mathbf{r}) \cdot \chi_m^{-1}(\mathbf{r}, \mathbf{r}') \cdot E(\mathbf{r}') \right] . $$

where “out” denotes that the integration is limited to the region outside the cavity and the modified solvent response function $\chi_m(\mathbf{r}, \mathbf{r}')$ can be obtained from the electric field fluctuations in the presence of the cavity.

The interaction Hamiltonian $H_{ps}$ in Eq. (1) is defined as

$$ H_{ps} = \sum_i \mathbf{m}_i \cdot \mathbf{E}(\mathbf{r}_i) , $$

where $q(\mathbf{r})$ is the charge density of the solvent and $\mathbf{E}(\mathbf{r}_i)$ is the electrostatic field at position $\mathbf{r}_i$ due to solvent molecules. The summation over $i$ runs over all of the residues in the protein. Gaussian electric field of the solvent can be integrated out to yield an effective Hamiltonian of the protein,

$$ H_{\text{eff}} = H_p + \frac{1}{2} \sum_{ij} \mathbf{m}_i \cdot \mathbf{Q}(\mathbf{r}_i, \mathbf{r}_j) \cdot \mathbf{m}_j . $$

The solvent-mediated interaction tensor between the dipoles $\mathbf{Q}(\mathbf{r}_i, \mathbf{r}_j)$ in Eq. (5) is given by

$$ \mathbf{Q}(\mathbf{r}_i, \mathbf{r}_j) = -\beta \left( \int \int_{\text{out}} d\mathbf{r} d\mathbf{r}' q(\mathbf{r}) \nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|} \nabla' \frac{1}{|\mathbf{r}' - \mathbf{r}_j|} q(\mathbf{r}') \right) , $$

where $\langle \rangle$ denotes the ensemble average in the present of the solute, where $1/\beta = k_B T$ is the Boltzmann constant times temperature. Equation (6) determines the solvent-mediated interaction between dipoles due to the reaction field from solvent polarization. From MD trajectories the ensemble average can be done and the reaction field tensor can be obtained directly from MD simulations. It should be noted that since the MD trajectories are generated in the presence of the solute and the ensemble average performed will be the ensemble average of the solvent molecules in the presence of the solute.

Combining Eqs. (2) and (5), we obtain an implicit relation determining the mean dipole moment $\langle \mathbf{m}_i \rangle$ of the $i$th residue in terms of intrinsic dipole moment of the corresponding amino acid,

$$ \bar{\mathbf{m}}_i = \langle \mathbf{m}_i \rangle + \alpha_i \sum_j \left[ (1 - \delta_{ij}) \mathbf{T}(\mathbf{r}_i, \mathbf{r}_j) + \mathbf{Q}(\mathbf{r}_i, \mathbf{r}_j) \right] \langle \mathbf{m}_j \rangle , $$

where $\delta_{ij}$ is the Kronecker delta for the residue indices $i$ and $j$. Mean dipole moment $\langle \mathbf{m}_i \rangle$ represents the dipole moment “modified” from $\bar{\mathbf{m}}_i$ due to the interaction with other dipoles and surrounding solvents. It would correspond to the equilibrium dipole moment of a residue of the protein in solution. Dipole moment correlation tensor $\mathbf{M}(\mathbf{r}_i, \mathbf{r}_j) = \beta (\delta \mathbf{m}_i, \delta \mathbf{m}_j)$, associated with the fluctuations $\delta \mathbf{m}_i = \mathbf{m}_i - \langle \mathbf{m}_i \rangle$, is given by due to the quadratic nature of Hamiltonian equation (5),

$$ \mathbf{M}^{-1}(\mathbf{r}_i, \mathbf{r}_j) = \delta_{ij} / \alpha_i + (1 - \delta_{ij}) \mathbf{T}(\mathbf{r}_i, \mathbf{r}_j) + \mathbf{Q}(\mathbf{r}_i, \mathbf{r}_j) . $$

Equation (8) shows that the correlation of the dipole moment fluctuations, and therefore the dielectric response of the protein is independent of the mean dipole moment $\langle \mathbf{m}_i \rangle$ or the intrinsic dipole moment $\bar{\mathbf{m}}_i$.

