

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 6286–6295

www.rsc.org/pccp

PAPER

A theory for the anisotropic and inhomogeneous dielectric properties of proteins

Will C. Guest,^{ab} Neil R. Cashman^b and Steven S. Plotkin^{*a}

Received 6th October 2010, Accepted 12th January 2011

DOI: 10.1039/c0cp02061c

Using results from the dielectric theory of polar solids and liquids, we calculate the mesoscopic, spatially-varying dielectric constant at points in and around a protein by combining a generalization of Kirkwood–Fröhlich theory along with short all-atom molecular dynamics simulations of equilibrium protein fluctuations. The resulting dielectric permittivity tensor is found to exhibit significant heterogeneity and anisotropy in the protein interior. Around the surface of the protein it may exceed the dielectric constant of bulk water, especially near the mobile side chains of polar residues, such as K, N, Q, and E. The anisotropic character of the protein dielectric selectively modulates the attractions and repulsions between charged groups in close proximity.

1. Introduction

A quantitatively accurate theory for the dielectric properties of polar liquids and solids took form only after several decades of research, starting with the early work of Lorentz¹ and reaching predictive power with the theories of Kirkwood and Oster for fluids^{2,3} and Mott and Littleton⁴ for solids. Debye's original formulation⁵ followed Langevin's theory of paramagnetism and quantifies the earlier observations of Clausius and Mossotti for the dielectric constant ϵ of gases:

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi}{3} \sum_i n_i \left(\alpha_i + \frac{\mu_i^2}{3k_B T} \right). \quad (1)$$

Here the sum is over species of molecules, α_i and μ_i are the electronic polarizability and permanent electric dipole moment of species i , n_i is the number per cm³ of species i , and $k_B T$ is Boltzmann's constant times the temperature. This analysis yields a dielectric constant which increases as temperature is decreased due to the more effective alignment of dipoles against thermal randomization.

However, for a positive dielectric constant the left hand side of eqn (1) is bounded by unity, while the right hand side is not. In fact if one substitutes known values for the electronic polarizability, molecular dipole moment, and density at room temperature for a substance such as water, eqn (1) can only be satisfied by a negative dielectric constant. This is a consequence of the assumption of the Lorentz field for the local field in the model, *i.e.* the model predicts ferro-electricity for a

substance such as water below a temperature $T_c = 4\pi n \mu^2 / 9k_B T \approx 1900$ K analogous to Weiss ferromagnetism.

Onsager's treatment of the local field resulted in a reaction field which could polarize a molecule but not align it, and a cavity field which could provide torque on a dipolar molecule.⁶ This theory removed the dielectric catastrophe, but still predicted dielectric constants about half of the experimental values for substances such as water. Oster and Kirkwood's more explicit treatment of dipole–dipole correlations² predicted dielectric constants within a few percent for water by treating the alignment of a water molecule dipole with that of its neighbors.³ This treatment results in an increased dielectric constant when molecules align their neighbors in a ferromagnetic fashion, with Kirkwood's expression for the dielectric constant (for 1 species)

$$\frac{(\epsilon - 1)(2\epsilon + 1)}{12\pi n \epsilon} = \left[\alpha + \frac{\mu^2}{3k_B T} \left(1 + n \int_{v_0} d\Omega \cos \gamma e^{-W/k_B T} \right) \right] \quad (2)$$

reducing to that in the Onsager theory when the local effective dipole moment is treated at the mean field level, determined by the reaction field. In eqn (2), W is the potential of the mean force acting on a pair of molecules, γ is the angle between the dipole moments of a pair of molecules, and the integration is over all relative orientation and positions of the molecules within a sphere of volume v_0 .

A natural application for the theory of polar dielectric media is the study of electrostatic effects in biomolecules such as proteins, as these effects are key to their stability and function. A complete understanding of these effects requires an accurate description of protein dielectric properties, which determine the strength of interactions between charges in the protein. However, unlike a homogeneous liquid whose dielectric constant does not vary throughout its volume, the

^a Dept. of Physics and Astronomy, University of British Columbia, Vancouver, Canada. Tel: +1 (604)822-8813; E-mail: steve@physics.ubc.ca

^b Brain Research Centre, University of British Columbia, Vancouver, Canada

dielectric response of a biomolecule varies from site to site depending on the local molecular structure. Furthermore, complex constraint forces within the molecule may cause only partial alignment of local dipoles with an applied external field, introducing anisotropic effects.

The importance of biomolecule dielectric behavior in such fields as protein–protein interactions and enzyme reaction catalysis has led to interest in a method for calculating protein dielectrics that accounts for their varying local behavior. Explicit microscopic approaches to the calculation of the dielectric constant by extracting fluctuations appearing in Kirkwood–Fröhlich theory from molecular dynamics (MD) simulations have been developed by Wada *et al.*,⁷ van Gunsteren *et al.*,⁸ and Simonson and Perahia *et al.*^{9–11} Masunov and Lazardis¹² have calculated the potentials of the mean force between pairs of charged amino acid side chains and found that a uniform dielectric is unsatisfactory for explaining the results of explicit simulation. The results of explicit and implicit solvation for calculating protein pK_a 's have been compared,¹³ with the conclusion that the larger scale, anisotropic structural reorganization that can accompany (de)protonation is difficult to capture using Poisson–Boltzmann (PB) methods, but may be captured using molecular dynamics with a generalized Born implicit solvent. Voges and Karshikoff¹⁴ have provided a theory that enables the iterative calculation of a heterogeneous (but isotropic) dielectric constant in a small cavity containing part of a protein and have applied it to the calculation of amino acid dielectric constants.

The advent of automated PB equation solvers like APBS¹⁵ and DelPhi¹⁶ has enabled the rapid calculation of protein electrostatic energies. PB methods are used extensively in biomolecular simulation, including ligand docking studies for high-throughput drug screening¹⁷ and implicit solvent molecular dynamics.¹⁸ However, PB calculations commonly assume a constant internal dielectric environment for proteins which neglects local variation in susceptibility, while measurements such as pK_a shifts indicate a much richer profile for the effective dielectric constant in proteins.^{19,53} It is thus desirable to calculate the local dielectric constant at all points in and around a protein to use as input for these programs to enable a more accurate and descriptive calculation of protein electrostatic energies. Moreover, knowledge of the local dielectric function allows for an understanding of the mesoscopic structure of the susceptibility within and around a protein, which has consequences for many aspects of stability, folding, binding, and biological function. For these reasons we have developed a mesoscopic-scale theory to calculate the spatially-varying and anisotropic static dielectric constant in and around a protein. This theory has been used previously to investigate electrostatic contributions to regional stability in the prion protein.²⁰

2. Results and discussion

2.1 Protein dipoles in an applied field

The first step in deriving the local dielectric constant of the protein is to characterize its response to applied electric fields. The protein can be thought of as an assembly of fluctuating

dipoles at locations determined by the native protein fold. Unlike in the liquid case, where these dipoles are relatively free to orient with the prevailing applied electric field (subject to local organization of the liquid), stereochemical intramolecular forces constrain the motion of the dipoles in the protein, so they may only partially align with an applied field. Furthermore, dipoles in the protein do not move independently, as the coupling of fluctuations due to the above-mentioned steric constraints in the protein may cause the dipoles to react in a coordinated fashion.

In principle, any two atoms participating in a covalent bond in the protein may be viewed as a dipole, but the motions of atoms within the backbone and side chains of residues in a protein are highly correlated by the covalent bond network. We will thus define the effective dipoles as groups of atoms in each residue backbone or side chain. Exceptions are made for glycine, alanine, and proline, since their side chains are structurally incapable of motion substantially independent of their backbones; all the atoms in each one of these residues are considered as a single dipole.

In the absence of an applied electric field, these dipoles undergo thermal fluctuations that are not necessarily isotropic: there may be greater average motion in some directions compared to others. For example, fluctuations perpendicular to the time-average dipole orientation are typically greater than those parallel to the average dipole. Thus each dipole has its own system of principal axes characterizing the response of its three components to an external field.

