Dielectric Properties of Proteins from Simulation: The Effects of Solvent, Ligands, pH, and Temperature

ARTICLE *in* BIOPHYSICAL JOURNAL · JULY 2001

Impact Factor: 3.97 · DOI: 10.1016/S0006-3495(01)76226-1 · Source: PubMed

3 AUTHORS, INCLUDING:

Michael Falta
University of Otago
18 PUBLICATIONS **121** CITATIONS
SEE PROFILE

Wilfred F van Gunsteren
ETH Zurich
559 PUBLICATIONS **41,206** CITATIONS
SEE PROFILE

Dielectric Properties of Proteins from Simulation: The Effects of Solvent, Ligands, pH, and Temperature

Jed W. Pitera, Michael Falta, and Wilfred F. van Gunsteren
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

ABSTRACT We have used a standard Fröhlich-Kirkwood dipole moment fluctuation model to calculate the static dielectric permittivity, $\epsilon(0)$, for four different proteins, each of which was simulated under at least two different conditions of pH, temperature, solvation, or ligand binding. For the range of proteins and conditions studied, we calculate values for $\epsilon(0)$ between 15 and 40. Our results show, in agreement with prior work, that the behavior of charged residues is the primary determinant of the effective permittivity. Furthermore, only environmental changes that alter the properties of charged residues exert a significant effect on ϵ. In contrast, buried water molecules or ligands have little or no effect on protein dielectric properties.

INTRODUCTION

Proteins are exceedingly complex molecules. Defined as heteropolymers of amino acids, they contain a mixture of neutral, polar, and charged side chains. Although the dielectric properties of simple condensed phases, such as neat liquids, are well understood (Scaife, 1989), the influence of the heterogeneous protein environment on electrostatic interactions is still the subject of some debate. Because many biological molecules (DNA, for example) are charged at physiological pH, there is significant interest in understanding the precise details of electrostatic interactions in biological contexts (Honig and Nicholls, 1995). The fundamental constant defining the strength of the electrostatic interaction between two charges separated by a fixed distance is the relative static dielectric permittivity of the medium that separates them, $\epsilon(0)$. Though water has a relatively high dielectric permittivity ($\epsilon_{wat} = 78$ at 298 K), other components of the cell have permittivities approaching the $\epsilon = 2$ of hydrocarbon crystals (CRC, 2000). Initial estimates of the protein dielectric permittivity ranged from 2 to 80 (Ramachandran and Sasisekharan, 1968; Gilson and Honig, 1986; Nakamura et al., 1988; Svensson et al., 1990; King et al., 1991). In the past decade, numerous papers have reported values for $\epsilon(0)$ in the vicinity of a protein in solution as calculated from computer simulation (Simonson and Perahia, 1992; Smith et al., 1993; Simonson and Brooks, 1995; Simonson, 1998), and a consensus value between 10 and 35 has emerged. For regions deep within the protein, the effective $\epsilon(0)$ drops to something between 2 and 4. The primary determinant of this value is the behavior of side chains bearing a formal charge (Simonson and Perahia, 1992).

To understand the dielectric properties of proteins more completely, we have explored the influence of environmental effects (solvation, pH, ligand binding, and temperature) by calculating effective static dielectric constants $\epsilon(0)$ for four different proteins simulated under different conditions. With one exception, the molecular dynamics (MD) simulations used in our analysis were carried out by other researchers and are discussed in detail elsewhere. In order, the four proteins we have studied are hen egg white lysozyme (HEWL, 129 residues), simulated in water solution (Stocker and van Gunsteren, 2000), as a protein crystal (Stocker et al., 2000), and in chloroform solution (J. Pitera, unpublished results); α-lactalbumin (α-LAC, 123 residues), simulated at two different pH values (Smith et al., 1999); rat fatty acid binding protein (FABP, 131 residues), simulated as a complex with palmitate (*holo*-FABP) and as an apoprotein with water filling its large ligand binding pocket (*apo*-FABP) (Bakowies and van Gunsteren, submitted for publication); and a llama antibody heavy-chain variable domain (LLAMA, 115 residues), simulated at both 300 K and 340 K (Voordijk et al., 2000). Each of these proteins has been simulated for at least 1 ns in their respective environments. Fig. 1 shows the structure of each protein and emphasizes that two of the proteins (HEWL and α-LAC) are predominantly α-helical, whereas the other two (FABP and LLAMA) have mainly β-sheet secondary structure.

