Studying the Dielectric Properties of a Protein Solution by Computer Simulation

S. Boresch, P. Höchtl, and O. Steinhauser*

Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstrasse 17, A-1090 Vienna, Austria

Received: March 7, 2000; In Final Form: June 13, 2000

We report the static and frequency-dependent dielectric properties of a 9 mmol/L ubiquitin solution based on the analysis of a 5 ns molecular dynamics (MD) simulation. In accord with available experimental results, we obtain a significant dielectric increment for the dielectric constant (DC) at low frequencies (including the static DC ($\omega = 0$)), but a decrement at higher frequencies. The overall dielectric properties were decomposed into the protein−protein and the water−water self-terms, as well as the protein−water cross-term. The most significant contributions arise from the two self-terms, approximately 65% from water and 21% from the protein. These two components, corresponding to what experimentalists often refer to as β- and γ-processes, also determine the bimodal shape of the dielectric loss function ($\chi''(\omega)$). The quantitatively smallest protein−water cross-term (14%) corresponds to the experimentally observed δ-relaxation; it accentuates the bimodal shape of $\chi''(\omega)$ even further. A finer partitioning of the solvent into two solvation shells and bulk reveals the special role and properties of the first hydration layer surrounding the protein. Our findings point to protein−water interactions and, in particular, bound biological water as the microscopic origin of the δ-relaxation.

1. Introduction

Aqueous solutions of proteins exhibit a number of interesting properties (see, e.g., refs 1−3 and the references given therein): In contrast to solutions of many other polar compounds, their static (and low-frequency) dielectric constants (DC) are noticeably higher than that of water, whereas at higher frequencies they fall below that of the pure solvent. The imaginary part of the frequency-dependent DC is characterized by a pronounced bimodal distribution. In addition, there are some concentration-dependent effects (crossover in the real part of the frequency-dependent DC). Detailed analyses are complicated by the fact that the only measurable quantity is the DC of the solution; interpretations (have to) rely heavily on fits of the raw data to analytic functions, which are considered to represent the processes occurring on the molecular level. Phenomenologically, there is agreement about three principal contributions: (i) a very slow motion corresponding to the overall tumbling of the protein (often referred to as the β-process), (ii) the fast relaxation of unperturbed water (γ-process), and (iii) the so-called δ-process. This δ-process or δ-relaxation is thought to originate from the interactions between the protein and the solvent, as well as water molecules bound to (or interacting tightly with) the biomolecule, and it is at the center of most experimental and theoretical work.[2−9] Although several theoretical models rationalizing the dielectric properties of protein solutions have been proposed,[2,3] no molecular theory seems to be generally accepted. For example, the authors of two detailed experimental studies[4,5] differ in the interpretation of their results substantially from the mechanism proposed in ref 3.

Given the difficulties involved in the analysis of the experimental measurements, computer simulations should be ideally suited to enhance our understanding of the dielectric properties of aqueous solutions of biomolecules. A recently introduced theoretical framework permits, in principle, the rational explanation of the frequency-dependent DC of a solution as the sum of contributions from individual components (e.g., the protein, solvation shells, and "bulk" water).[10] An application of this approach to solutions of monosaccharides and disaccharides demonstrated that these systems are best understood as consisting of three components: the solute, a first solvation shell, and free bulk water.[11] However, close reexamination of the results of an analogous study of a zincfinger peptide[10] reveals an unexpected discrepancy: the value of 60.3 found for the static DC of the solution is slightly *lower* than that of the water model used; i.e., a dielectric decrement rather than an increment was obtained. With some restrictions (awkward treatment of the ions during the analysis, missing cross-term contributions), it is also possible to estimate solution DCs based on raw data reported by Yang and co-workers;[12,13] they also indicate a dielectric decrement rather than an increment. Thus, one faces somewhat of a quandary. Computer simulations coupled with the theoretical framework presented in ref 10 should provide a valuable tool to help understand and interpret the dielectric properties of biomolecular solutions. On the other hand, it is clearly unsatisfying that the few existing simulation studies gave qualitatively wrong results for the static DC (at least in comparison with the bulk of experimental evidence).

Evidently, this dilemma must be resolved before computer simulations can be relied upon to help interpret the dielectric properties of solvated biomolecules. The issue is also important in connection with several molecular dynamics (MD) simulations that were concerned with the static DC of proteins.[14−18] In these studies, the focus was on the solute (the protein or DNA molecule) rather than on the solution, so they are of little direct relevance in the present context. However, the significance of the results obtained is linked to the question of whether the experimental dielectric properties of macromolecular solutions can be reproduced. Providing this answer is, thus, the primary

* To whom correspondence should be addressed. Tel., +43-1-4277-52716; Fax, +43-1-4277-52790; E-mail, os@mdy.univie.ac.at.

10.1021/jp0008905 CCC: $19.00
Published on Web 08/19/2000

objective of this work. It should be understood as a pilot study investigating the usefulness of MD simulations in connection with the analysis of the dielectric relaxation of protein−water systems. Rather than extending or repeating the analysis of the zincfinger peptide, we chose a globular protein to avoid artifacts that might arise from too small a size of the solute or the presence of the Zn ion. Since the computational effort of simulating a well-solvated protein for several nanoseconds increases rapidly with the size of the biomolecule, we selected ubiquitin. It is unfortunate that there are no experimental data available for this protein, but our choice was limited by what was doable with the computational resources available to us while at the same time maintaining certain standards concerning the quality of the simulation (in particular, length of the simulation, treatment of electrostatic interactions, and size of the solvent box). Comparison to previous nanosecond simulations of ubiquitin in solution[19,20] can serve as a gauge to judge the quality of the trajectory. A further advantage of the system is that at neutral pH it contains the same number of positively and negatively charged amino acids; i.e., the simulation system is electrically neutral, and, consequently, there is no need to add counterions. This simplifies the analysis of the dielectric properties significantly.[10,21]

In the remainder of this paper we proceed as follows. We first outline the necessary theory to compute the dielectric properties of multicomponent systems from a MD simulation. Next, the technical details of the simulation and of the analyses are described. We report the frequency-dependent DC of the ubiquitin solution, as well as its decomposition into contributions from the protein self-term, the water self-term, and the protein−water cross-term. Since the last term is of particular interest in connection with the δ-process (cf. above), it was investigated further in a twofold manner. (i) Static self- and cross-term contributions to the overall static DC were computed for an even finer partitioning of the system into protein, two solvation shells, and bulk water. (ii) Application of the recently introduced dielectric self-consistent field (DSCF) method[22] permitted one to assess the degree of convergence for the cross-term achieved in the 5 ns simulation. The results are discussed from several angles. Since most analyses carried out are rather novel, close attention was paid at all stages to possible practical pitfalls and limitations of the methodologies employed. The insights concerning the molecular origin of the observed dielectric properties of the DC of the solution are summarized. Finally, we compare the present results with those obtained for the zincfinger peptide.[10]

2. Theory

In this section we summarize the theory required to compute the frequency-dependent dielectric constant of a solute−solvent system, as well as the contributions from individual components to the dielectric relaxation. Our goal is to make clear the essential elements of a computer simulation based analysis. The complete derivations and additional details can be found in the original work;[10,23,24] we also refer the reader to a detailed study of water−methanol mixtures.[25,26]

2.1. Calculation of the Frequency-Dependent Dielectric Susceptibility. The quantity of interest is the frequency-dependent dielectric constant $\epsilon(\omega)$. For multicomponent systems it is advantageous to formulate the theory in terms of the dielectric susceptibility $\chi(\omega)$, which is related to the DC by

$$4\pi\chi(\omega) = \epsilon(\omega) - 1 \tag{1}$$

We note that in the following the two terms will sometimes be used interchangeably; the connection is always given by eq 1. The susceptibility $\chi(\omega)$ is obtained from the time correlation function of the system dipole moment $\Phi(t) = \langle\vec{M}(t)\cdot\vec{M}(0)\rangle$ according to

$$\chi(\omega) = \frac{1}{3Vk_BT}\mathcal{L}\ [-\dot{\Phi}(t)] \tag{2}$$

where $\vec{M}$ is the dipole moment of the simulation system, V is its volume, T is the temperature, and k_B is the Boltzmann constant. The Fourier-Laplace transform $\mathcal{L}\ [f]$ of a function f is given by

$$\mathcal{L}\ [f] = \int_0^\infty \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t} f(t) \tag{3}$$

The notation $\dot{f}$ denotes the derivative of a function with respect to time, $\dot{f} = \mathrm{d}f/\mathrm{d}t$. Equation 2 is the expression appropriate for tinfoil dielectric boundary conditions,[10] which were used in this work. One sees from eq 3 that $\chi(\omega)$ is complex, i.e.

$$\chi(\omega) = \chi'(\omega) + i\chi''(\omega) \tag{4}$$

Utilizing a property of the Fourier−Laplace transform, $\mathcal{L}\ [-\dot{\Phi}] = \Phi(0) + i\omega\mathcal{L}\ [\Phi]$, one finds for the real ($\mathrm{Re}\{\chi(\omega)\} = \chi'(\omega)$) and imaginary part ($\mathrm{Im}\{\chi(\omega)\} = \chi''(\omega)$) of the susceptibility

$$\chi'(\omega) = \chi(\omega = 0) - (\omega/3Vk_BT)\mathrm{Im}\{\mathcal{L}\ [\Phi]\} \tag{5}$$

$$\chi''(\omega) = (\omega/3Vk_BT)\mathrm{Re}\{\mathcal{L}\ [\Phi]\} \tag{6}$$

Frequently, one refers to $\chi'(\omega)$ as the frequency-dependent susceptibility and to $\chi''(\omega)$ as the dielectric loss or the dielectric absorption.