2.2 Internal protein constraints

Let the protein be composed of N dipoles, each with dipole moment components in the x , y , and z directions. Construct a vector $\boldsymbol{\mu}$ of length $3N$ that contains the *deviations* of all the protein dipole moment components from their equilibrium values, such that the x , y , and z components of the i^{th} dipole's deviations are μ_{3i-2} , μ_{3i-1} , and μ_{3i} respectively. This means that $\langle \boldsymbol{\mu} \rangle_0 = 0$, where the angle brackets refer to the thermal average in zero external field.

In the presence of a local electric field $\mathbf{E} = (E_x, E_y, E_z)$ which can vary dipole to dipole, the change in free energy by perturbing the configuration of dipoles from their equilibrium positions is (to 2nd order in $\boldsymbol{\mu}$)

$$\Delta G = \frac{1}{2} \sum_{i,j=1}^{3N} K_{ij} \mu_i \mu_j - \sum_{i=1}^{3N} E_i \mu_i \quad (3)$$

where E_i are the components of the vector $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_N)$ representing the local field on all N dipoles, and K_{ij} are the second derivative matrix elements $\delta^2 G / \delta \mu_i \delta \mu_j|_0$ evaluated at the equilibrium position $\boldsymbol{\mu} = 0$.

The probability for the system to occupy such a configuration is proportional to $\exp(-\Delta G/k_B T)$. Thus the averages of the induced dipole moment and cumulant matrix elements can be found by diagonalizing the free energy,²¹ and are given by

$$\langle \mu_i \mu_j \rangle_c = k_B T (K^{-1})_{ij} \quad (4a)$$

$$\langle \mu_i \rangle = \sum_j (K^{-1})_{ij} E_j = \frac{1}{k_B T} \sum_j \langle \mu_i \mu_j \rangle_c E_j. \quad (4b)$$

Since the average of $\boldsymbol{\mu}$ is 0 in the absence of an external field, the cumulants above may be replaced by the unperturbed averages $\langle \dots \rangle_0$.

So long as the free energy in the above analysis is linear in the field strength (linear response), the statistics of the dipole fluctuations need not be Gaussian, and consequently the local potential of the mean force need not be harmonic. To see this, it is sufficient to consider an isolated dipole $\boldsymbol{\mu}$ in the presence of a local field \mathbf{E} , with an arbitrary unperturbed probability distribution $P_0(\boldsymbol{\mu})$ and $\langle \boldsymbol{\mu} \rangle_0 = 0$. In the presence of a weak field the probability distribution becomes $P_0 e^{-\boldsymbol{\mu} \cdot \mathbf{E}/k_B T} \approx P_0 (1 - \boldsymbol{\mu} \cdot \mathbf{E}/k_B T)$. The thermal average of the dipole moment $\boldsymbol{\mu} = \mu_1 \hat{\mathbf{i}} + \mu_2 \hat{\mathbf{j}} + \mu_3 \hat{\mathbf{k}}$ then has components $\langle \mu_i \rangle = \sum_j \langle \mu_i \mu_j \rangle E_j / k_B T$, which is precisely eqn (4b). Thus, whatever the statistics of the dipoles, eqn (4b) gives the average moment in the presence of a weak field.

In the all-atom molecular dynamics simulations of proteins described below, it was observed that most dipoles were tightly bound by harmonic potentials, with mean fluctuations much less than the total dipole magnitude. However, some polar amino acids near the protein surface underwent significant rearrangement due to the lack of steric constraints. In this sense the protein is more liquid-like on its surface than in its interior. The probability distributions of all dipoles in simulation were compared to normal distributions by the Lilliefors test.²² Fig. 1A shows the distribution of Lilliefors test statistics from all of the dipole probability distributions of ubiquitin, taken from a 20 ns all-atom classical molecular dynamics (MD) simulation in an explicit solvent of the native state of ubiquitin. Simulations were performed using the NAMD simulation package,²³ with the CHARMM22 force field.^{24,25} More details of the simulation protocol are described in the section ‘‘Implementation in a Protein System’’ below. The significant majority of dipole probability distributions closely followed the normal distribution, although some dipole modes exhibited decidedly non-Gaussian potentials, often bi- or tri-modal, indicating multiple energetic minima for these dipoles. These multiple minima correspond to different metastable configurations of the amino acid side chain (Fig. 1A inset), however as mentioned above linear response does not require the fluctuation distribution to be Gaussian.

2.3 Collective dipole fluctuation modes

If the motion of each dipole were uncorrelated with other dipoles, then the eigenbasis of fluctuations would be the principal axes of the individual dipoles. The actual eigenbasis taken from all-atom molecular dynamics simulations can be projected onto this individual-dipole basis to determine the degree of coupling between dipoles.

In the individual-dipole basis $|\phi\rangle$, define a fluctuation eigenvector $|\psi\rangle = \sum_{\phi} a_{\phi} |\phi\rangle$, where by normalization $\sum_{\phi} |a_{\phi}|^2 = 1$. Each modulus $|a_{\phi}|^2$ can be interpreted as the probabilistic weight p_{ϕ} of the eigenvector $|\psi\rangle$ in the basis vector $|\phi\rangle$. Borrowing the concept from spin-glass theory of the average cluster size of spin glass states,²⁶ we let

$$M = \frac{1}{3} \left(\sum_{\phi=1}^{3N} p_{\phi}^2 \right)^{-1} \quad (5)$$

denote the degree of coupling between individual-dipole fluctuation modes. When a mode is uncoupled, it has a weight distribution given by a Kronecker delta and so $M = 1/3$. If a mode is fully coupled to all individual-dipole modes, $p_{\phi} = 1/(3N)$ and so $M = N$, the total number of dipoles.

Fig. 1B plots the distribution of M for ubiquitin. The number of dipoles participating in each mode varies from 1 to 20, with 70% of modes containing less than 10 dipoles. Thus dipole motion exhibits coordination between moderately sized groups of neighboring dipoles, and only relatively few dipoles move independently of other dipoles; these independent dipoles tend to be less sterically constrained and reside on the protein surface, as seen in the inset of Fig. 1B. It is therefore important to consider collective dipole modes in proteins to arrive at an accurate response relation.

2.4 Linear response relation for induced moments

We calculate the effective local dielectric constant at a point in a protein by considering the equivalence between microscopic and macroscopic descriptions of the electric response of nearby media as depicted schematically in Fig. 2. In the microscopic description, we imagine the matter within a region of radius a of this point to have a dipole moment \mathbf{m} and tensor polarizability $\bar{\alpha}$, placed in a cavity of the same radius within an environment consisting of various scalar dielectric constants, accounting for the response of the water and/or protein surrounding the cavity and represented generically as $\epsilon_A, \epsilon_B, \epsilon_C, \dots$ in Fig. 2. The average dielectric of the regions surrounding the cavity is ϵ_1 . In the macroscopic description, this cavity is instead filled with a dielectric medium of permittivity $\bar{\epsilon}_2$, again surrounded by the dielectric ϵ_1 (see Fig. 2). Following the approach taken by Voges and Karshikoff,¹⁴ we solve for $\bar{\epsilon}_2$ in terms of ϵ_1 , \mathbf{m} , $\bar{\alpha}$, and a .