THEORY AND METHODS

Our analysis closely follows that outlined by Smith et al. (1993) and, like that work, makes use of the Fröhlich-Kirkwood theory (Kirkwood, 1939; Fröhlich, 1958; Neumann et al., 1984) of solute dielectric properties. To calculate the effective dielectric permittivity of a complex solute from computer simulation, it is necessary to map the properties of the solute observed in the simulation onto a simpler geometry, one amenable to analytical treatment. Specifically, the solute is approximated as a spherical cavity of volume V and permittivity ϵ embedded in a uniform dielectric continuum with permittivity ϵ_{RF}. The charge distribution of the solute is represented as a point charge and point dipole placed at the center of the spherical cavity. From the fluctuations of the solute dipole moment, $\mathbf{M}$, observed during a computer simulation, the temperature, T, the volume, V,

Received for publication 10 October 2000 and in final form 2 February 2001.

Address reprint requests to Dr. W. F. van Gunsteren, Physikalische Chemie, ETH Zürich, ETH Zentrum, CH-8092 Zürich, Switzerland. Tel.: 41-1-632-5501; Fax: 41-1-632-1039; E-mail: wfvgn@igc.phys.chem.ethz.ch.

0006-3495/01/06/2546/10 $2.00

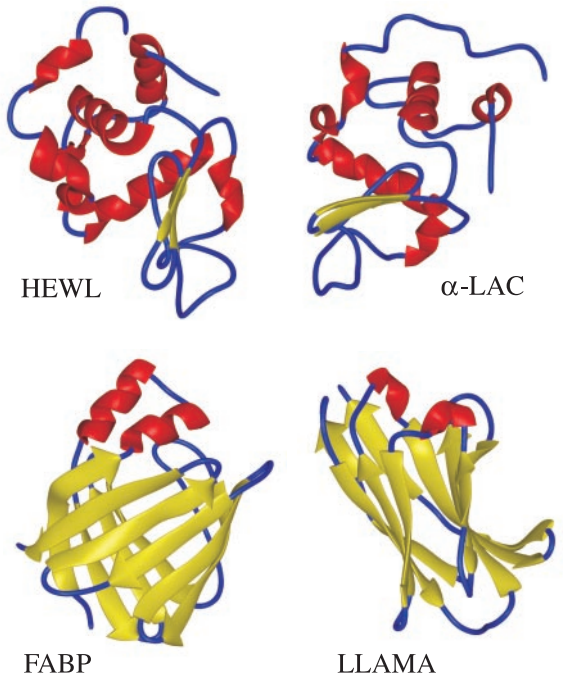

FIGURE 1 The four proteins studied in our analysis. In order, they are hen egg white lysozyme (HEWL; *upper left*), α-lactalbumin (α-LAC, *upper right*), rat fatty acid binding protein (FABP, *lower left*), and a llama antibody heavy-chain variable domain (LLAMA, *lower right*). Secondary structural elements are colored in red (α-helices), yellow (β-sheets), or blue (loops).

of the solute cavity, and the external dielectric, ϵ_{RF}, it is then possible to determine the permittivity ϵ inside the solute cavity.

In this Fröhlich-Kirkwood model, the dielectric permittivity of a system is specified as a function of the probability distribution of its total dipole moment, p(**M**), specifically by its second moment: the average fluctuation $\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2$. The total dipole moment **M** is simply:

$$\mathbf{M} = \sum_{i=1}^{N} q_i \mathbf{r}_i, \tag{1}$$

Where q is the partial charge of each atom of the system of interest and r is the distance from a fixed origin. For systems with a net charge, the value of **M** is dependent on the choice of the origin, so we have uniformly used the center of mass of the protein in our calculations. The fluctuations of this dipole moment are related to the local dielectric permittivity ϵ through:

$$\frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 V k_B T} = \frac{(2\varepsilon_{RF} + 1)(\varepsilon - 1)}{(2\varepsilon_{RF} + \varepsilon)}, \tag{2}$$

where ϵ_0 is the permittivity of vacuum, k_B is Boltzmann's constant, and SI units are used. It is straightforward to rearrange Eq. 2 to yield an expression for ϵ:

$$\varepsilon = \frac{1 + \dfrac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 V k_B T} \dfrac{2\varepsilon_{RF}}{(2\varepsilon_{RF} + 1)}}{1 - \dfrac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 V k_B T} \dfrac{1}{(2\varepsilon_{RF} + 1)}} \tag{3}$$

There are three methods (Heinz et al., 2001) that are generally applied to calculate the distribution of **M**, and from that the effective dielectric permittivity ϵ of a system: 1) umbrella sampling, 2) sampling of constraint forces, and 3) sampling of equilibrium fluctuations.