In the present analysis the static susceptibility $\chi_0 = \chi(\omega = 0)$ is of special interest since for protein solutions it is characterized by a dielectric increment (cf. the Introduction). For $\omega = 0$, eq 2 reduces to

$$\chi_0 = \chi(\omega = 0) = \frac{\Phi(0)}{3Vk_BT} = \frac{\langle\vec{M}^2\rangle}{3Vk_BT} \tag{7}$$

introducing and defining χ_0. Formally, the expression for χ_0 should contain $\langle\vec{M}^2\rangle - \langle\vec{M}\rangle^2$ instead of just $\langle\vec{M}^2\rangle$.[27] For sufficiently long simulations, however, the time average of the dipole moment $\langle\vec{M}\rangle$ has decayed to zero and, therefore, can be omitted. This is true for molecular liquids[28] and for small solutes.[11] In simulations of protein systems, on the other hand, $\langle\vec{M}\rangle$ may still be far from zero. The large solute influences $\langle\vec{M}^2\rangle - \langle\vec{M}\rangle^2$ of the system in a twofold manner. First, there is a contribution from the fluctuation of the dipolar groups in the protein about their average position. Second, and much more important for the dielectric properties of the solution, the biomolecule has a strong net dipole moment $\vec{M}_P$, considerably larger than that of a solvent molecule. The rate at which $\langle\vec{M}_P\rangle$ decays and, thus, the rate at which $\langle\vec{M}^2\rangle - \langle\vec{M}\rangle^2$ converges are mainly determined by the rotational tumbling of the protein, a process occurring on the nanosecond time scale. If $\langle\vec{M}_P\rangle$ has not yet become zero (hence, if $\langle\vec{M}^2\rangle - \langle\vec{M}\rangle^2$ has not yet converged) since rotation of the solute has not yet been adequately sampled, it is necessary to estimate the limiting value for the variance of the dipole moment fluctuation from the available data. This is a practical problem that needs to be resolved on a case by case basis. The tumbling contribution, on the other hand, is of minor interest when one is interested in the internal DC of the biomolecule. In this case one can remove

the rotation of the protein prior to the analysis (cf., e.g., ref 14); then $\langle \vec{M}_P^2 \rangle - \langle \vec{M}_P \rangle^2$ is the quantity to use in eq 7 and its convergence should pose no particular problems.

2.2. Dielectric Component Analysis. Equations 2 and 7 describe the overall dielectric properties of a multicomponent system. As shown in ref 10, the total susceptibility can be expressed as the sum of component susceptibilities of all species present in the protein solution (ubiquitin and water). We note that an identical formalism has already been applied earlier to mixtures of molecular liquids.[25,26] For the ubiquitin system we, thus, have (cf. eq 2 for the second step)

$$\chi(\omega) = \chi_P(\omega) + \chi_W(\omega) = \frac{1}{3Vk_BT}\{L\ [-\dot{\Phi}_P] + L\ [-\dot{\Phi}_W]\} \tag{8}$$

where *P* and *W* denote protein and water, respectively. The two-component susceptibilities, $\chi_P(\omega)$ and $\chi_W(\omega)$, each consist of two terms, e.g., $\chi_P(\omega)$ is given by

$$\chi_P(\omega) = \chi_{PP}(\omega) + \chi_{PW}(\omega) = \frac{1}{3Vk_BT}\{L\ [-\dot{\Phi}_{PP}] + L\ [-\Phi_{PW}]\} \tag{9}$$

with an analogous expression for $\chi_W(\omega)$. In generalization of the notation introduced above, $\Phi_{PP}(t) = \langle \vec{M}_P(0)\cdot\vec{M}_P(t)\rangle$ is the autocorrelation function of the protein dipole moment, $\Phi_{PW}(t)$ the cross-correlation function between the dipole moment of the protein and the dipole moment of the water, etc. Equation 8 shows that the two-component susceptibilities, $\chi_P(\omega)$ and $\chi_W(\omega)$, are directly related to the two-component time correlation functions $\Phi_P(t)$ and $\Phi_W(t)$ of the protein solution. These are given by

$$\Phi_P(t) = \Phi_{PP}(t) + \Phi_{PW}(t) = \langle \vec{M}_P(0)\cdot\vec{M}_P(t)\rangle + \langle \vec{M}_P(0)\cdot\vec{M}_W(t)\rangle \tag{10}$$

$$\Phi_W(t) = \Phi_{WW}(t) + \Phi_{PW}(t) = \langle \vec{M}_W(0)\cdot\vec{M}_W(t)\rangle + \langle \vec{M}_P(0)\cdot\vec{M}_W(t)\rangle \tag{11}$$

making clear how one arrives at eq 9. Each component, P or W, consists of a self-term resulting from the respective component autocorrelation functions, $\Phi_{PP}(t)$ or $\Phi_{WW}(t)$, as well as an identical cross-term contribution, $\Phi_{PW}(t)$. While the former accounts for the dipolar correlations in the protein and in water, the latter adds the contributions from the interactions between P and W. Following ref 10, we sometimes refer to the subcomponents χ_{PP}, χ_{PW}, and χ_{WW} as pair susceptibilities. The generalization to more than two components (as is necessary, e.g., for the decomposition of the system into protein, solvation shells and bulk water) is straightforward (see eqs 62−65 of ref 10).

Two properties of the component and pair susceptibilities introduced in eqs 8 and 9 make them the key to an analysis of dielectric properties based on computer simulations. (i) Equations 10 and 11 show that, e.g, $\Phi_P(t)$ and, hence, $\chi_P(\omega)$ depend on both components present (i.e., P and W). Generalizing to more than two components, one sees that $\Phi_i(t)$ ($\chi_i(\omega)$) of the *i*th component is a function of all other components in the system.[10] In addition, (ii) in our approach the component and pair susceptibilities are uniformly scaled by the volume of the system (cf. eqs 8 and 9), i.e.

$$\chi_i \propto \frac{1}{V}\Phi_i \quad \text{and} \quad \chi_{ij} \propto \frac{1}{V}\Phi_{ij} \tag{12}$$

The theoretical justification of eq 12 is shown in detail in refs 10 and 21; other approaches exist[18] but do not seem fruitful in the present context.

These two properties make clear that a component susceptibility provides information about the magnitude of its respective contribution to the overall susceptibility of the protein solution.[10,21] This meaning of the term *component susceptibility*, used as a shorthand, should always be kept in mind. As has also been pointed out by Tom Simonson recently,[18] it is *not* to be confused with a spatially local dielectric constant or susceptibility.

In practice, the theoretical framework outlined above is applied in the following manner. Once a (sufficiently long) trajectory has been accumulated in a MD simulation, one evaluates the time correlation functions of eqs 10 and 11. The dielectric susceptibility of the solution, as well as the relative contributions of the components present in the system, are obtained directly without the need for involved fits, as is the case in the analysis of experimental data. If one suspects that, e.g., $\chi_W(\omega)$ is masking a superposition of effects, one can either look at its subcomponents ($\chi_{W-W}(\omega)$ and $\chi_{P-W}(\omega)$, cf. eq 9) or carry the partitioning of the system one step further. An example of the latter would be the distinction between waters in contact with the protein and bulk water. Obviously, everything that has been said here for the frequency-dependent component susceptibilities is also valid for the static case, $\omega = 0$.