In practice, we discretize the space in and around the protein into a lattice with spacing b , with b typically about 1 Å. For each lattice point at \mathbf{r} we consider a spherical cavity centered at \mathbf{r} of radius a , with a typically a few Å. The cavity may contain parts of several dipoles, and has inside it a local field $\mathbf{E}(\mathbf{r}|a)$ due to both the external field and the system’s response. All dipoles in a given cavity are taken to experience the same total field. We take the contribution of each dipole to the induced cavity dipole moment to depend on the volume fraction of the backbone or side chain containing the dipole that is within the cavity. Let $f_A(\mathbf{r}|a)$ be the volume fraction of residue A inside the cavity centered at position \mathbf{r} , given the cavity has radius a . To obtain the static dielectric response, $f_A(\mathbf{r}|a)$ should be the time-averaged fraction of A in the cavity. The i^{th} component of the field-induced moment inside the cavity is given by a sum over both residues and components. It is clearest to write the sums separately, rewriting eqn (4b) as $\langle \mu_i^A \rangle = (k_B T)^{-1} \sum_{B=1}^N \sum_{j=1}^3 \langle \mu_i^A \mu_j^B \rangle E_j^B$ for the i^{th} component of the dipole of residue A . The induced moment of the protein dipoles in the cavity, $\mathbf{m}_p(\mathbf{r}|a)$, is given by the sum of the induced moments of all residues weighted by the fraction of those residues inside the cavity:

$$\mathbf{m}_p(\mathbf{r}|a) = \sum_{A=1}^N \langle \mathbf{m}^A(\mathbf{r}|a) \rangle \equiv \sum_{A=1}^N f_A(\mathbf{r}|a) \langle \boldsymbol{\mu}^A \rangle. \quad (6)$$

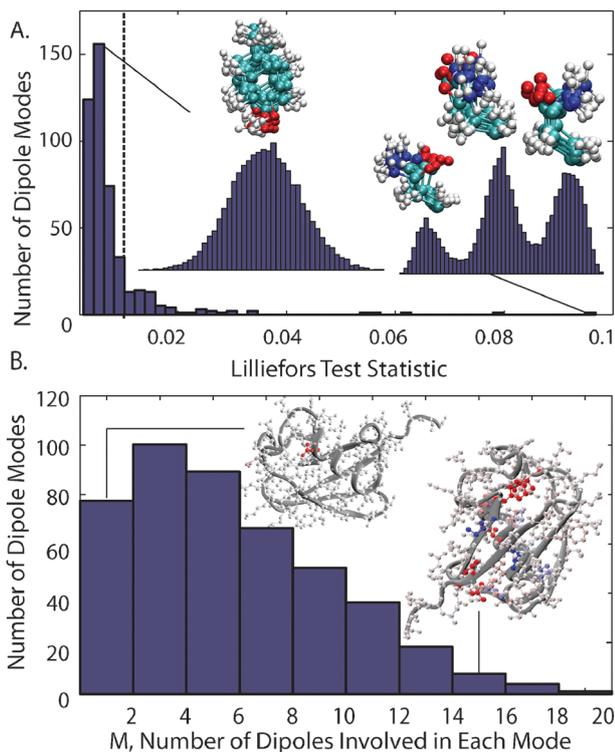


Fig. 1 Dipole fluctuation statistics. (A) The distribution of Lilliefors test statistics for ubiquitin dipoles. Values greater than the dotted line indicate non-normal distributions with 95% confidence. Representative normal and non-normal dipole distributions for the side chains of Y59 and Q31, respectively, along with several molecular configurations, are shown in the insets. The values of the Lilliefors test statistic for these side chains are also indicated. (B) The distribution of the spin-glass parameter M (a measure of the effective number of dipoles involved in each fluctuation mode, see text) for the dipole fluctuation modes in ubiquitin. Inset images show the residues involved in localized and collective modes. Residues are color-coded so as to indicate whether their motions are correlated (same color) or anticorrelated (different colors).

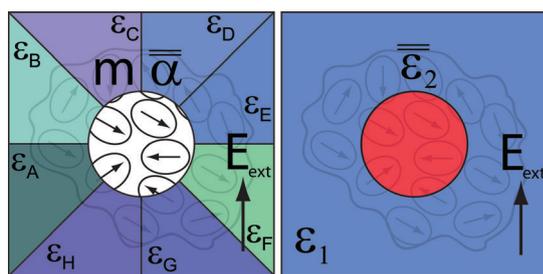


Fig. 2 Schematic representation of the approach used to calculate the dielectric constant $\bar{\epsilon}_2$. In the microscopic view, a cavity of radius a containing media with induced dipole moment m and tensor polarizability $\bar{\alpha}$ is surrounded by a heterogeneous dielectric of various permittivities (in this particular case $\epsilon_A \dots \epsilon_H$; the number of different dielectric regions will depend on the sphere size and lattice point spacing). In the macroscopic view, this cavity is instead filled with an anisotropic dielectric tensor of permittivity $\bar{\epsilon}_2$ surrounded by an effective homogeneous isotropic dielectric ϵ_1 . In both views, an arbitrary external electric field \mathbf{E}_{ext} is applied.

The cavity that determines the local dielectric may also be near the surface of the protein, where it will contain a number of solvent molecules (usually water). In this case the sum on residues in the cavity includes a contribution due to the water molecules inside it. The number of water molecules $n_w(\mathbf{r}|a)$ in a cavity is determined by taking the available volume in the sphere but outside the protein, and dividing by the average volume of a water molecule at STP.

The electronic polarizability of the media depends on the proportions of protein backbones, side chains, and water in the cavity. Analogously with the permanent dipole response, the total electronic polarizability in the cavity is weighted by volume fractions:

$$\alpha(\mathbf{r}|a) = \sum_{A=1}^N f_A(\mathbf{r}|a)\alpha^A + n_w(\mathbf{r}|a)\alpha^w. \quad (7)$$

Scalar values of the electronic polarizability α for each residue are taken from the literature²⁷; they cannot be measured directly from traditional classical MD simulations because atomic partial charges are fixed by the CHARMM22 parameter set.

We take Kirkwood's analysis as a starting point to determine the contribution of water to the cavity's dipole moment, in which the induced moment due to permanent dipole reorientation is given by

$$\mathbf{m}_w(\mathbf{r}|a) = n_w(\mathbf{r}|a) \frac{gp^2}{3kT} \mathbf{E}_e. \quad (8)$$

In this equation, p is the permanent dipole moment of water and \mathbf{E}_e is the local effective field orienting the molecules. The constant g arises from the Kirkwood–Oster nearest-neighbor approximation of the term in parentheses on the right-hand side of eqn (2). It has been calculated previously for water and found to be 2.67.³ Constrained motion of water molecules, relative to that in bulk,²⁸ has been observed at the surfaces of proteins.^{29,30} The heterogeneity of hydrogen bonding between protein and water has been studied by Bagchi and co-workers²⁹ with particularly long lifetimes observed near positively charged residues, as well as reduced hydration layer rigidity near functionally-relevant sites on a villin headpiece subdomain. Such constrained and correlated motions may effectively increase the local dielectric of water near the protein surface if the local moments are positively correlated, and several examples of this effect are discussed below. A detailed investigation of the water structure at the protein surface and its consequences on the dielectric is a topic of future work. Here we make the simplifying assumption that water dipole correlations are essentially the same as those occurring in bulk.

In what follows, a relationship analogous to eqn (8) is derived with the refinement that the local electric field be reinterpreted as a field proportional to the cavity field. In a cavity containing protein and water, the local field \mathbf{E}_e experienced by the water and protein dipoles consists of a cavity field \mathbf{G} due to the externally applied perturbing field \mathbf{E}_{ext} and a reaction field \mathbf{R} due to the response of the medium outside the cavity (with an assumed average scalar dielectric ϵ_1) to the induced dipole within the cavity:

$$\mathbf{E}_e = \mathbf{G} + \mathbf{R} = \frac{3\epsilon_1}{2\epsilon_1 + 1} \mathbf{E}_{\text{ext}} + \mathcal{F}(\alpha \mathbf{E}_e + \mathbf{m}) \quad (9)$$

Here, $\mathcal{F} = 2(\varepsilon_1 - 1)/((2\varepsilon_1 + 1)a^3)$, α is the total electronic polarizability of the cavity from eqn (7), and $\mathbf{m} = \mathbf{m}_p + \mathbf{m}_w$ is the total dipole moment due to the positions of atomic nuclei inside the cavity. Solving for \mathbf{E}_e ,

$$\mathbf{E}_e = \frac{1}{1 - \alpha\mathcal{F}} (\mathbf{G} + \mathcal{F}\mathbf{m}) \equiv \gamma(\mathbf{G} + \mathcal{F}\mathbf{m}), \quad (10)$$

where for convenience, $\gamma \equiv 1/(1 - \alpha\mathcal{F})$.