In the first two methods the free energy is calculated as a function of **M**, and from this p(**M**) and $\epsilon(0)$ are derived. Unfortunately, each determination of ϵ with these two methods requires a separate simulation or series of simulations. It is thus difficult to determine the contributions of various components (atoms) of the system to the permittivity. Consequently, we make use of the third method in this work, fluctuations of **M** observed from equilibrium simulation. After recording the trajectories from a single equilibrium MD simulation of the full system of interest, a number of different permittivities corresponding to the contribution of some or all of the atoms in the system can be calculated. Typically, extensive (>1 ns) simulation times are necessary to obtain converged values of the observed fluctuations and thus attain accurate estimates of $\epsilon(0)$.

The MD simulations of HEWL, α-LAC, FABP, and LLAMA that were analyzed in this work were all carried out previously by other authors, with the exception of the simulation of HEWL in chloroform. All simulations were carried out using the GROMOS96 united atom 43A1 force field (van Gunsteren et al., 1996) and the GROMOS96 biomolecular simulation program (Scott et al., 1999). The details of each simulated system are summarized in Table 1, and interested readers are referred to the relevant original papers for further information. In contrast to the simulations analyzed in the prior work of Smith et al. (1993), a reaction-field force

TABLE 1 Details of simulated systems

Protein	Solvent	System specification	Total protein mass (g/mol)	Number of amino acids	Number of protein atoms	Protein charge (*e*)	Periodic box boundaries	Box dimensions (nm)	Number of solvent molecules	Temperature (K)	Simulation time analyzed (ns)
HEWL	$CHCL_3$	Solution	14,313	129	1322	+9	t.o.	8.46	2159	300	0–5
	Water	Solution	14,313	129	1322	+9	t.o.	7.74	7113	300	0.5–2
	Water	Crystal (4 chains)	14,313	129	1322	+9	Rect.	5.9 × 6.9 × 3.1	1715	300	0.5–2
α-LAC	Water	Low pH (2.0)	14,095	123	1267	+16	t.o.	7.21	5582	300	0.3–2
	Water	High pH (8.0)	14,071	123	1244	−7.74* (−8)	t.o.	7.20	5574	300	0.3–1.5
FABP	Water	*apo* (25 waters)	14,273	131	1376	0	t.o.	6.75	4506	300	0–5
	Water	*holo* (palmitate)	14,273	131	1376	0 (−1)†	t.o.	6.75	4506	300	0–5
LLAMA	Water	Room temperature (300 K)	12,251	115	1134	+3	t.o.	6.97	5042	300	0–3
	Water	High temperature (340 K)	12,251	115	1134	+3	t.o.	6.96	5042	340	0–2

*Incorrect C-terminal changes were used in the simulation. All analysis used the corrected charges of −8.
†The palmitate ligand has a charge of −1.
t.o., truncated octahedron; Rect., rectangular prism.

(Tironi et al., 1995) was used in each simulation to approximate the effects of Coulomb interactions outside of the twin-range cutoff sphere. For the simulations with water as a solvent, a value of $\epsilon_{RF} = 54$ was used, whereas $\epsilon_{RF} = 5$ for the chloroform simulation. Also, the simulations analyzed in that prior work made use of the earlier GROMOS87 force field (van Gunsteren and Berendsen, 1987). Although different numbers of counterions and solvent simple point charge (SPC) (Berendsen et al., 1981) waters were explicitly included in each MD simulation, they were not included in our analysis.