With the solvent-mediated interaction tensor $\mathbf{Q}(\mathbf{r}_i, \mathbf{r}_j)$ obtained from Eq. (6), Eq. (8) completely determines the dielectric response of a protein within the framework of the model, given the coordinates of the dipoles $\{ \mathbf{r}_i \}$ based on the equilibrium solution structure of the protein, and the set of values of the intrinsic polarizabilities $\{ \alpha_i \}$. The crucial point of the model lies in the determination of the intrinsic amino acid polarizabilities. If available, experimental data corresponding to the left-hand side of Eq. (8) would provide use-
ful guidance to these intrinsic polarizabilities. Alternatively, MD simulation results can be used instead, provided that the right-hand side of Eq. (8) is also calculated within the model with conditions corresponding to the simulation. We have adopted the latter strategy, with adjustments of the boundary element calculations to fit MD results with Ewald summation.

MD simulations of three small proteins, bovine pancreatic trypsin inhibitor (BPTI), ribonuclease inhibitor, and lysozyme with water molecules have been performed to obtain the dipole moment correlation tensor \( \mathbf{M}(\mathbf{r}, \mathbf{r}') \). Initial coordinates of the proteins were taken as crystal structures from the Protein Data Bank.\(^{13}\) Periodic boundary conditions have been used with cubic boxes of different lengths filled with TIP3P\(^{14}\) water molecules of bulk density 1.00 g/cm\(^3\). A CHARMM9 force field was used,\(^{15}\) with the integration time step 2 fs. Covalent bonds between hydrogen and heavy atoms were fixed by SHAKE algorithm.\(^{16}\) The particle mesh Ewald method\(^{17}\) was used for electrostatic interactions. For each system, data were collected over 1.0 ns after initial 0.3 ns equilibration. Temperature was maintained near 300 K by periodic rescaling of particle velocities every 1000 steps. Since dipole moments of charged residues are not well defined the partial charges of the charged residues are assigned based upon Lazaridis and Karplus scheme\(^{18}\) to render the residue neutral. Table I shows the box sizes and the number of water molecules in a unit cell used for the simulations.

To make comparisons with the model calculation of dipole moment correlation tensor based on the geometry of a single cavity inside an infinitely extended dielectric medium, MD simulations of a protein inside a continuous dielectric material of infinite size would have to be performed. For MD simulations with periodic boundary conditions an Ewald summation has to be performed to calculate the reaction field based on the same configuration. The polarization reaction field inside the cavity in this case can be written as

\[
Q(\mathbf{r}, \mathbf{r}') = -\beta \sum_n \left( \int \int_{\text{out}} d\mathbf{r} d\mathbf{r}' q(\mathbf{r}) \times \nabla \left[ \frac{1}{|\mathbf{r} - \mathbf{r} + \mathbf{n}|} \nabla' \left[ \frac{1}{|\mathbf{r}' - \mathbf{r}' + \mathbf{n}|} q(\mathbf{r}') \right] \right] \right),
\]

where \( \mathbf{n} \) spans the integer multiples of the lattice vector corresponding to the periodic box.\(^{19}\) The formula of the Ewald potential with a spherical shell summation scheme on conducting boundary condition was used.\(^{20}\)

With the reaction field tensor \( Q(\mathbf{r}, \mathbf{r}') \) and the dipole moment correlation tensor \( \mathbf{M}(\mathbf{r}, \mathbf{r}') \), Eq. (8) can be represented as a 3N×3N matrix equation with the set of 20 unknowns \( \{\alpha_i\} \), where \( N \) is the number of residues in the protein. As can be seen from Eq. (8), only diagonal elements of the inverse of dipole moment correlation tensor is affected by the values of polarizabilities within the framework of the quadratic Hamiltonian defined by Eq. (2). It is readily verified that to minimize the mean square difference between both sides of the diagonal part of Eq. (8), one should take the average of \( 1/\alpha_i \) values for each residue of a particular type of amino acid. Amino acid compositions of the three proteins simulated are shown in Table II.

For each of the 20 amino acids, the averages and standard deviations of \( 1/\alpha_i \) were determined from the diagonal part of Eq. (8) with the solvent-mediated contribution calculated by Eq. (9), within the cumulative database of the three proteins shown in the last row of Table II.