2.3. Practical Calculation of the Dielectric Spectra. So far the decomposition of the susceptibility involved no approximations. However, the Fourier−Laplace transformation (cf. eq 2) of correlation functions derived from simulations can be problematic in practice, mainly because the correlation functions become very noisy at longer time intervals. One way to partially circumvent this difficulty is to fit the correlation function to appropriate target functions. For the correlation functions used in this work, a biexponential ansatz was sufficient, i.e.

$$\frac{4\pi}{3Vk_BT}\Phi(t) \approx \Phi_{fit}(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2} \tag{13}$$

Sums of exponentials as in eq 13 have the advantage that the Fourier−Laplace transform can be calculated analytically. We consider the possibility to transform back and forth between time and frequency domain a requirement for suitable fit functions. To make the fitting procedure more robust, we imposed the additional constraint

$$A_2 = \frac{4\pi}{3Vk_BT}\Phi(t=0) - A_1 = 4\pi\chi_0 - A_1 \tag{14}$$

The resulting three-parameter approach automatically normalizes the fitted correlation function $\Phi_{fit}(t)$ to $\chi_0 = \Phi(t=0)/(3Vk_BT)$, the quantity with the smallest error.

The Fourier−Laplace transform of Φ_{fit} can now be carried out analytically. Inserting eq 13 into eqs 5 and 6 one obtains

$$\chi'(\omega) = \chi(0) - \omega\left(\frac{A_1\tau_1^2\omega}{1+\tau_1^2\omega^2} + \frac{A_2\tau_2^2\omega}{1+\tau_2^2\omega^2}\right) \tag{15}$$

$$\chi''(\omega) = \omega\left(\frac{A_1\tau_1}{1+\tau_1^2\omega^2} + \frac{A_2\tau_2}{1+\tau_2^2\omega^2}\right) \tag{16}$$

where A_1, A_2, τ_1, and τ_2 are the parameters entering eq 13, subject to the condition of eq 14.

At this point is seems appropriate to clearly distinguish between the fitting required in the analysis of experimental data and the dielectric component analysis described here. As already mentioned in the Introduction, the only measurable quantity is the DC of the solution. Any further interpretation requires the extraction of the contributing processes. This can only be accomplished by fitting the data to a theoretical model; one commonly used ansatz is the sum of four Debye processes (i.e., exponential decays in the time domain). (Recently, measured spectra were also fitted by a distribution of relaxation times rather than a sum of Debye decay processes (see, e.g., ref 5). While such functions do model the measured data excellently, no corresponding representation in the time domain exists.) In the simulation-based approach, on the other hand, fits are solely needed to remove noise from the correlation functions. Since one has direct access to individual components, simpler functional forms (e.g., a biexponential (eq 13) rather than the tetraexponential ansatz just mentioned) suffice. *Additional information is not obtained from a more complicated target function, but from a more detailed decomposition.*

3. Methods

3.1. Details of the MD Simulation. Ubiquitin is a small, globular protein (76 amino acids, 1231 atoms, molecular weight of 8565 D) that plays a key role in the ATP-dependent protein degeneration pathway.[29] Because of its size and ease of availability, it is frequently used as a model compound, both in computer simulations[19,20] as well as in a variety of experimental applications.[30−33] Starting coordinates were taken from the crystal structure refined to 1.8 Å (PDB entry 1ubq).[34] All MD simulations were carried out with the SANDER module from the AMBER 4.1 suite of programs;[35] the Cornell et al. all-atom force field was employed.[36] Missing hydrogen positions were generated using standard bond lengths and bond angles. Assuming that the intrinsic pK's of the individual residues are unshifted by the protein environment, we chose the protonation state of the charged residues as it would be at pH 7. The glutamates and aspartates were negatively charged, and the lysines and arginines were assigned a charge of +1. The hydrogen of the single histidine residue was placed at the δ-position. Since the protein is electrically neutral under these conditions, no ions were added.

A cubic box of 5523 TIP3P water molecules[37] was generated around the protein with the EDIT program. Equilibration of the system was carried out as follows: First, the protein solution was heated to 300 K in four consecutive simulations of 1 ps length at 10, 150, 200, and 300 K with a time step of 0.1 fs. This heating phase was followed by 500 ps of Berendsen constant pressure (1 atm, $\tau_P = 1$ ps) and temperature (300 K, $\tau_T = 1$ ps) MD[38] with a time step of 1 fs. All further simulations were carried out at constant volume, using the average box size of the previous equilibration run ($56.2 \times 56.2 \times 56.2$ Å^3). At this volume the system corresponds to a protein solution of 0.0093 mol/L or 79.7 g/L. Equilibration was completed by additional 400 ps of simulation under the same conditions as the production run of 5 ns length (2 fs time step, $\tau_T = 10$ ps). In all simulations SHAKE[39,40] was used to constrain all bond lengths. Coordinates were saved to disk every 20 fs.

Electrostatic interactions were computed by the PME method[41,42] with a $60 \times 60 \times 60$ grid. The cutoff radius for the real space Ewald term and the van der Waals interactions was 10 Å. The implementation of the Ewald summation used in AMBER corresponds to the use of dielectric tinfoil boundary conditions.[23] The nonbonded pair list was updated at every time step to minimize energy loss during the simulation.

3.2. Details of the Analysis. *3.2.1. Frequency-Dependent Properties of the Solution—Contributions from Protein, Water, and Protein−Water Components.* The dielectric properties of the ubiquitin solution and of ubiquitin, the solvent, and the solvent−protein cross-term were computed from the trajectory saved during the 5 ns MD simulation. The static component susceptibilities $4\pi\chi_{0,W}$, $4\pi\chi_{0,P}$, and $4\pi\chi_{0,P-W}$ were calculated in generalization of eq 7 either from the self-averages $\langle\vec{M}_W^2\rangle$ and $\langle\vec{M}_P^2\rangle$ or from the cross-average $\langle\vec{M}_P\cdot\vec{M}_W\rangle$.

The time correlation functions Φ_{PP}, Φ_{PW}, and Φ_{WW} (cf. section 2.2) as well as the dipole moment correlation function of the whole system were fitted to a biexponential function with three independent parameters (see section 2.3, eqs 13 and 14) by the Levenberg−Marquardt algorithm.[43,44] Since the accuracy of the correlation functions becomes rapidly lower for longer time intervals, only the first 250 ps were used as input for the fits. The corresponding dielectric spectra were obtained by analytical Fourier−Laplace transform of the fitted functions according to eqs 15 and 16. A thorough error analysis of a nonlinear fit is quite involved. To judge at least the reliability of the fit itself, we proceeded as follows. Since we do not have error estimates for the individual data points in the correlation functions, we make the common assumption that the standard deviations for all points of the correlation function are the same ($\sigma_i = \sigma$). One can then calculate a standard deviation after the (nonlinear) least-squares fit has been carried out according to[43]

$$\sigma^2 = \sum_{i=1}^{N}[y_i - y(x_i)]^2/(N - M) \qquad (17)$$

Here the y_i are the data points of the original function and the $y(x_i)$ are the corresponding values obtained from the fit. N is the number of data points, and M is the number of independent parameters. In practice, we replaced $N - M$ by N, since we have only three parameters (cf. eq 14) and at least several thousand data points. The absolute magnitude of the correlation functions studied varied significantly, which made it difficult to compare the estimates obtained by eq 17 directly. Therefore, we report a normalized error E_{rms} related to σ by $E_{rms} = \sigma/y_1$. (E_{rms} is equivalent to σ that would be obtained from fits of the corresponding normalized correlation functions.) We note again that E_{rms} is only intended to reflect the reliability of the fit; a low value does not necessarily guarantee high accuracy of the parameters.

3.2.2. Further Partitioning Using the Voronoi Method. To better understand the role of the protein−water cross-term, we carried out a detailed spatial decomposition of the dielectric properties of the first and second water shell surrounding the solute. Using the technique of Voronoi polyhedra,[45] the first solvation shell (S1) was defined as all water molecules having at least one polyhedral face in common with the solute. The second solvation shell (S2) comprised all waters having one or more common surfaces with a S1 water molecule. All remaining waters were considered as bulk water (B). The Voronoi analysis was carried out with a program package by Gerstein et al.[46,47] To reduce the computational effort, only every fifth frame in the trajectory was subjected to the Voronoi analysis. Only static susceptibilities $4\pi\chi_{0,ij}$ were calculated, where i and j now stand for P, S1, S2, and B as just defined.