For the simulation of HEWL in chloroform solution, the system was identical to the water solution of HEWL simulated by Stocker and van Gunsteren (2000) except that the solvent SPC water was replaced by rigid $CHCl_3$ molecules. Although the protein is not necessarily expected to maintain its normal folded structure in the nonpolar chloroform, we wanted to test the effect of immersion in a low-dielectric solvent on the protein's dielectric properties. It should be noted that all ionizable protein side chains were simulated in the same protonation states as the original water simulations, rather than the neutral states that might be expected in a low-dielectric medium. The system of one HEWL protein chain, nine Cl^- counterions, and 2159 chloroform molecules was equilibrated and simulated using a protocol identical to that of Stocker and van Gunsteren (2000). The only deviations from this protocol were the use of a reaction field permittivity of 5 and a compressibility of 8.816×10^{-4} nm^3 (kJ/mol)$^{-1}$ to reflect the chloroform solvent (Tironi and van Gunsteren, 1994). The final production simulation consisted of 5 ns of MD at 300 K and 1 atmosphere pressure. Visual inspection showed that the overall protein structure was well preserved, even after 5 ns of simulation. Protein all-atom and C_α atom-positional root-mean-square (RMS) deviations from the starting structure were roughly stable over the entire simulation at 0.3 and 0.2 nm, respectively.

To calculate the effective static dielectric permittivity, $\epsilon(0)$, from the equilibrium simulations, we have exactly followed the analysis of Smith et al. (1993) Specifically, the initial configuration of each MD simulation was selected as a reference configuration, defining the origin and axes of the coordinate system. Each subsequent protein coordinate from the MD trajectory was superimposed on the reference coordinate by translation of the center of mass and a least-squares fit of the positions of the C_α atoms of each residue. Least-squares superpositions using all atoms of the protein were not found to change the results significantly. Following the superimposition, the dipole moment for the selected atoms of each configuration was calculated.

To eliminate the influence of the choice of coordinate axes on the calculated fluctuation, the total dipole moment fluctuation ($\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2$) was calculated from the components of $\mathbf{M}$ along each axis:

$$\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2 = \langle \mathbf{M}_x^2 \rangle - \langle \mathbf{M}_x \rangle^2 + \langle \mathbf{M}_y^2 \rangle - \langle \mathbf{M}_y \rangle^2 + \langle \mathbf{M}_z^2 \rangle - \langle \mathbf{M}_z \rangle^2 \quad (4)$$

Each protein was decomposed into three different groups of atoms for the analysis. First, the contributions from all atoms of the protein to the dipole moment fluctuation were analyzed, yielding an overall dielectric permittivity $\epsilon(0)$. Next, the contributions from charged residues and the charged N and C termini were excluded from the analysis, yielding the effective permittivity in the absence of formal charges, $\epsilon(0)_{protein,e.c.r.}$. Finally, the dipole moment and its fluctuation were calculated for just the atoms of the protein backbone (H, N, C_α, C, and O) to yield $\epsilon(0)_{backbone}$. The atoms that were included in each of these three calculations are graphically displayed for each protein in Fig. 2. In the case of FABP, where we wished to evaluate the effects of buried ligands or water molecules on $\epsilon(0)$, a further distinction was made. From the trajectories of *holo*-FABP, a complex with a palmitate ligand, and of the *apo*-FABP, a complex with 25 buried water molecules, we carried out analyses in which the buried species (palmitate or water) were either included with or excluded from the dipole moment calculation. The different contents of the FABP binding cavity are depicted in Fig. 3.