Inverse intrinsic polarizabilities of amino acids determined are plotted in Fig. 1, and the corresponding average polarizabilities are shown in Table III. Amino acids with relatively long, polar side chains have larger polarizabilities than those with short, nonpolar side chains. Polarizabilities range from \( \alpha_i \sim 1 \text{Å}^3 \) to \( \alpha_i \sim 40 \text{Å}^3 \), in contrast to the implicit assumption of the dielectric continuum theory, \( \alpha_i = \text{const} \). Also shown in Table III are the electronic polarizabilities given by the model of Voges and Karshikoff.\(^{7}\) It is clear that the nuclear and electronic polarizabilities have different origin. For example, the nonpolar side chain of tryptophan gives a small nuclear polarizability (1.8 Å\(^3\)), on the other hand the electronic polarizability (21 Å\(^3\)) is rather large due to the electron rich aromatic ring. In the Voges and Karshikoff model the nuclear polarizability is determined from the Onsager model based on the permanent dipole moment of the amino acid. Their values are substantially larger (Table III) than ours given here. This is not unexpected since the dipole fluctuations of residues in a protein are severely constrained in contrast to the unconstrained assumption in

<table>
<thead>
<tr>
<th>Protein</th>
<th>Box size (Å)</th>
<th>No. of water</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPTI</td>
<td>46.7</td>
<td>2955</td>
</tr>
<tr>
<td>RNI</td>
<td>51.2</td>
<td>3871</td>
</tr>
<tr>
<td>LYZ</td>
<td>56.6</td>
<td>5014</td>
</tr>
</tbody>
</table>

| Protein | N | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
| BPTI | 58 | 6 | 6 | 3 | 2 | 6 | 1 | 2 | 6 | 6 | 1 | 2 | 2 | 2 | 4 | 1 | 4 | 4 | 1 | 3 | 3 | 4 | 5 |
| RNI | 89 | 6 | 3 | 3 | 4 | 2 | 6 | 11 | 5 | 1 | 6 | 12 | 6 | 6 | 2 | 2 | 5 | 4 | 3 | 3 | 3 | 5 |
| LYZ | 129 | 13 | 10 | 14 | 6 | 8 | 1 | 2 | 13 | 2 | 6 | 9 | 7 | 2 | 2 | 2 | 10 | 7 | 6 | 4 | 5 |
| Total | 276 | 25 | 19 | 20 | 12 | 16 | 8 | 15 | 24 | 3 | 14 | 23 | 17 | 3 | 8 | 8 | 16 | 14 | 9 | 11 | 11 |
Onsager’s model. If the molecular volume $v$ of each amino acid is known we can define a dimensionless polarizability $\gamma = 4 \pi \alpha/3v$ and Clausius–Mossotti relation or other models yield an effective dielectric constant, which can be viewed as a local dielectric constant used in the Vögels and Karshikoff model. In our model, once these intrinsic polarizabilities are known the dielectric properties of a protein can be calculated and there is no need to invoke the local dielectric constant concept.

Fluctuations of polarizability values for each amino acid arise from the effect of the particular local environment not accounted for within the framework of the model we have adopted. Although it would be of interest to see how the statistics change as we sample more proteins, we expect that the major contribution to the uncertainties observed in Fig. 1 is due to the effect of heterogeneous local environments, which can be captured only partially with the point-dipole picture of protein residues. However, considering the exponential distribution of polarizabilities within 20 amino acids as seen in Fig. 1, and that even a simple dielectric continuum picture already provides a qualitative guide to the problem, it is reasonable to assume that the dielectric properties of a protein on a length scale comparable to or larger than the size of each residue would be relatively insensitive to the uncertainties of the intrinsic polarizabilities.

We also note that the relatively uniform distribution of the polarizabilities of amino acids in logarithmic scale shown in Fig. 1 suggests the possibility that the natural amino acid set has been evolutionarily selected partially to maximize nature’s flexibility in protein design in terms of dielectric properties.

As a first step toward an inhomogeneous model on the dielectric response of proteins, it seems the 20 intrinsic polarizabilities of the amino acids did capture the salient features of protein’s inhomogeneity. There are many possible ways by which the model can be extended. The model can easily be generalized to the dynamical case. Evaluation of the frequency-dependent intrinsic polarizabilities would allow for predictions of the time-dependent dielectric response of systems such as chromophores embedded in large proteins. In this report all of the residues are neutral and it will be interesting to see how these intrinsic polarizabilities will hold if charge residues are used in the calculation.

The author is grateful for the contributions and discussions of Hyung-June Woo at the initial stage of this work and for the financial support by a Petroleum Research Fund, administered by American Chemical Society.

1 M. F. Perutz, Science 201, 1187 (1978).
Analogous summation of the dipole–dipole interaction $T(r,r')$ would have to be done in Eq. (8). This modification is not necessary, since determination of $a_i$ involves the diagonal parts of Eq. (8) only, and the reaction field on a dipole due to its own replicas vanishes.