3.2.3. Calculations with the Dielectric Self-Consistent Field Method (DSCF). For reasons that will become clear in the Results section, we needed an independent assessment of the convergence of the protein−water cross-term. Recently, the dielectric self-consistent field method (DSCF), a tool to study

solvated systems based on continuum electrostatics, was introduced.[22] Since it permits a semiquantitative calculation of orientational correlation functions, it is well suited to investigate the P−W cross-term. In DSCF one solves the integral equation

$$\vec{P}(\vec{r}) = \frac{3(\epsilon(\vec{r}) - 1)}{4\pi(\epsilon(\vec{r}) + 2)}\{\vec{E}_0(\vec{r}) + \lim_{\sigma \to 0}\int_{|\vec{r}-\vec{r}|>\sigma} \mathrm{d}\vec{r}\,' \,\overleftrightarrow{T}_{\mathrm{EW}}(\vec{r} - \vec{r}\,')\vec{P}(\vec{r}\,')\} \quad (18)$$

for the polarization $\vec{P}$ under periodic boundary conditions. $\vec{E}_0(\vec{r})$ is the electric field in vacuum (calculated using Ewald summation) at point $\vec{r}$ due to the protein, modeled by the same force field as in the MD simulations. The dipole−dipole tensor $\overleftrightarrow{T}_{\mathrm{EW}}(\vec{r}) = \vec{\nabla}\vec{\nabla}\varphi_{\mathrm{EW}}(\vec{r})$ is the double gradient of the Green function for the Ewald summation with tinfoil dielectric boundary conditions, $\varphi_{\mathrm{EW}}(\vec{r})$. The DC $\epsilon(\vec{r})$ in the prefactor is a function of position. For regions within the solute it was set to one; outside the value for the static DC of TIP3P water (97) was used.[28] Having solved eq 18 numerically, thereby replacing $\vec{P}(\vec{r})$ by point dipoles $\vec{\mu}_i = \vec{\mu}(\vec{r}_i)$ located on equidistant mesh points $\vec{r}_i$, one can compute the distance dependent Kirkwood *g*-factor

$$g_{\mathrm{PW}}(r) = \frac{1}{|\vec{M}_{\mathrm{P}}^2|}\vec{M}_{\mathrm{P}} \cdot \sum_{\substack{i \\ |\vec{r}_{\mathrm{P}} - r_i| \le r}} \vec{\mu}_i \quad (19)$$

The quantity $g_{\mathrm{PW}}(r)$ describes the angular correlation of the dipole moment of the protein $\vec{M}_{\mathrm{P}}$ with all "solvent" dipoles $\vec{\mu}_i$ located in a sphere of radius *r* about $\vec{r}_{\mathrm{P}}$, the center of geometry of the protein. Additional details can be found in ref 22. The value of $g_{\mathrm{PS}}(r)$ for $r \to \sqrt{3}L/2$ obtained in this manner corresponds qualitatively to the static susceptibility $4\pi\chi_{0,\mathrm{P-W}}$ computed from the MD simulation. The distance-dependent Kirkwood *g*-factor was calculated for several orientations of the protein with respect to the space fixed axis of the coordinate system (the coordinates of the protein used in the DSCF calculations were those of the crystal structure). Results are reported for the following two geometric arrangements. In the first, the protein was aligned with the axes of the coordinate system, so that, viewing the protein as an ellipsoid, the axis of the longest elongation of the biomolecule coincided with the *x*-axis, the second longest one coincided with the *y*-axis, and the third one coincided with the *z*-axis. The second orientation was chosen so that the dipole moment of the protein, $\vec{M}_{\mathrm{P}}$, lay in the *x*-axis and had no components in the *y*- and *z*-directions. In both cases, the center of geometry of ubiquitin was located at the origin. These two arrangements are connected by a rotation of 117.62° about the axis of rotation (0.462, 0.819, 0.341). The numerical solution of the DSCF was obtained for a cubic box of 56.2 × 56.2 × 56.2 Å^3 (as in the MD simulation), using a 64 × 64 × 64 grid.

4. Results

4.1. RMSD Deviation. Figure 1 shows the root-mean-square deviation (rmsd) of the backbone atoms of ubiquitin as a function of simulation time. Residues 73−76 were not included in the analysis as their flexibility[34] might distort the results. The rmsd starts at 0.9 Å; this is the deviation from the X-ray structure after 900 ps of equilibration (cf. section 3.1). During the 5 ns of production, the rmsd oscillates about 1 Å and never rises beyond 1.3 Å, indicating a very well behaved system. The development of the rmsd as a function of simulation time depicted in Figure 1 is in accord with the findings by Fox and Kollman, who reported a 1.5 ns simulation of ubiquitin under very similar conditions.

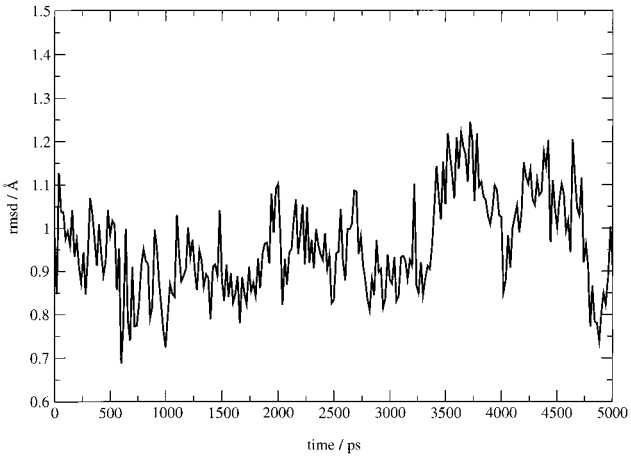

Figure 1. Rmsd of the heavy backbone atoms of residues 1-72 of ubiquitin as a function of simulation time. The highly mobile residues 73-76 (cf. ref 34) were omitted from the analysis.

4.2. Dielectric Properties of the Solution−Static Properties. In the following two sections we report the dielectric properties of the full solution, as well as the results of the decomposition into contributions from the solute (ubiquitin), the solvent (water), and the water−protein cross-term. We begin with the static results; the frequency-dependent results are presented separately in the next section.

When introducing the expression for the static dielectric constant in section 2.1, it was pointed out that the formally correct quantity to use in eq 7 is $\langle\vec{M}^2\rangle - \langle\vec{M}\rangle^2$ instead of $\langle\vec{M}^2\rangle$. For the water component $\langle M_{\mathrm{W}}\rangle^2$ has practically vanished after 500 ps of simulation (data not shown); thus, this distinction is not relevant. However, for the average dipole moment of the protein this is not the case. The cumulative averages of $\langle\vec{M}_{\mathrm{P}}^2\rangle$, $\langle\vec{M}_{\mathrm{P}}\rangle^2$, and $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$ as a function of simulation time are shown in Figure 2. The square of the average dipole moment ($\langle\vec{M}_{\mathrm{P}}\rangle^2$, solid line) gradually approaches zero, but even after 5 ns has a residual value of approximately 250 Å^2 e^2. This behavior was to be expected since the decay of $\langle\vec{M}_{\mathrm{P}}\rangle$ is determined by the overall tumbling of the molecule. The time scale for such processes is on the order of several nanoseconds—the experimental rate constant for ubiquitin is 4.1 ns.[30] One might, therefore, conclude that, clearly, in this case $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$ ought to be inserted in eq 7. However, a more careful look at Figure 2 indicates that this is not true. During the simulation time of 5 ns, $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$ (dashed line) has not fully converged. By contrast, $\langle M_{\mathrm{P}}^2\rangle$ (dotted line) has reached a plateau after approximately 1 ns, and its value remains practically unchanged after 3 ns. Almost all of the variation in $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$, therefore, results from the ongoing change of $\langle\vec{M}_{\mathrm{P}}\rangle^2$. If $\epsilon_{0,\mathrm{P}}$ or $\chi_{0,\mathrm{P}}$ were calculated with $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$, the values obtained would depend systematically on the length of the simulation; i.e., very different results would be obtained after 2, 3, or 5 ns. On the other hand, since $\langle\vec{M}_{\mathrm{P}}\rangle^2$ is clearly going to zero, $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$ will asymptotically approach the value of $\langle M_{\mathrm{P}}^2\rangle$. Thus, $\langle M_{\mathrm{P}}^2\rangle$ appears to be the best estimate for the limiting value of $\langle\vec{M}_{\mathrm{P}}^2\rangle - \langle\vec{M}_{\mathrm{P}}\rangle^2$. We note that this contrasts with what was done in most published work where the focus was on the internal dielectric constant of the biomolecular solute.[12−18] The behavior of the system dipole moment (protein plus water) is very similar to that of the protein (data not shown); we also used $\langle\vec{M}^2\rangle$ instead of $\langle\vec{M}^2\rangle - \langle\vec{M}\rangle^2$.