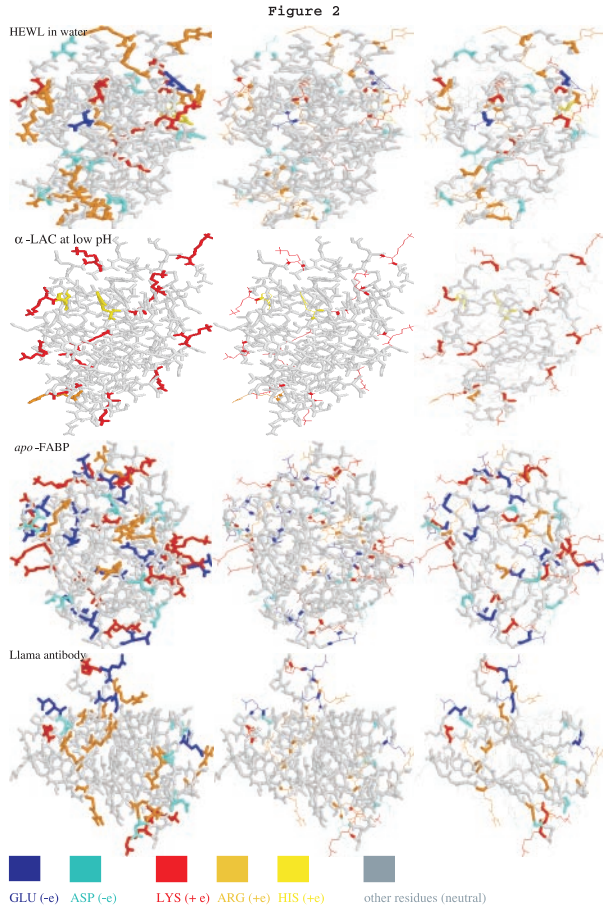

FIGURE 2 Schematic depiction of the atoms included (*thick lines*) or excluded (*thin lines*) in the different calculations of the dielectric permittivity $\epsilon(0)$ for the four proteins studied. Charged residues are shown in shades of red (LYS, ARG, and HIS) or blue (GLU and ASP). Other (neutral) residues are displayed in gray. The left column shows the entire protein, as used in the calculation of $\epsilon(0)_{protein}$. The middle column shows the protein with termini and the side chains of charged residues excluded ($\epsilon(0)_{protein,e.c.r.}$), while the right column depicts the protein backbone by itself ($\epsilon(0)_{backbone}$). In order, the four proteins depicted are HEWL, α-LAC, FABP, and LLAMA.

The volume term (V) in Eq. 3 was calculated from the total mass of each protein molecule and a typical partial specific volume for proteins of 0.72 cm^3/g (Creighton, 1984). The volume calculated in this manner was

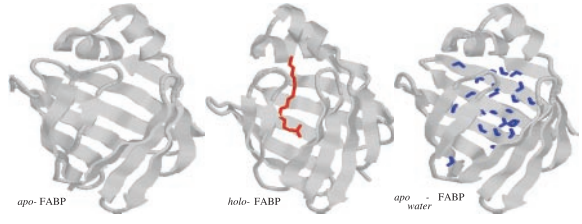

FIGURE 3 Schematic depiction of the rat fatty acid binding protein (*apo*-FABP), alone (*left*), complexed with palmitate (*holo*-FABP), and with 25 water molecules occupying the binding cavity (apo_{water}-FABP). The protein is drawn in gray, and the molecules occupying the binding cavity are in red (palmitate) or blue (water).

compared with the volume enclosed by the molecular surface of the reference configuration of each protein, as calculated by the program MSMS (Sanner et al., 1996) using a probe atom diameter of 1.5 nm. The two volumes determined for each protein are presented in Table 2. All subsequent analysis used the volume derived from the specific volume, column 1 of Table 2.

The final undetermined quantity in Eq. 3 is the dielectric permittivity ϵ_{RF} of the continuum surrounding the protein,. Again following the work of Smith et al. (1993), we used a value of $\epsilon_{RF} = 68$ for all simulations carried out in SPC water. A value of $\epsilon_{RF} = 5$ was used for the simulation of HEWL in chloroform, based on the data of Tironi and van Gunsteren (1994). Sensitivity analysis showed that a 25% decrease in the value of ϵ_{RF} used for the SPC water calculations shifts the calculated permittivities by only 4–7%, but does not change any of our conclusions. The influence of temperature or pH on the solvent dielectric permittivity was not taken into account in our analyses. In addition, the influence of the low-volume fraction of water in the HEWL crystal on the dielectric permittivity was not considered, either in the original simulation (Stocker and van Gunsteren, 2000) or in our analyses. A volume-weighted average of the protein and solvent permittivities for ϵ_{RF} is inappropriate because the dielectric permittivity appears in the denominator of the expression for the electric energy. Although a geometric mean of $\epsilon_{solvent}$ and $\epsilon_{protein}$ might serve, due to the complex nature of dielectric response and the highly heterogeneous composition of protein crystals, any single permittivity chosen as a uniform ϵ_{RF} is probably incorrect. Taking the solvent permittivity for ϵ_{RF} is a conservative choice that prevents protein-protein interactions from being artificially exaggerated by the reaction field correction.

RESULTS AND DISCUSSION

Fig. 4 shows the dipole moment fluctuations per unit volume $(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 for the simulation of HEWL in water. Clearly, the fluctuations observed when the entire protein is included in the analysis are much larger than for either the protein without charged residues and termini or the protein backbone alone. The latter two are re-displayed at a different vertical scale in Fig. 4 *b*. For HEWL in water, we find a value of $\epsilon(0)_{protein}$ of 25.7, in contrast to the $\epsilon(0)_{protein,e.c.r.}$ and $\epsilon(0)_{backbone}$ values of roughly 2.6 and 1.9, respectively.