The total potential energy of the protein dipole component in the cavity is the sum of the electric potential energy $-\mathbf{m}_p \cdot \mathbf{E}_e$ and the steric potential energy $U_S(\mu_p) = \frac{1}{2}K_{ij}^{AB} \mu_i^A \mu_j^B$, where the implied sums on A and B run from 1 to N (the number of BB + SC moieties) and the sums on i and j run from 1 to 3. The steric potential constants K_{ij}^{AB} taken from simulation implicitly include the protein dipole self-interaction term $\gamma\mathcal{F}f_A f_B \mu_i^A \mu_j^B$ due to the effect of the protein dipole reaction field on the moment \mathbf{m}_p itself. Using eqn (3), (6), and (10),

$$U(\mathbf{m}_p) = \frac{1}{2}K_{ij}^{AB} \mu_i^A \mu_j^B - \gamma\mathcal{F}f_A \mu_i^A m_i^w - \gamma f_A \mu_i^A G_i, \quad (11)$$

where the cavity field is applied only to dipoles in the cavity, resulting in the prefactor f_A in the third term of eqn (11). The total steric potential energy must be included to properly account for the statistics of dipoles inside the cavity. Thus eqn (11) can be thought of as a hybrid potential energy. The potential energy of the water, on the other hand, is determined only by the total effective field, as there are no internal steric constraints assumed on its motion (except as embodied in the Kirkwood g -factor in eqn (8)). This approximation can be refined by explicitly considering the simulated statistics of water molecules at the protein surface. The water dipole self-interaction term $\gamma\mathcal{F}\mathbf{m}_i^w \cdot \mathbf{m}_i^w$ is zero in the case of isotropic polarizability α since the reaction field produced by the water dipole is parallel to the dipole itself and therefore cannot apply a torque to it. We neglect effects of the distensibility in magnitude of the nuclear part of the water dipole moment. The water dipole potential energy is

$$U(\mathbf{m}_w) = -\gamma\mathcal{F}f_A \mu_i^A m_i^w - \gamma m_i^w G_i. \quad (12)$$

Thus the potential energy of water and protein dipoles is a sum of terms bilinear in the water and protein dipoles and linear in the effective cavity field $\gamma\mathbf{G}$. This has the form of the problem solved above in eqn (4b), so the i^{th} component of the field-induced water and protein moments in the cavity are

$$\langle m_i^A \rangle = \sum_{j=1}^3 \left(\sum_{B=1}^N \langle m_i^A m_j^B \rangle_0 + \langle m_i^A m_j^w \rangle_0 \right) \frac{\gamma G_j}{k_B T} \quad (13a)$$

$$\langle m_i^w \rangle = \sum_{j=1}^3 \left(\sum_{B=1}^N \langle m_i^w m_j^B \rangle_0 + \langle m_i^w m_j^w \rangle_0 \right) \frac{\gamma G_j}{k_B T}, \quad (13b)$$

where the time average $\langle \dots \rangle_0$ is taken in the absence of an external perturbing field. We have used the generality of the potential elaborated in the comments below eqn (4b). The dipole polarizability to the cavity field is a property only of correlations within the system itself. Only mutual reaction fields influence the motion of the dipoles in this case. The correlation functions involving water in eqn (13a) and (13b) can be evaluated by direct integration. The integration for protein dipoles is over all space, while it is confined to a sphere

of radius p for the water dipoles, which can reorient but are fixed in magnitude. The potential energies in eqn (11) and (12) appear in a Boltzmann factor with the cavity field $\mathbf{G} = 0$; those Boltzmann factors not containing K_{ij} are small compared to $k_B T$ and may be linearized to give

$$\langle m_i^A \rangle = \sum_{B=1}^N \sum_{j=1}^3 f_A f_B \left(1 + \mathcal{F} \frac{n_w g p^2}{3k_B T} \right) \frac{\langle \mu_i^A \mu_j^B \rangle_0}{k_B T} \gamma G_j \quad (14a)$$

$$\langle m_i^w \rangle = \sum_{j=1}^3 \left(\frac{n_w g p^2}{3k_B T} \right) \left(\delta_{ij} + \sum_{A,B=1}^N \mathcal{F} f_A f_B \frac{\langle \mu_i^A \mu_j^B \rangle_0}{k_B T} \right) \gamma G_j \quad (14b)$$

Eqn (14a) and (14b) can be combined and written generally as a matrix equation, with a nuclear polarizability tensor $\bar{\alpha}$ relating the effective field $\gamma\mathbf{G}$ and induced moment $\mathbf{m} = \mathbf{m}_w + \mathbf{m}_p$:

$$\mathbf{m}(\mathbf{r}|a) = \bar{\alpha}(\mathbf{r}|a) \cdot \gamma(\mathbf{r}|a) \mathbf{G}(\mathbf{r}|a), \quad (15a)$$

with components

$$\bar{\alpha}_{ij}^{AB}(\mathbf{r}|a) = f_A f_B \left(1 + 2\mathcal{F} \frac{n_w g p^2}{3k_B T} \right) \frac{\langle \mu_i^A \mu_j^B \rangle_0}{k_B T} + \delta_{ij} \frac{n_w g p^2}{3k_B T} \quad (15b)$$

The quantities $\langle \mu_i^A \mu_j^B \rangle_0$ may be obtained directly from MD simulations of the protein in the absence of an external field, as described below.

2.5 The dielectric constant at an arbitrary point

With the response of permanent dipoles and polarizable media in a cavity now established, we calculate the dielectric permittivity tensor at a location \mathbf{r} by following the recipe outlined in Fig. 2. In a microscopic description, a set of polarizable constituents with induced and permanent dipole moments exist in a cavity of a heterogeneous dielectric medium with various scalar dielectric constants. The present theory approximates the medium external to the cavity by a single scalar dielectric ε_1 equal to the average of the neighboring effective scalar dielectrics over the surface of the cavity ($\varepsilon_A, \varepsilon_B, \varepsilon_C, \dots$ in Fig. 2). The medium inside the cavity is assigned a single tensor dielectric $\bar{\varepsilon}_2$ because the polarizability in the cavity is a tensor. As described below, we take the effective scalar value of the cavity's dielectric to be the geometrical average of its principal components.

The total field \mathbf{E}_{in} inside the cavity at a position \mathbf{r} is the superposition of the cavity field, the reaction field, and the permanent and induced dipole fields from the water and protein,¹⁴ so

$$\mathbf{E}_{\text{in}} = \mathbf{G} + \mathcal{F}(\alpha\mathbf{E}_e + \langle \mathbf{m} \rangle) - \nabla \left(\frac{\alpha\mathbf{E}_e \cdot \mathbf{r}}{r^3} + \frac{\langle \mathbf{m} \rangle \cdot \mathbf{r}}{r^3} \right). \quad (16)$$

On substituting for $\langle \mathbf{m} \rangle$ from (15a) and \mathbf{E}_e from (10), the potential inside the cavity is

$$\Phi_{\text{in}} = [(1 + \gamma\mathcal{F}\alpha)(I + \gamma\mathcal{F}\bar{\alpha})\mathbf{G}] \cdot \mathbf{r} + [(\bar{\alpha}\gamma + \mathcal{F}\bar{\alpha}\gamma\alpha + \alpha\gamma)\mathbf{G}] \cdot \frac{\mathbf{r}}{r^3}. \quad (17)$$