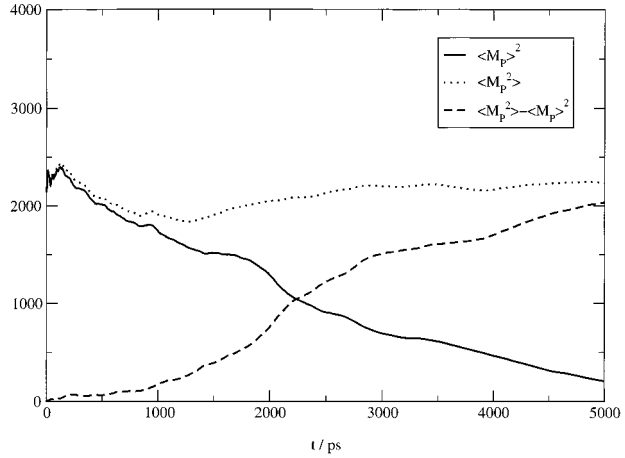

Figure 2. Cumulative averages $\langle \vec{M}_P^2 \rangle$, $\langle \vec{M}_P \rangle^2$, and $\langle \vec{M}_P^2 \rangle - \langle \vec{M}_P \rangle^2$ of ubiquitin in Å^2 e^2 as a function of simulation time; $\vec{M}_P$ is the dipole moment of the protein (ubiquitin).

TABLE 1: Static and Frequency-Dependent Dielectric Properties of a 0.0093 M Aqueous Ubiquitin Solution, as well as the Protein−Protein (P), Water−Water (W), and Protein−Water (P−W) Contributions[a]

contribution	$4\pi\chi_0 \pm \sigma$[b]	A_1	τ_1	A_2	τ_2	E_{rms}
P	29.4 ± 1.1	0.2	3.5	29.2	2568.0	0.0003
W	89.9 ± 3.3	87.0	6.8	2.9	553.1	0.016
P−W	9.6 ± 1.4	0.9	17.2	8.7	1907.9	0.013
total[a]	138.4 ± 5.4	87.9	6.9	50.5	1919.8	0.011

[a] The static susceptibility of the total solution is obtained as the sum $\chi_{0,P} + 2\chi_{0,P-W} + \chi_{0,W}$ (cf. eqs 8 and 9). The parameters A_1, A_2, τ_1, and τ_2 are defined in eqs 13 and 14. The definition of E_{rms} can be found in section 3.2.1. [b] Standard deviation of the mean calculated from twenty 250 ps subaverages taken from the full trajectory.

The results of the dielectric analysis of the ubiquitin solution based on a decomposition in protein (P), water (W), and protein−water (P−W) cross term are summarized in Table 1. It contains both static and frequency-dependent properties; we first concentrate on the static DCs or susceptibilities (the two quantities are related by eq 1). The second column of Table 1 lists the static component susceptibilities as well as the static susceptibility of the solution (last line). We reiterate that a component susceptibility is not a spatially local dielectric constant or susceptibility; instead, it provides information about the magnitude of its respective contribution to the overall susceptibility of the protein solution. All values are multiplied by 4π for easier comparison with the DC. We obtain a static susceptibility ($4\pi\chi_0$) of 138.4 for the solution; compared to the DC of the water model used (97),[28] this is a strong dielectric increment. Solute and solvent contribute to the overall DC/susceptibility as follows: Water makes the largest contribution (89.9); the protein adds 29.4, and the cross-term contributes two times 9.6 (cf. section 2.2). Qualitatively, these three components correspond to the experimentally observed γ-process (water relaxation), β-process (tumbling of the protein), and δ-process (protein−water interaction); further interpretation requires the frequency-dependent analysis. The relative magnitude of the three terms would, of course, change for different concentrations.

4.3. Dielectric Properties of the Solution−Frequency Dependent Dielectric Properties. As described in the Methods section, the time correlation functions of the total and component dipole moments could be fitted to biexponential functions. The fit parameters are also listed in Table 1. The preexponential factors A_1 and A_2 are scaled so that they add up to $4\pi\chi_0$ (cf. eq 14); thus, they directly reflect the weight of the two processes

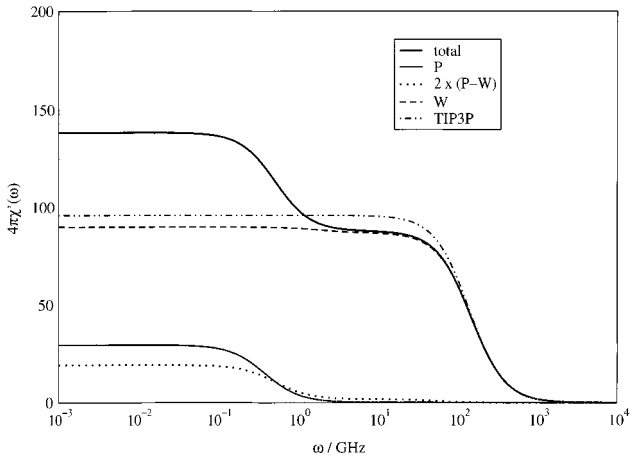

Figure 3. Frequency-dependent susceptibility $4\pi\chi'(\omega)$ of the components of the 0.0093 M aqueous ubiquitin solution: protein−protein (P) and water−water (W) self-terms, as well as two times the protein−water cross-term (P−W). In addition, $4\pi\chi'(\omega)$ of the solution as a whole (total) and of pure TIP3P water[28] is given.

described by the biexponential decay. The overall dipole moment of the solution (last row in Table 1) is characterized by two quite different time scales, a short one (6.9 ps), whose value is very similar to that of pure TIP3P water (7.3 ps),[28] and a much longer one (1920 ps). The respective weights are approximately two to one and can be rationalized by the static component susceptibilities. A_1 is very similar to $4\pi\chi_{0,W}$, and A_2 is close to the sum $4\pi\chi_{0,P} + 2 \times 4\pi\chi_{0,P-W}$. This justifies the interpretation of the fast process as the contribution from the solvent, and the slow process as the combination of contributions from the protein, from protein−solvent interactions, and, possibly, of waters bound to the biomolecule.

The component dipole moment time correlation functions provide a more detailed insight into the origin of the slow relaxation process. The water component is dominated by a fast decay corresponding to that of pure water; in addition, there is a small contribution from a much slower process (553 ps) not found for pure TIP3P water.[28] The water−protein cross-term consists of a fairly fast component of low, but nonnegligible weight ($A_{1,PW} = 0.9$, $\tau_{1,PW} = 17.2$ ps) and a dominant term with a relaxation time of 1908 ps. Finally, the protein has a relaxation time of 2568 ps, accompanied by a very fast process ($\tau_1 = 3.5$ ps) of extremely low weight ($A_1 = 0.2$). The protein relaxation time of 2.5 ns is comparable to the experimental estimates of the overall tumbling of ubiquitin (4.1 ns),[30] and it is also in line with the observed slow decay of $\langle M_P \rangle^2$ (see Figure 2 and the earlier discussion).

Upon Fourier−Laplace transformation of the (fitted) time correlation functions, one obtains the real and complex parts of the frequency-dependent (component) dielectric susceptibilities, which are displayed in Figures 3 and 4, respectively. Note that the plots are for the angular frequency $\omega = 2\pi\nu$ rather than the frequency ν. Figure 3 permits one to follow the changes of the DC of the solution, as well as that of the three components (water, protein, water−protein cross-term) as a function of ω. To facilitate comparisons, we also included $4\pi\chi'(\omega)$ of TIP3P water in the plot.[28] At low frequencies, one clearly sees the dielectric increment for the solution (bold solid line). While the water component (dashed line) is slightly lower than that of pure water (dash−dotted line), the protein contribution (solid line) and the cross-term contribution (dotted line) lead to the observed significant increment. At higher frequencies, however, the large protein cannot reorient itself fast enough, and its contribution to the DC vanishes. This corresponds to the large

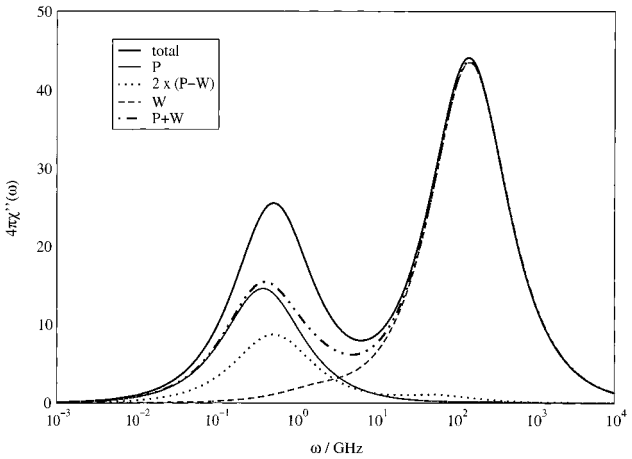

Figure 4. Frequency-dependent dielectric loss $4\pi\chi''(\omega)$ of the components of the 0.0093 M aqueous ubiquitin solution: protein−protein (P) and water−water (W) self-terms, as well as two times the protein−water cross-term (P−W). In addition, $4\pi\chi''(\omega)$ of the solution as a whole (total) and the sum of the protein and water self-term, i.e., the overall spectrum minus the protein−water cross-term, is given (P+W).

drop at approximately 0.5 GHz, which results from the disappearance of both the protein and the slow cross-term contribution. Above 1 GHz only the water term and the fast cross-term contribution remain; thus, we now have a dielectric decrement compared to water. Finally, at frequencies $>10^3$ GHz the water contribution, just as $4\pi\chi'(\omega)$ of pure water, vanishes.