Because our simulation of HEWL in chloroform was one of the most extensive data sets available to us, we used these coordinates to evaluate the convergence of the calculated dipole moment fluctuations as a function of the sampling time. The results of this analysis are shown in Fig. 5. Each panel shows a number of curves that correspond to starting the analysis at different time points along the simulation trajectory. The dotted curves correspond to analysis begun after the first 100 ps of simulation and were used for all calculations of $\epsilon(0)$ for this system. From these graphs, it is clear that multi-nanosecond simulation times are required to accurately determine the value of the dipole moment fluctuations. The step-like jump seen in each graph at ~3.5 ns appears to correspond to the movement of a six-residue loop (44ASN-50SER) of the protein. Coincident with this shift, increased movements are also seen in the C terminus. We have found no clear structural reason for the sharp rise in each curve at the end of the 5-ns simulation period.

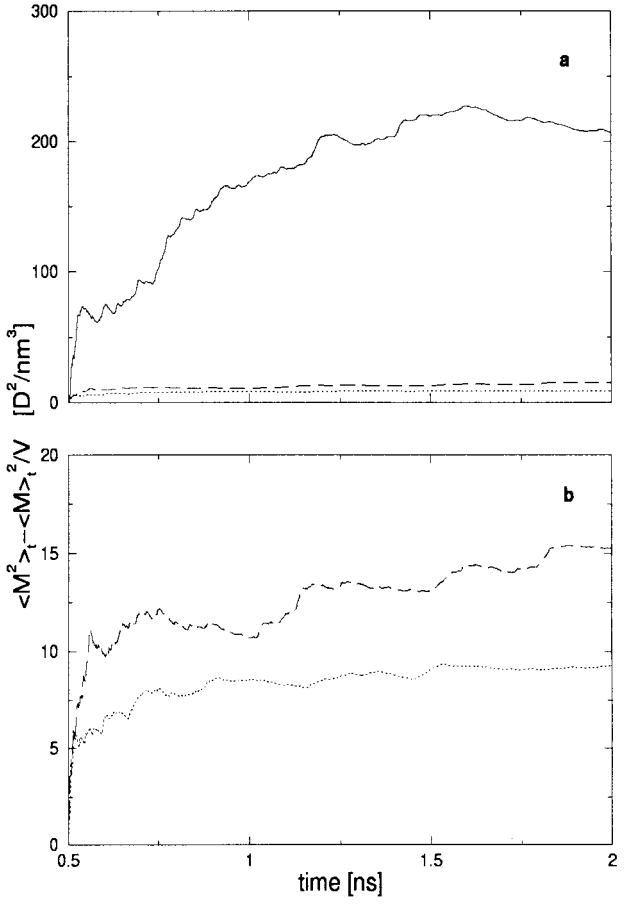

FIGURE 4 Total dipole moment fluctuation per unit volume $(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time for the simulation of HEWL in water, depicting the relative magnitude of dipole moment fluctuations calculated for the entire protein (——), the protein without charged residues or termini (— — —), and the protein backbone only (· · ·). (*a*) All three curves, emphasizing the greater magnitude of fluctuations calculated for the entire protein; (*b*) The lower two curves on an expanded vertical scale.

Our third HEWL simulation is a crystal of four chains of the HEWL protein, corresponding to the crystallographic unit cell. As with all protein crystals, there is a large amount of water present, which was included in the simulation but excluded from our dipole moment analysis. The four separate protein chains permit a nice estimate of the uncertainty in our calculated $\epsilon(0)$ values. By treating them as separate samples, it is possible to estimate a mean and standard

TABLE 2 Volumes determined for each protein

Protein	V (nm^3) based on a partial specific volume of 0.72 cm^3/g	V (nm^3) enclosed by the molecular surface
HEWL	17.113	16.734
α-LAC*	16.914 (16.885)	16.483 (16.483)
FABP	17.126	18.529
LLAMA	14.701	14.334

*at pH 2.0. Values in parentheses are for pH 8.0.