The potential outside the cavity is formed by the superposition of the potentials from the external field (taken to be uniform),

the field due the cavity in the dielectric, and the field due to the dipole in the cavity:

$$\Phi_{\text{out}} = -\mathbf{E}_{\text{ext}} \cdot \mathbf{r} + a^3 \frac{\mathbf{r}}{r^3} \cdot \mathbb{W} \mathbf{E}_{\text{ext}}, \quad (18)$$

where \mathbb{W} is defined by

$$\mathbb{W} \equiv \frac{9\epsilon_1}{(2\epsilon_1 + 1)^2 a^3} (\bar{\alpha}\gamma + \mathcal{F} \bar{\alpha}\gamma\alpha\gamma + \alpha\gamma) - \frac{\epsilon_1 - 1}{2\epsilon_1 + 1} I. \quad (19)$$

Shifting to the equivalent macroscopic description of the system as a dielectric with permittivity tensor $\bar{\epsilon}_2$ surrounded by a dielectric with scalar permittivity ϵ_1 , the potential in the surrounding dielectric is found to be³¹

$$\Phi_{\text{out}} = -\mathbf{E}_{\text{ext}} \cdot \mathbf{r} + a^3 (2I + \epsilon_1^{-1} \bar{\epsilon}_2)^{-1} (\epsilon_1^{-1} \bar{\epsilon}_2 - I) \mathbf{E} \cdot \frac{\mathbf{r}}{r^3}. \quad (20)$$

Equating the microscopic and macroscopic expressions for the potential outside the cavity and solving for $\bar{\epsilon}_2$,

$$\bar{\epsilon}_2 = \epsilon_1 (2\mathbb{W} + I) (I - \mathbb{W})^{-1}. \quad (21)$$

In the case where \mathbb{W} is a scalar and the dipole response in eqn (15a) is approximated by freely rotating Langevin dipoles (a potentially severe approximation due to steric constraints in the protein as mentioned above), eqn (21) reduces to that in the theory of Voges and Karshikoff.¹⁴

2.6 Implementation in a protein system

To calculate the local dielectric constant according to the approach described here, the first step is to obtain the matrix of correlations $\langle \mu_i^A \mu_j^B \rangle_0$ appearing in eqn (15b) that describes coupled fluctuations between the dipoles in the native state. This is done with an all-atom classical MD simulation of various proteins at 298 K using the CHARMM22 force field,^{23–25} with particle-mesh Ewald electrostatics and a Lennard-Jones cut-off distance of 13.5 Angstroms. Proteins are solvated in a box of explicit water molecules that exceeds the dimensions of the native protein by 10 Angstroms on all sides and has periodic boundary conditions. Basic residues (Lys and Arg) are protonated, acidic residues (Asp and Glu) are deprotonated, and histidines are neutrally charged to reflect ionization conditions at pH 7. Na^+ and Cl^- ions are added to the solvent to achieve overall system charge neutrality and an ionic strength of 150 mM. The simulation time step is 2 fs, and snapshots are taken every 1 ps for ensemble averaging. As seen in Fig. 3B, a 1 ps interval is sufficient for dipole positions to decorrelate from those in the previous snapshot. The total simulation time required for convergence to reliable values was typically 1 ns; longer simulations did not appreciably change the distribution of correlations (Fig. 3C). All simulations to calculate the spatially-varying dielectric function were run for 2 ns. The dipole moment of each side chain and backbone is calculated from the partial charges assigned to atoms in the CHARMM force field, with distances to atoms measured from the center of mass of the set of atoms.

As one would expect, components of the same dipole exhibit significant autocorrelation (Fig. 3A diagonal), but there are also significant cross-correlations between dipoles distant in sequence but spatially close in the native protein structure.

A band of large fluctuations is typically observed at the N- and C-terminus of the protein due to its high flexibility.

After MD simulation of dipole fluctuations is complete, the protein is centered in a rectangular box with dimensions that exceed the minimum and maximum x , y , and z coordinates of the protein on all sides by at least the cavity radius a . The dielectric constant is calculated at each lattice point within this box (generally with a 1 Å spacing). The effective surrounding dielectric constant ϵ_1 used for each point is determined by averaging the scalar dielectric constant of all points in a shell within 0.5 Å of the cavity boundary. An iterative solution is necessary, since the dielectric constant at one point depends on the dielectric constant at surrounding points. An approximate function of $\epsilon = 10$ inside the protein and $\epsilon = 78$ outside was used as an initial dielectric function for the first iterations, and the system was then iteratively relaxed until the spatially-varying dielectric function converged, typically after less than 20 iterations. We found that the final values of the dielectric function were independent of the choice of initial conditions.

The dielectric calculation program is implemented in Tcl and MATLAB (The MathWorks, Natick, MA). For a protein of length 100 residues, the initial simulation to obtain the dipole correlation matrix takes roughly 12 hours on an 8-core 2.5 GHz Intel Xeon workstation, while the calculation of the dielectric constant afterwards takes less than 30 minutes running on one core with a lattice point spacing of 1 Å. The time needed to calculate the dielectric function varies linearly with the number of lattice points.

2.7 Dielectric anisotropy

As demonstrated above, to properly capture the behavior of a protein the local dielectric constant must be a tensor. However, for many practical applications, it is desirable to have an equivalent scalar dielectric constant that replicates the behavior of the tensor as well as possible. The best choice of method for converting the tensor $\bar{\epsilon}_2$ to the scalar ϵ_2 may depend on the situation, but as shown by Mele,³² the transmission of free charge fields into an anisotropic dielectric depends on the geometric mean of the dielectric constants in each direction. That is, if λ_1 , λ_2 , and λ_3 are the eigenvalues of $\bar{\epsilon}_2$, we define the equivalent scalar ϵ to be $\epsilon = (\lambda_1 \lambda_2 \lambda_3)^{1/3}$. (It is worth noting that if all three eigenvalues are equal, their harmonic, geometric, and arithmetic means are the same. Even if one eigenvalue exceeds the other two by 50%, among the most pronounced anisotropy we have seen in our dielectric calculations for 25 proteins, the difference between the harmonic, geometric, and arithmetic means is still less than 5%. Thus the choice of approach for averaging the principal axes of the dielectric scalar to produce a scalar does not significantly affect the quantitative result.)

2.8 Spatial variation in protein dielectric response

An example of a dielectric map for adenylate kinase (PDB 1AKY) is shown in Fig. 4. Panel A depicts the scalar dielectric function as a surface plot for a slice through the protein, while panel B shows the scalar dielectric function as a 3D isosurface plot. Panel C depicts the regions of anisotropy