The imaginary component (Figure 4) provides an alternative illustration of the same underlying processes. Two well-separated peaks can be discerned in the spectrum of the solution (bold solid line). From the component susceptibilities one sees that the large peak at higher frequencies is caused by water (dashed line), which also makes a small contribution in the frequency range between 0.1 and 5 GHz. The largest contribution to the peak at lower frequencies comes from the protein (solid line); the cross-term (dotted line) also adds to the overall result. To differentiate between the relative importance of the protein self-term and of the cross-term, we have also included the spectrum that one obtains if the cross-term is omitted (dash−dotted line). While there is a significant quantitative difference (in particular, the low-frequency peak is much smaller), the qualitative result remains unchanged. The spectrum, therefore, confirms the conclusion that could already be drawn from the static susceptibilities. The two major contributions to the dielectric constant of the protein solution are the protein self-term and the water component. The protein−water cross-term affects the absolute height, but not the bimodal shape of the spectrum.

Several of the most important results of this work have been presented in this and the previous section, and it seems appropriate to give a short summary at this point. Before doing so, a few remarks concerning the quality of the fits are in order. First, if one takes into account that correlation times ≥ 2 ns were deduced from a 5 ns simulation, the two long correlation times ($\tau_{2,PW} = 1908$ ps and $\tau_{2,P} = 2568$ ps) ought to be considered as reasonable estimates rather than as accurate and precise values, despite the low E_{rms} values reported in Table 1. As mentioned in section 3.2.1, low E_{rms} values indicate a satisfactory reproduction of the raw data by the fits, but do not guarantee that the parameters are necessarily reliable. Some clarification can, however, be obtained by repeating the fits using a different amount of raw data; we did this using the correlation functions up to 200, 300, and 500 ps as input (results not shown). There is overall little variation in the values reported in Table 1, with

TABLE 2: Static Pair Susceptibilities χ_{ij} and Component Susceptibilities χ_i of the Voronoi Decomposition of an Aqueous Ubiquitin Solution into Contributions from Protein (P), First (S1) and Second Solvation Shell (S2), and Bulk Water (B)[a]

	P	S1	S2	B	$4\pi\chi_{0,i}$[b]
P	29.4 ± 1.1	-2.8 ± 0.4	1.0 ± 0.3	11.4 ± 1.2	39.0
S1	-2.8 ± 0.4	3.9 ± 0.1	1.4 ± 0.1	0.8 ± 0.4	3.3
S2	1.0 ± 0.3	1.4 ± 0.1	5.5 ± 0.1	5.1 ± 0.5	13.0
B	11.4 ± 1.2	0.8 ± 0.4	5.1 ± 0.5	65.8 ± 2.1	83.1

[a] All susceptibilities are multiplied by 4π to facilitate comparison to component DCs. Standard deviations of the mean calculated from 20 250 ps subaverages taken from the full trajectory are given as error estimates. [b] The component susceptibilities are obtained as the sum of the respective row.

two exceptions. The short relaxation time of the protein ($A_{1,P} = 0.2$, $\tau_{1,P} = 3.5$ ps in Table 1) is not discernible in all fits (i.e., the protein is better fitted monoexponentially in these cases). The slow relaxation time of water ($A_{2,W} = 2.9$, $\tau_{2,W} = 553$ ps) is always present, but the values obtained for $\tau_{2,W}$ range from 200 ps to >2 ns. It is not too surprising that the reliability of a tiny component with a long relaxation time, superimposed on a dominant fast process, is low.

Thus, four relaxation processes, including $\tau_{2,PW} = 1.9$ ns and $\tau_{2,P} = 2.5$ ns mentioned above, are robust with respect to the details of the fit. This is an interesting finding because of the similarity to the tetraexponential fits frequently used in the analysis of experimental data. By looking at Table 1, the four terms can be unambiguously attributed to processes on the molecular level: a water contribution ($\tau_{1,W}$), a protein contribution ($\tau_{2,P}$), and two contributions originating from the correlation between solute and solvent ($\tau_{1,PW}$, $\tau_{2,PW}$). Figures 3 and 4 also are in good qualitative agreement with experimental results for the frequency-dependent dielectric constant; see, e.g., refs 4−7. Together with the albeit extremely high dielectric increment found for the static DC (this issue will be addressed in the next two subsections), the calculation, therefore, reproduces the main features of the dielectric properties of protein solutions found experimentally.

4.4. Additional Results from a Voronoi Analysis of the First Two Solvation Shells. As mentioned in the Introduction, the so-called δ-process, which corresponds to the protein−water contributions characterized by $\tau_{1,PW}$ and $\tau_{2,PW}$ (cf. section 4.3 and Table 1), is the focus of most experimental and theoretical work.[2−9] To obtain a deeper understanding of this term, a decomposition of the ubiquitin solution into protein (P), first solvation shell (S1), second solvation shell (S2), and bulk water (B) was carried out. The static pair and component susceptibilities obtained from the Voronoi analysis are listed in Table 2. The first four columns of numbers are the static pair susceptibilities $\chi_{0,ij}$, where i and j stand for P, S1, S2, and B, respectively. These entries should be read as a matrix; for example, the second number in the first row of the table (-2.8 ± 0.4) is the pair susceptibility $4\pi\chi_{0,P-S1}$, etc. The last column (labeled $\chi_{0,i}$) contains the four static component susceptibilities; each is the sum of the four values in the corresponding row. These four $\chi_{0,i}$ add up to the static DC of the solution (138.4), as they must.

The component results $\chi_{0,i}$ mainly reflect the varying weights of the respective terms. Since there are much fewer waters in S1 than in S2 (or B), it is not too surprising that $\chi_{0,S1} < \chi_{0,S2} \ll \chi_{0,B}$. By contrast, the pair susceptibilities $\chi_{0,ij}$ shed some light on the contributions to the overall protein−water cross-term, as well as on the differences between waters close to proteins and those in the bulk. The first notable result is the anticorrelation between P and S1 (-2.8 ± 0.4). The negative $\chi_{0,P-S1}$

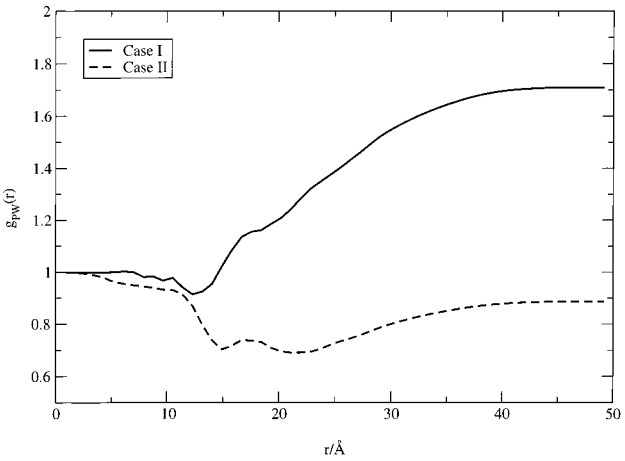

Figure 5. Distance-dependent Kirkwood *g*-factor $g_{PW}(r)$ (eq 19) for two orientations of the protein with respect to the simulation box. The center of geometry of ubiquitin lies in the origin of the coordinate system in both cases. Case I: the protein is aligned with the axes of the coordinate system, so that the axis of the longest elongation of the biomolecule coincides with the *x*-axis, the second longest one coincided with the *y*-axis, and the third one coincided with the *z*-axis. Case II: The protein is rotated so that its dipole moment $\vec{M}_P$ lies in the *x*-axis.

highlights the role of waters in the first solvation shell, which partially dampen (or shield) the large dipole moment of the ubiquitin solute. The low $\chi_{0,S1-B}$ cross-correlation shows the differences between waters in these two spatial regions. The second shell, on the other hand, has fairly low correlations with P (1.0 ± 0.3) and S1 (1.4 ± 0.1), but a relatively large correlation with B (5.1 ± 0.5). These values suggest that S2 acts as a transition zone between waters strongly interacting with the solute (S1) and free waters (B).