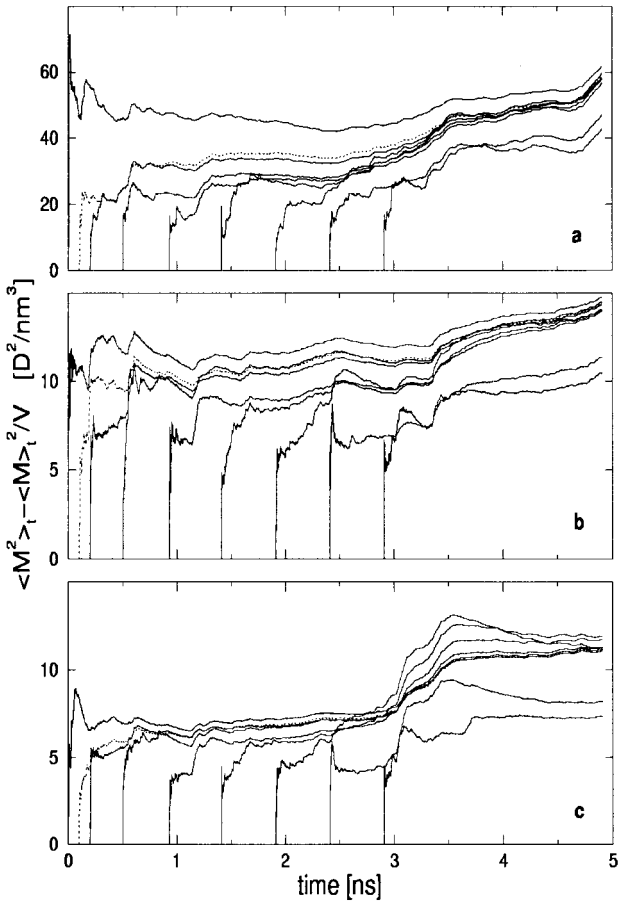

FIGURE 5 Total dipole moment fluctuation per unit volume ($\langle \mathbf{M^2} \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time for the simulation of HEWL in chloroform, based on analysis starting from different points in the simulation. (*a*) Total dipole moment fluctuation per unit volume calculated over all protein atoms; (*b*) Total dipole moment fluctuation per unit volume calculated for all protein atoms except for charged residues and the charged N and C termini; (*c*) Restricts the calculation further to depict only the contribution of the protein backbone. The dotted line corresponds to analysis begun after the initial 100 ps of MD simulation and corresponds to the results for this system discussed in the remainder of the text.

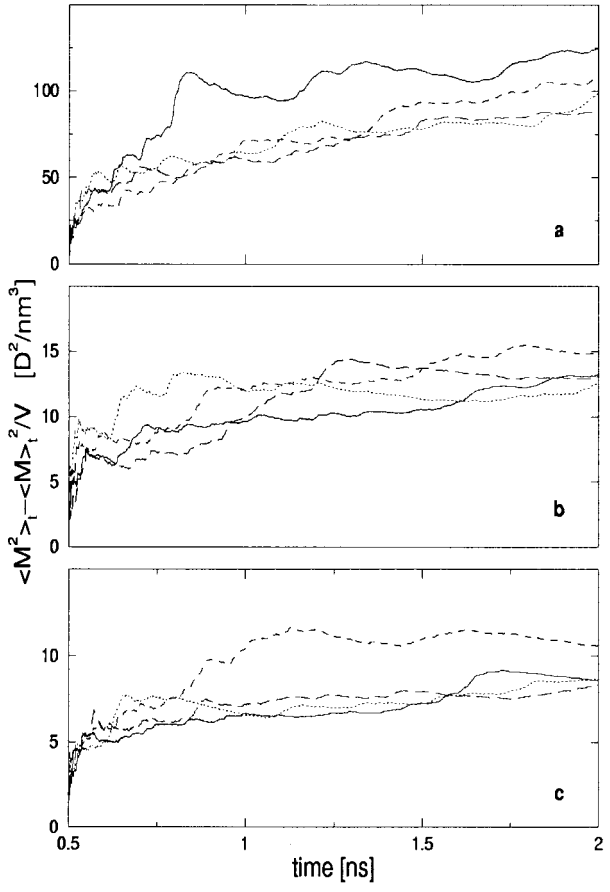

FIGURE 6 Total dipole moment fluctuation per unit volume ($\langle \mathbf{M^2} \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time for the simulation of four HEWL chains in a crystal unit cell. Results for each chain are graphed separately as solid lines (chain 1), dotted lines (chain 2), short-dashed lines (chain 3), and long-dashed lines (chain 4). (*a*) Total dipole moment fluctuation per unit volume calculated over all protein atoms; (*b*) Total dipole moment fluctuation per unit volume calculated for all protein atoms except for charged residues and the charged N and C termini; (*c*) restricts the calculation further to include only the contribution of the protein backbone. Because there are four independent curves for each graph, the variation between curves can be interpreted as an indication of the minimal uncertainty in the values of $(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2)/V$.