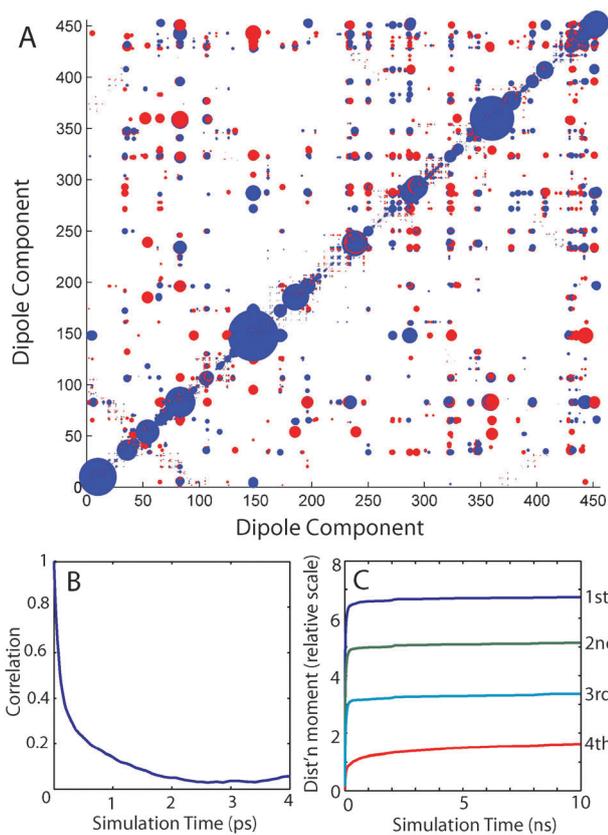


Fig. 3 Dipole correlations and convergence in simulation. (A) Correlation map for dipoles in ubiquitin. The area of the circles at each coordinate (i, j) indicates the magnitude of the correlation function $\langle \mu_i \mu_j \rangle$ averaged over snapshots of the system, for all pairs of dipole components $\{\mu_i, \mu_j\}$ in ubiquitin. The numbering of the components runs from the N- to the C-terminus and accounts for the x -, y -, and z -components of each backbone and side chain dipole. Blue indicates $\langle \mu_i \mu_j \rangle > 0$; red indicates $\langle \mu_i \mu_j \rangle < 0$ (anticorrelation). (B) The correlation coefficient between dipole products, $\text{corr}(\mu_i(0)\mu_j(0), \mu_i(t)\mu_j(t))$ as a function of the time t between frames from the simulation. The dipole motion decorrelates with a time constant of < 0.5 ps. (C) The 1st, 2nd, 3rd, and 4th moments of the distribution of dipole correlation values for a 10 ns simulation of ubiquitin. The dipole correlations converge to a stable distribution after a total simulation length of 1 ns.

in the dielectric function, which tend to be localized around the surface of the protein.

A notable feature of the heterogeneous protein dielectric theory is the presence of regions with relative permittivity comparable to or exceeding that of water on the surface of the protein, as can be seen from Fig. 4B. The solvation energy of a charged group varies inversely with the solvent dielectric constant, so the presence of these regions lowers the potential energy of protein surface charges and enhances protein stability. They arise from the presence of charged or polar groups with large dipole moments on the protein surface that can fluctuate extensively, as there are fewer native steric constraints restricting their motion.

Large values of the dielectric constant approaching that of the solvent have been observed on the surface of proteins,^{8,10} and values of the effective dielectric constant approaching 150 have been seen for salt bridges on the surface of barnase.^{33,34}

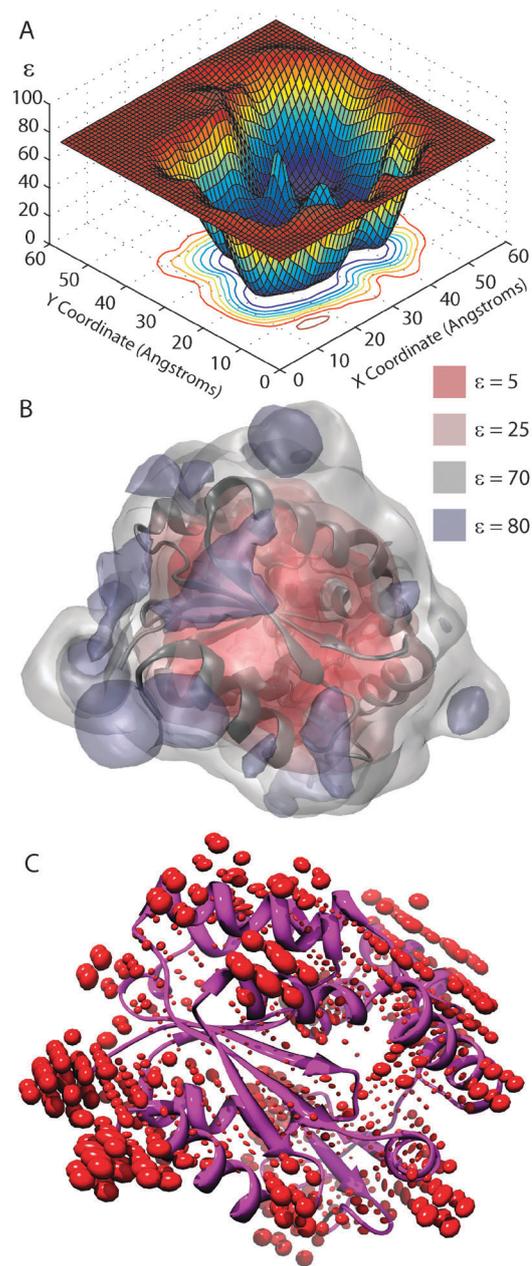


Fig. 4 The spatially-varying dielectric function for adenylate kinase (1AKY). (A) The effective scalar dielectric constant on a horizontal plane through the geometric center of the protein. (B) Dielectric contours around the 1AKY structure, showing surfaces of $\epsilon = 5, 25, 70$ and 80 . Regions inside the blue globules have dielectric constants larger than that of water. (C) Representation of the anisotropic dielectric constant $\bar{\epsilon}$. The orientation of each ellipsoid is given by the eigenbasis of the dielectric tensor at that point; the lengths of the semimajor axes are directly proportional to the eigenvalues of the tensor. Only ellipsoids with a difference between eigenvalues of $> 25\%$ are shown.

Dielectric values much greater than water have also been observed just outside the charged head groups of lipid bilayers.^{35,36} Having the surface of the protein surrounded by this region of high dielectric constant would attenuate the projection of electric fields from solution into the protein and *vice versa*, potentially reducing electrostatic attractions or repulsions

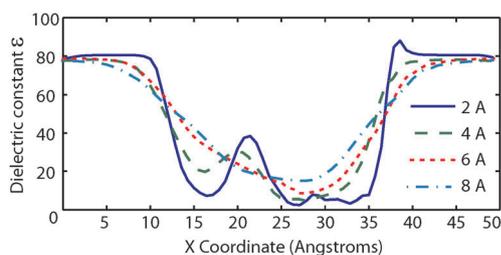


Fig. 5 The effect of cavity sphere radius a on the calculated dielectric function. Plots are taken on a line through the geometric center of ubiquitin (1UBQ), for $a = 2, 4, 6,$ and 8 \AA .

between nearby proteins. This reduces the effects of generic electrostatically-driven interactions which may be involved in protein aggregation, potentially allowing for higher intracellular protein concentration. A variable dielectric profile on the protein surface may also allow for optimization of specific binding strategies for ligands or protein complexes.

Additionally, the dielectric function calculated by the present theory varies from over 100 in places on the protein surface to as low as 2 in the hydrophobic protein interior (see Fig. 4B and Fig. 6). Other studies have reported values in this range for the effective dielectric constant of proteins: $\epsilon \approx 2$ for PARSE parameter sets,³⁷ $\epsilon \approx 4$ from bulk measurements of anhydrous protein on application of the Kirkwood–Frohlich theory to an idealized protein,³⁸ $\epsilon = 2\text{--}8.9$ by site-dependent thermodynamic integration/molecular dynamics studies,³⁹ and $\epsilon \approx 20$ for best agreement with the experimental pK_a s of titrable groups in proteins.⁴⁰ The wide range of measured and calculated dielectric constants in previous studies reaffirms the considerable

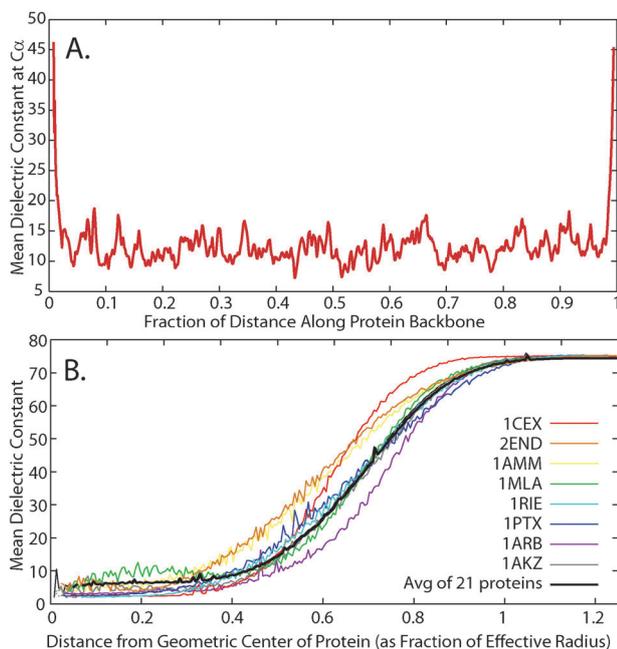


Fig. 6 Mean properties of the protein dielectric function. (A) Average dielectric constant as a function of fractional distance along the protein backbone from a set of 21 proteins.⁴⁸ Note the increased permittivity at the N- and C-termini due to their large flexibility. (B) Average dielectric constant as a function of the fractional distance from the protein geometric centre.

heterogeneity of protein dielectric response and may arise from differences in the local environment where the processes investigated take place.