Since the negative $\chi_{0,P-S1}$ contribution is part of the overall P−W cross-term, other terms have to compensate; as one sees from Table 2 this is accomplished mainly by a large cross-correlation between P and B (11.4 ± 1.2). While the anticorrelation between P and S1 can be rationalized on physical grounds, the high value of $\chi_{0,P-B}$ is puzzling since one would expect a low value. Considering the details of the simulation system, the following explanation of this finding seems most logical to us. In this work (as well as in the vast majority of all simulations of solvated biomolecules), a single solute with a large dipole moment interacts with the waters in the system. To avoid artificial directing influences of this single solute, a simulation should ideally be sufficiently long as to populate all possible orientations of the solute with respect to the space-fixed axes of the coordinate system. It seems likely that very long calculations (certainly much longer than 5 ns) are required to achieve this goal. In nature this orientational diversity is realized by the presence of multiple, orientationally not equivalent solutes. The DSCF calculations reported next allow one to probe semiquantitatively the consequences of insufficient rotational sampling and to investigate whether it could be responsible for the high correlation between P and B.

4.5. Convergence Considerations Regarding the Protein−Water Cross-Term. If the correlation between protein and bulk were indeed too high, then the cross-susceptibility $\chi_{0,P-W}$ would be overestimated as well; this could provide an explanation for the sizeable dielectric increment found for the solution as a whole (+40% compared to the static DC of 97 of TIP3P).[28] As described in section 3.2.3, we used the DSCF method to compute $g_{PW}(r)$, which for $r \rightarrow \sqrt{3}L/2$ is proportional to $\chi_{0,P-W}$, for several orientations of the protein. The results for the two most interesting cases are depicted in Figure 5. In the first (case I), the solute was arranged according to geometric criteria; in the second (case II), the solute was rotated so that its dipole moment coincided with the *x*-axis (see section 3.2.3 for the details). In both curves one observes an anticorrelation for small *r*; this agrees with the negative $\chi_{0,P-S1}$ obtained for the P−S1 cross-susceptibility (cf. previous section). In case I, $g_{PW}(r)$ then rises gradually to 1.7, whereas in case II the *g*-factor remains below unity. The ratio of the final values of $g_{PW}(r \rightarrow \sqrt{3}L/2)$ is approximately 1.8. Figure 5, therefore, shows that the contributions to $\chi_{0,P-W}$ can vary significantly as a function of the orientation of the solute. While the large difference between cases I and II may not be quantitatively accurate, it strongly indicates that the value of $\chi_{0,P-W}$ suffers from insufficient sampling since it is unlikely that all relevant orientations of the solute in the simulation box have been sampled by the present 5 ns simulation.

The considerations just given concern the protein−water cross-contribution to the overall DC, $\chi_{0,P-W}$. They do not affect the other results of this work, in particular the protein and water self-terms. Considering that the relaxation time of the water component is 6.8 ps (cf. Table 1), a simulation time of 5 ns is certainly sufficient to describe $\chi_W(\omega)$ correctly. In addition, the solvent is also unlikely to introduce errors in the cross-term since its relaxation is much faster (by 3 orders of magnitude) than the tumbling motion of the protein; i.e., the waters have no difficulty following the much bigger solute quasi-instantaneously. To a good approximation the protein self-contribution (in contrast to the cross-term) does not depend on the orientation of the solute in the simulation box. The dipole moment of ubiquitin is mainly determined by its average structure, which varies little over 5 ns (cf. Figure 1); this is unlikely to change for longer simulation times. Although $\langle \vec{M}_P \rangle^2$ has not gone to zero as it ideally should (cf. Figure 2), the analysis given in section 4.2 shows that, nevertheless, reasonable values for $\chi_{0,P}$ can be derived. While it has been pointed out recently that the internal fluctuations of a solute could depend on orientation if the simulation box is not large enough,[48] any such effects will have only a small influence on the dielectric properties of the solution. Concerning the overall DC of the system it should further be kept in mind that the cross-term makes the smallest contribution (14% compared with 65% for the solvent and 21% for the protein). Returning finally to the P−W cross-term, the following can be concluded. While the available 5 ns trajectory is long enough to derive values for the time scales of the processes contributing to $\chi_{0,P-W}(\omega)$, i.e., $\tau_{1,PW}$ and $\tau_{2,PW}$, they are (probably by far) too short to accurately obtain the protein−water contribution to the solution DC owing to insufficient sampling.

5. Discussion

We have calculated the dielectric properties of an aqueous solution of the small, globular protein ubiquitin. We find a sizeable increment for the static DC (Table 1) and for the DC at low frequencies (Figure 3). At higher frequencies there is a sharp drop in $\chi'(\omega)$; i.e., for $\omega > 1$ GHz a dielectric decrement is obtained (Figure 3). The imaginary component $\chi''(\omega)$ is characterized by two clearly separated peaks (Figure 4). This is in good accord with the experimental findings and interpretations for the vast majority of protein solutions. For example, Dachwitz et al. studied the frequency-dependent dielectric properties of myoglobin solutions.[4] They fitted their data to a sum of five exponentials. One of these terms (which was also of very low weight) can be discarded from the comparison since it results from a very fast relaxation process of water not

reproduced by the TIP3P water model. The four remaining terms correspond well to the four principal components obtained in our analysis (cf. Table 1 and section 4.3). The two major contributions to the bimodal dielectric loss spectrum were attributed by Dachwitz et al. to the overall tumbling of myoglobin at low frequencies (β-process) and to the relaxation of the solvent (i.e., water) at high frequencies (γ-process). Similarly, the major contributions to the DC found in this study are the protein and water self-terms (χ_P and χ_W in Table 1, solid and dashed lines in Figures 3 and 4). The δ-relaxation was found to consist of two small contributions, which the authors interpreted as the result of hydration water (i.e., water interacting with the protein) and fluctuations of polar side chains of myoglobin.[4] We find two contributions to the protein−water cross-term, having rather different time scales ($\tau_{1,PW} = 17$ ps and $\tau_{2,PW} = 1.9$ ns; see Table 1). The Voronoi decomposition of the static DC showed that particularly S1 behaves differently from bulk; this is a clear and independent evidence for the importance of hydration water proposed by Dachwitz et al. in connection with the δ-process. In a more recent study by Miura et al.,[5] frequency-dependent data for several proteins were reported. Again, the main contributions were thought to arise from the protein and water (β and γ-process). Similar to the findings by Dachwitz et al. and the results presented here, the δ-relaxation, attributed to bound water, qualitatively modified the spectra but was not the dominating influence. The good qualitative agreement with experimental results for a number of solutions of biological macromolecules[4−7] proves that the primary goal of this study was accomplished. (Since only one simulation was carried out, the concentration dependence of the frequency-dependent DC could not be investigated.) It is unfortunate that no experimental data for ubiquitin exist, but as we pointed out in the Introduction, a simulation of a larger protein (such as myoglobin) was not practical.

From the results presented here, a number of conclusions concerning the molecular origin of the dielectric properties of the ubiquitin solution can be drawn. These can be immediately deduced from the raw data and do not rely on assumptions of theoretical models of any kind. The major contributions to the frequency-dependent dielectric susceptibility and dielectric loss (Figures 3 and 4) come from the protein−protein and water−water self-terms; the same holds true for the static susceptibility (cf. Table 1). The protein contribution is mainly responsible for the dielectric increment at low frequencies. Since this term vanishes at frequencies $\omega > 1$ GHz, only the water contribution remains, thus leading to the observed change from dielectric increment to decrement. The protein−water cross-term (although most likely too high, see section 4.5 and next paragraph) enhances this effect: it contributes mainly at low frequencies, but disappears at higher frequencies when the protein cannot reorient fast enough anymore. In addition, the more detailed spatial decomposition of the solvent into two solvation shells and bulk showed the very different properties of the first solvation shell and gave clear indications as to its physical role: The anticorrelation of the P−S1 cross-term dampens the large dipole moment of the protein. These insights are good examples of how simulation studies can complement experimental measurements in a fruitful manner.

Despite the impossibility of a direct comparison to experimental data for ubiquitin, the magnitude of the dielectric increment found is somewhat surprising. Comparing, for example, to the work of Miura et al.,[5] solutions of most proteins have an increment of the static DC between zero and 15%. This is much lower than the 40% found in this study (compared to the dielectric constant of the water model used). We note, however, that Oncley reported an even higher increment for horse serum γ-pseudo globulin.[1] One issue to keep in mind is the concentration of the protein in the simulation (80 g/L or approximately 9 mM). The dielectric measurements of myoglobin by Dachwitz et al. were carried out with a 60 g/L solution.[4] NMR studies of ubiquitin are carried out with 5−6 mM solutions (R. Konrat, personal communication). The artifically high concentration of the simulation could be responsible for the increment found in the present study; however, it seems unlikely that it can account for all of the observed increment. Until computational studies of more dilute solutions become feasible, and until there are experimental data to compare to, it is difficult to determine the origin of this discrepancy. One indication as to a possible source of the large increment came from the detailed analysis of the solute−solvent cross-term contribution by a decomposition of the system into protein, first and second solvation shell, and bulk water. It revealed an unexpectedly high correlation between ubiquitin and bulk water (section 4.4). Calculation of the protein−solvent cross g-factor with the recently introduced DSCF method[22] (section 4.5) showed that depending on the orientation of the protein in the simulation box very different protein−water cross-term contributions were obtained. These results indicate that the contribution to the DC from protein−water correlations has not converged in our 5 ns simulation. While the present calculation permitted one to estimate the time scale of the relaxation processes of the cross-term, a quantitatively accurate determination of its magnitude was not possible. We repeat that this type of sampling problem (at least to a good approximation) does not affect the protein and water self-terms. Unfortunately, better convergence of the P−W cross-term will not be easy to achieve. We found only marginal differences in the results between an analysis of 4 and 5 ns of MD (data not shown). More likely, an order of magnitude longer simulation times (i.e., >50 ns) may be required for better sampling.