The timescale for relaxation processes is also important. Nuclear polarization due to protein dipole relaxation, which dominates the dielectric response on the protein surface, happens on a timescale of several picoseconds (see Fig. 3B); electronic polarization due to electron motion, which occurs throughout the molecule, happens much faster. Processes occurring on a timescale longer than several picoseconds would therefore experience both nuclear and electronic polarization effects, while processes on shorter timescales would experience only electronic polarization effects and a consequently lower value for the effective dielectric constant. Frequency-dependent relaxation plays a role in the electrostatics of enzyme catalysis^{41,42}; similarly, a frequency-dependent friction coefficient has been seen to strongly affect reaction rates or transition states in diverse systems ranging from gas-phase and condensed phase reactions^{43–45} to protein folding.^{46,47} A frequency-dependent relaxation response could be obtained from the spectrum of normal mode relaxation in Fig. 1; investigation of these topics is reserved for future studies.

2.9 Dependence on sphere size & lattice point spacing

The cavity radius a is an adjustable parameter in this approach. A smaller value of a gives a more local description of the dielectric response of the protein but suffers from the application of a macroscopic description to the atomic-scale behavior within a smaller cavity. Conversely, a larger value of a may properly capture the effective macroscopic response of a protein region but conceal important shorter-length phenomena. In Fig. 5, the calculated dielectric function on a line through the middle of ubiquitin is shown for various cavity radii a . Based on these observations, we use a cavity radius of 3–4 Å as an optimal length scale to capture both the locally average behavior and the mesoscopic dielectric structure.

The choice of cavity radius determines the spacing of lattice points in the calculation, since it is necessary to have an adequate density of them near the surface of the cavity to accurately reflect the nature of the surrounding dielectric. We have found that once the lattice point spacing is 1/4 of the cavity radius, the dielectric map thus produced has converged in that it no longer changes with an increasing density of lattice points. We thus choose a lattice spacing of 1 Å.

2.10 Averaged dielectric properties

Protein N- and C-termini tend to have high flexibility, and for 1AKY the ends also have high net charge, so the dielectric function tends to be larger in these regions as well. To see whether this is a general trend, we plot in Fig. 6A the dielectric constant averaged over proteins, as a function of sequence index. So that different length proteins may be compared, the index is chosen to start at zero, and is normalized by $N - 1$ where N is the number of residues. One can see from the plot that the dielectric constant is, on average, significantly larger at the ends of the protein.

To investigate how the dielectric function varies as one moves from a protein's interior to its surface, we plotted the scalar dielectric constant at a distance r from the geometric center of the protein, averaged over the surface of the sphere of radius r , *i.e.* $\langle \epsilon(r) \rangle \equiv \sum' \epsilon(\mathbf{r}) / 4\pi r^2 \Delta r$ where all points in a spherical shell between r and $r + \Delta r$ are summed. A plot of this is shown in Fig. 6B for several proteins indicated. To investigate the general trend across proteins, this quantity was then averaged again over a dataset of 21 proteins⁴⁸ to obtain a protein-averaged dielectric constant as a function of radius. To compare differently sized proteins, the radius r was normalized by the effective protein radius r_p , defined as the radius of a sphere that would have the same radius of gyration r_G as the protein, *i.e.* $r_p = \sqrt{5/2} r_G$. The resulting quantity $\langle \epsilon(r/r_p) \rangle_{\text{prot}}$ is also plotted in Fig. 6B. It is worth noting that at a given radius r within a given protein, there is significant scatter in the data.

3. Conclusions

Electrostatic interactions between charges are critical in determining the stability of a protein. The strength of these interactions is modulated by the local environment around the charges, which can relax or polarize in response to the electric fields. This “dielectric screening” weakens forces between charges. We found that in the interior of a protein, the dielectric is not constant but instead is spatially heterogeneous, with many local minima and maxima. Moreover, our studies show that the polarizability of an amino acid is context-specific and large on the surface of the protein, where the local dielectric constant can be even larger than that of water. These regions can thus act as “stability shells” for charges, because charges tend to migrate towards higher dielectrics. We found that the dielectric response inside a protein tended to be direction-dependent.

This theory fits in the middle of the microscopic-to-macroscopic continuum of techniques to describe biomolecule electrostatic properties. It is not fully microscopic in that individual atoms are collected into backbone and side chain dipoles to improve computational efficiency, and the applied fields are assumed to be approximately uniform over distances of a few Å; conversely, by allowing the dielectric characteristics of a protein to vary throughout its volume it captures subtleties in electric effects that a purely macroscopic model would efface. It is useful to have a robust and versatile tool for capturing much of the microscopic electrostatic behavior in a simple parameter like a locally-varying dielectric constant, which may then be refined by MD simulation or density functional methods to explore interesting or noteworthy effects identified by the mesoscale method. An alternative approach of extracting local electrostatic properties directly from all-atom MD simulation often requires the subtraction of large quantities of comparable magnitude, introducing large errors that require long simulations to satisfactorily average. Moreover, the most popular MD force fields are non-polarizable, so they do not account for the effects of electronic polarizability which are integrated into this method. Polarizable MD force fields tend to be computationally demanding due to the need to frequently recalculate charge distributions, so that correlations between electronic polarization and relatively

long time-scale nuclear motions are difficult to characterize at present.

On a practical level, this theory requires only a brief (1–2 ns) equilibrium simulation of the protein of interest. Calculation of the protein dielectric map can therefore be accomplished on a single workstation in 1 day, enabling a rapid analysis of several targets when needed. The low computational cost of this method is particularly important when studying large proteins or oligomeric protein complexes, for which longer-length MD simulations as a means of obtaining electrostatic energies may be impractically slow and a continuum electrostatics approach is therefore more appropriate. To increase the speed for large multiprotein simulations, the correlation matrices for individual proteins, or subsets of the system including protein interfaces, may be obtained in isolation and then appropriately combined to produce an overall dielectric map.

Salt bridge formation and disruption is known to play an important role in the misfolding of amyloid- β ,^{49–52} it is also instructive to investigate the significance of electrostatics in other misfolding-prone proteins. The theory developed here has been applied to the prion protein to elucidate the role that salt bridges and hydrophobic transfer energies may play in its misfolding.²⁰ Salt bridges known to be absent in disease-causing human mutants of the prion protein were found to be among the strongest present in the protein, so that the human mutants were electrostatically the least stable of those proteins studied. Conversely, the prion protein with the most stable salt bridges belonged to a species known to be resistant to prion disease (frog). It was also demonstrated that a Coulomb law with a single local effective dielectric constant was insufficient to fully capture salt bridge energetics, which necessitated the calculation of the full dielectric map to accurately predict the strength of salt bridges.

The utility of the dielectric calculator extends to any protein system in which electrostatics may play a role. Prominent examples include protein interactions with polyanions like DNA or RNA, protein–protein recognition and binding, oligomerization and aggregation, and membrane protein transport and selectivity. Furthermore, this approach is not limited to using water as a solvent, as solvent conditions in simulations may be tuned to reflect different environments where needed. We have at present only calculated dielectric profiles for natively folded ensembles, but the same technique could be applied to partially folded or misfolded structural ensembles.