Finally, it is necessary to compare the present results with those obtained for the zincfinger peptide solution[10,21] that prompted the present work. Upon examination of the respective data one finds several agreements. (When doing so, one should keep in mind that in the zincfinger study the SPC/E water model ($\epsilon_{0,SPC/E} = 68$) instead of TIP3P ($\epsilon_{0,TIP3P} = 97$) was used—the values given are those reported in ref 28). In both systems the dipole moment of S1 is anticorrelated to the protein as reflected by the negative $\chi_{0,P-S1}$, and there is a fairly strong correlation (possibly artificially so) between the protein and bulk (see Table 2 of this work and Table 2 of ref 10, listing the contributions to the static DC from P, S1, S2, and B, respectively). Further, both component DCs of the solvent itself are lower than pure water. We emphasize this accord since the two simulations were carried out with different programs and force fields. The main difference between the two systems is the magnitude of the protein self-contribution to the DC, 10.6 for the zincfinger peptide and 29.4 for ubiquitin. The ratio of almost three between these two values is in good agreement with the ratio of the square of the dipole moments of the two biomolecules (excluding the Zn^{+2} from the calculation, cf. refs 10 and 21), which is 3.1. The difference in dipole moments is mainly caused by the size of the two solutes; ubiquitin (76 residues) is significantly larger than the 18-residue peptide. While in the case of ubiquitin the solute more than compensates for the lowering in DC compared to pure water (cf. above), the smaller solute contribution in the zincfinger system is insufficient to lead to an overall dielectric increment. Thus, it seems reasonable to attribute the

dielectric decrement found for the solution of the zincfinger peptide to the specific nature of the solute, in particular, to the difference in size and, hence, the lower dipole moment compared to ubiquitin.

6. Conclusion

On the basis of the results presented, we believe that the usefulness of computer simulations for the analysis of the dielectric properties of biomolecular solutions has been demonstrated. It was shown that an MD-based approach can qualitatively reproduce measured dielectric properties of protein solutions and aid in the interpretation of the experimental data. Somewhat contradictory findings concerning the static DC of the solution obtained in an earlier simulation study of a zincfinger peptide could be rationalized. In addition, a critical assessment of the reliability of the results was given. While the protein and water self-contributions could be obtained with reasonable accuracy, some problems were detected for the protein−water cross-term. To address this issue, we have recently started to continue the MD simulations to improve the accuracy of the cross-term and to permit the investigation of the frequency-dependent dielectric properties of the solvation shells.

Acknowledgment. This work was funded by project P12537−CHE of the Austrian Fond zur Förderung der wissenschaftlichen Forschung.

References and Notes

(1) Oncley, J. L. *Proteins, Amino Acids and Peptides*. Cohn, E. J., Edsall, J. T., Eds.; Reinhold: New York, 1943; Chapter 22.
(2) Pethig, R. *Dielectric and Electronic Properties of Biological Materials*; John Wiley & Sons: Chichester, New York, Brisbane, Toronto, 1979.
(3) Nandi, N.; Bagchi, B. *J. Phys. Chem. A* **1998**, *102*, 8217−8221.
(4) Dachwitz, E.; Parak, F.; Stockhausen, M. *Ber. Bunsen-Ges. Phys. Chem.* **1989**, *93*, 1454−1458.
(5) Miura, N.; Nobuyuki, A.; Shinyashiki, N.; Mashimo, S. *Biopolymers* **1994**, *34*, 357−364.
(6) South, G. P.; Grant, E. H. *Proc. R. Soc. London Ser. A* **1972**, *328*, 371−387.
(7) Miura, N.; Hayashi, Y.; Shinyashiki, N.; Mashimo, S. *Biopolymers* **1995**, *36*, 9−16.
(8) Suzuki, M.; Shigematsu, J.; Kodama, T. *J. Phys. Chem*. **1996**, *100*, 7279−7282.
(9) Nandi, N.; Bagchi, B. *J. Phys. Chem. B* **1997**, *101*, 10954−10961.
(10) Boresch, S.; Ringhofer, S.; Höchtl, P.; Steinhauser, O. *Biophys. Chem*. **1999**, *78*, 43−68.
(11) Höchtl, P.; Boresch, S.; Steinhauser, O. *J. Chem. Phys*. **2000**, *112*, 9810−9821.
(12) Yang, L.; Weerasinghe, S.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **1995**, *69*, 1519−1527.
(13) Yang, L.; Valdeavella, C. V.; Blatt, D. H.; Pettitt, B. M. *Biophys. J*. **1996**, *71*, 3022−3029.
(14) Smith, P. E.; Brunne, M.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem*. **1993**, *97*, 2009−2014.
(15) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082−1086.
(16) Simonson, T.; Perahia, D. *J. Am. Chem. Soc*. **1995**, *117*, 7987−8000.
(17) Simonson, T.; Brooks,C. L., III. *J. Am. Chem. Soc*. **1996**, *118*, 8452−8458.
(18) Simonson, T. *J. Am. Chem. Soc*. **1998**, *120*, 4875−4876.
(19) Fox, T.; Kollman, P. A. *Proteins: Struct., Funct., Genet*. **1996**, *25*, 315−334.
(20) Abseher, R.; Lüdemann, S.; Schreiber, H.; Steinhauser, O. *J. Mol. Biol*. **1995**, *249*, 604−624.
(21) Löffler, G.; Schreiber, H.; Steinhauser, O. *J. Mol. Biol*. **1997**, *270*, 520.
(22) Boresch, S.; Steinhauser, O. *J. Chem. Phys*. **1999**, *111*, 8271.
(23) Neumann, M.; Steinhauser, O. *Chem. Phys. Lett*. **1983**, *102*, 508−513.
(24) Neumann, M.; Steinhauser, O. *Chem. Phys. Lett.* **1983**, *95*, 417−422.
(25) Skaf, M. S.; Ladanyi, B. M. *J. Chem. Phys*. **1995**, *102*, 6542−6551.
(26) Ladanyi, B. M.; Skaf, M. S. *J. Phys. Chem*. **1996**, *100*, 1368−1380.
(27) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(28) Höchtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. *J. Chem. Phys*. **1998**, *109*, 4927−4937.
(29) Hershko, A. *Biochem. Sci*. **1991**, *16*, 265.
(30) Schneider, D. M.; Dellwo, M. J.; Wand, A. J. Biochemistry **1992**, *31*, 3645−3652.
(31) Di Stefano, D. L.; Wand, A. J. *Biochemistry* **1992**, *26*, 7272.
(32) Yang, D.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc*. **1997**, *119*, 11938−11940.
(33) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc*. **1998**, *120*, 9870−9879.
(34) Vijay, S.; Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol*. **1987**, *194*, 531.
(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, S. W.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 4.1*; University of California: San Francisco, 1995.
(36) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K., Jr.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc*. **1995**, *117*, 5179−5197.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys*. **1983**, *79*, 926−935.
(38) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys*. **1984**, *81*, 3684−3690.
(39) van Gunsteren, W.; Berendsen, H. *Mol. Phys*. **1977**, *34*, 1311−1327.
(40) Rykaert, J.-P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys*. **1977**, *23*, 327−341.
(41) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(42) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys*. **1995**, *103*, 8577−8593.
(43) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: Cambridge, 1992.
(44) Moré, J. J.; Garbow, B. S.; Hillstrom, K. E. *User Guide for MINIPACK-1*; Argonne National Laboratory Report ANL-80-74; Argonne National Laboratory: Argonne, IL 1980.
(45) Voronoi, G. F. *J. Reine Angew. Math*. **1908**, *134*, 198−287.
(46) Gerstein, M.; Chothia, C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10167−10172.
(47) Gerstein, M.; Tsai, J.; Levitt, M. *J. Mol. Biol*. **1995**, *249*, 955−966.
(48) de Souza, O. N.; Ornstein, R. L. *Biophys. J*. **1997**, *72*, 2395−2397.