In principle, screened interaction energies may be obtained directly from all-atom MD; however, the present theoretical framework may provide a clearer intuitive picture of why certain interactions may be strong or weak within the protein. While the present theory improves the quantitative description of protein dielectric response, we are currently limited in the application of our theory by the absence of a Poisson–Boltzmann equation solver capable of handling an anisotropic dielectric function; only an approximate isotropic (though heterogeneous) dielectric function can be used at present. We are currently developing a tool capable of accounting for such anisotropy to enable the accurate calculation of electrostatic energies and plan to apply it to the study of protein pK_a prediction, salt bridge energies, and protein thermodynamic stability.

Calculation of electrostatic energies by continuum electrostatics methods requires a description of the spatially-varying dielectric constant for the system under study. We have presented a robust tool to calculate this dielectric function in a protein–water system that accounts explicitly for the complex dynamic properties of protein and solvent dipoles. The method may be straightforwardly generalized to any biomolecule–solvent system. Heterogeneity and anisotropy are important characteristics of the protein dielectric, and strongly affect the electrostatic interactions that govern protein stability. Modulation of dielectric heterogeneity and anisotropy, through the evolution of residue fluctuations tailored to specific tasks, may provide a mechanism to simultaneously satisfy requirements for protein stability and function.

Acknowledgements

WCG acknowledges a Vanier Canada Graduate Scholarship from the Canadian Institutes for Health Research. NRC acknowledges funding from PrioNet Canada and a donation from William Lambert. SSP acknowledges funding from the A.P. Sloan Foundation, PrioNet Canada, the Natural Sciences and Engineering Research Council, and the Canada Research Chairs program. The authors are grateful for access to the WestGrid high-performance computing consortium, and we thank Andrey Karshikoff and George Sawatzky for helpful discussions.

References

- H. Lorentz, *The theory of electrons and its applications to the phenomena of light and radiant heat*, B.G. Teubner, Leipzig, 1916.
- J. G. Kirkwood, *J. Chem. Phys.*, 1939, **7**, 911–919.
- G. Oster and J. G. Kirkwood, *J. Chem. Phys.*, 1943, **11**, 175–178.
- N. F. Mott and M. J. Littleton, *Trans. Faraday Soc.*, 1938, **34**, 485–489.
- P. Debye, *Phys. Z.*, 1935, **36**, 103.
- L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486–1493.
- H. Nakamura, T. Sakamoto and A. Wada, *Protein Eng., Des. Sel.*, 1988, **2**, 177.
- J. W. Pitera, M. Falta and W. F. van Gunsteren, *Biophys. J.*, 2001, **80**, 2546.
- T. Simonson, D. Perahia and A. T. Brunger, *Biophys. J.*, 1991, **59**, 670–690.
- T. Simonson and D. Perahia, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 1082–1086.
- T. Simonson, *Curr. Opin. Struct. Biol.*, 2001, **11**, 243–252.
- A. Masunov and T. Lazaridis, *J. Am. Chem. Soc.*, 2003, **125**, 1722–1730.
- T. Simonson, J. Carlsson and D. A. Case, *J. Am. Chem. Soc.*, 2004, **126**, 4167–4180.
- D. Voges and A. Karshikoff, *J. Chem. Phys.*, 2000, **108**, 2219–2227.
- N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10037–10041.
- W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera and B. Honig, *J. Comput. Chem.*, 2002, **23**, 128–137.
- N. Okimoto, N. Futatsugi, H. Fujii, A. Suenaga, G. Morimoto, R. Yanai, Y. Ohno, T. Narumi and M. Tajiri, *PLoS Comput. Biol.*, 2009, **5**, e1000528.
- N. A. Baker, *Curr. Opin. Struct. Biol.*, 2005, **15**, 137–143.
- A. R. Fersht and M. J. Sternberg, *Protein Eng., Des. Sel.*, 1989, **2**, 527–530.
- W. C. Guest, N. R. Cashman and S. S. Plotkin, *Biochem. Cell Biol.*, 2010, **88**, 371–381.
- M. Kardar, *Statistical Physics of Fields*, Cambridge, New York, 2007.
- H. Lilliefors, *J. Am. Stat. Assoc.*, 1967, **62**, 399–402.
- J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkorshtid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
- B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- A. D. MacKerel Jr, B. Brooks, C. Brooks III, L. Nilsson, B. Roux, Y. Won and M. Karplus, in *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, ed. P. v. R. Schleyer *et al.*, John Wiley & Sons, Chichester, 1998, vol. 1, pp. 271–277.
- M. Mezard, G. Parisi and M. A. Virasoro, *Spin Glass Theory and Beyond*, World Scientific Press, Singapore, 1986.
- X. Song, *J. Chem. Phys.*, 2002, **116**, 9359–9363.
- U. Kaatzke, R. Behrends and R. Pottel, *J. Non-Cryst. Solids*, 2002, **305**, 19–28.
- S. Bandyopadhyay, S. Chakraborty and B. Bagchi, *J. Am. Chem. Soc.*, 2005, **127**, 16660–16667.
- A. Das and C. Mukhopadhyay, *J. Phys. Chem. B*, 2009, **113**, 12816–12824.
- R. C. Jones, *Phys. Rev.*, 1945, **68**, 93–96.
- E. Mele, *Am. J. Phys.*, 2001, **69**, 557–562.
- L. Serrano, A. Horovitz, B. Avron, M. Bycroft and A. R. Fersht, *Biochemistry*, 1990, **29**, 9343–9352.
- A. Horovitz and A. R. Fersht, *J. Mol. Biol.*, 1990, **214**, 613–617.
- H. A. Stern and S. E. Feller, *J. Chem. Phys.*, 2003, **118**, 3401–3412.
- H. Nymeyer and H.-X. Zhou, *Biophys. J.*, 2008, **94**, 1185–1193.
- Z. S. Hendsch, C. V. Sindelar and B. Tidor, *J. Phys. Chem. B*, 1998, **102**, 4404–4410.
- M. K. Gilson and B. H. Honig, *Biopolymers*, 1986, **25**, 2097–2119.
- G. N. Patargias, S. A. Harris and J. H. Harding, *J. Chem. Phys.*, 2010, **132**, 235103.
- J. Antosiewicz, J. A. McCammon and M. K. Gilson, *J. Mol. Biol.*, 1994, **238**, 415–436.
- H. Frohlich, *Proc. Natl. Acad. Sci. U. S. A.*, 1975, **72**, 4211–4215.
- A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons, New York, 1991.
- R. F. Grote and J. T. Hynes, *J. Chem. Phys.*, 1980, **73**, 2715–2732.
- W. H. Miller, *J. Chem. Phys.*, 1974, **61**, 1823–1834.
- E. Pollak, S. Tucker and B. J. Berne, *Phys. Rev. Lett.*, 1990, **65**, 1399–1402.
- S. S. Plotkin and P. G. Wolynes, *Phys. Rev. Lett.*, 1998, **80**, 5015–5018.
- S. S. Plotkin and J. N. Onuchic, *Q. Rev. Biophys.*, 2002, **35**, 205–286.
- The structures used were lads, laky, lakz, lamm, larb, lbfq, lcx, ldim, ledg, lhm, lmla, lorc, lphc, lptx, lrie, lrro, ltca, 2ayh, 2dri, 2end, and 3pte.
- B. Tarus, J. E. Straub and D. Thirumalai, *J. Am. Chem. Soc.*, 2006, **128**, 16159–16168.
- G. Reddy, J. E. Straub and D. Thirumalai, *J. Phys. Chem. B*, 2009, **113**, 1162–1172.
- J. A. Lemkul and D. R. Bevan, *J. Phys. Chem. B*, 2010, **114**, 1652–1660.
- J. Luo, J. D. Marechal, S. Warmlander, A. Graslund and A. Peralvarez-Marín, *PLoS Comput. Biol.*, 2010, **6**, e1000663.
- D. G. Isom, C. A. Castaneda, B. R. Cannon, P. D. Velu and B. Garcia-Moreno E., *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16096–16100.