deviation for each $\epsilon(0)$ value, admittedly with a very small number of samples. The results for each chain are graphed as separate curves in Fig. 6. We estimate an average value for $\epsilon(0)_{\text{protein}}$ of 12.5, with a standard deviation of 1.6. Again, the values for the protein excluding formally charged groups and for the protein backbone are much lower, 2.3 and 1.9, with corresponding standard deviations of 0.17 and 0.11. It is clear from these results that although the dipole moment fluctuations may appear to have an uncertainty of several hundreds of Debye^2, the corresponding variations in ϵ can be relatively small.

The results for all three HEWL simulations are compared in Fig. 7. From these results, it is clear that the highest fluctuations for the dipole of the entire protein are seen for the protein in water, and these are progressively more damped upon going to the hydrated crystal and the chloroform solution. In contrast, the dipole moment fluctuations when charged residues are excluded and the dipole moment fluctuations for the protein backbone are roughly similar in all three environments, where they vary by less than 5 Debye^2/nm^3.

The significance of contributions from charged residues is also evident in Fig. 8, which shows the results from the simulations of α-LAC at low pH (2.0) and high pH (8.0). In the upper graph (Fig. 8 *a*), which depicts the total dipole moment fluctuations per unit volume for the whole protein, higher fluctuations are quite clearly seen for the high-pH simulation. The volume-normalized fluctuations in this case

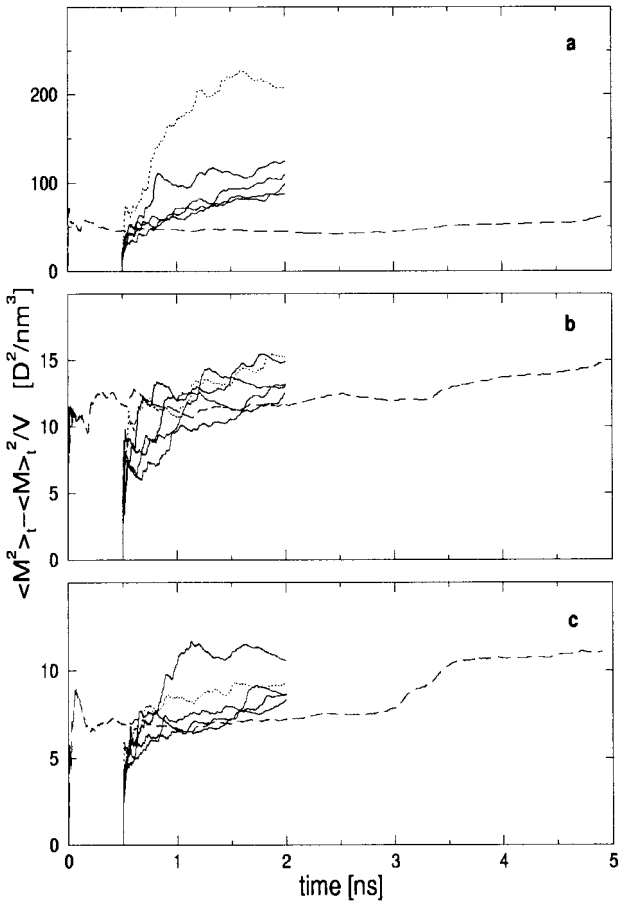

FIGURE 7 Summary of total dipole moment fluctuations per unit volume calculated for HEWL in three different environments. Results are shown for HEWL simulated in water solution (···), in a crystal (——, four molecules), and in chloroform solution (— — —). (*a*) Total dipole moment fluctuation per unit volume $(\langle \mathbf{M^2} \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time, calculated over the entire protein; (*b* and *c*) Results for the protein excluding charged residues and termini and for the protein backbone only, respectively.

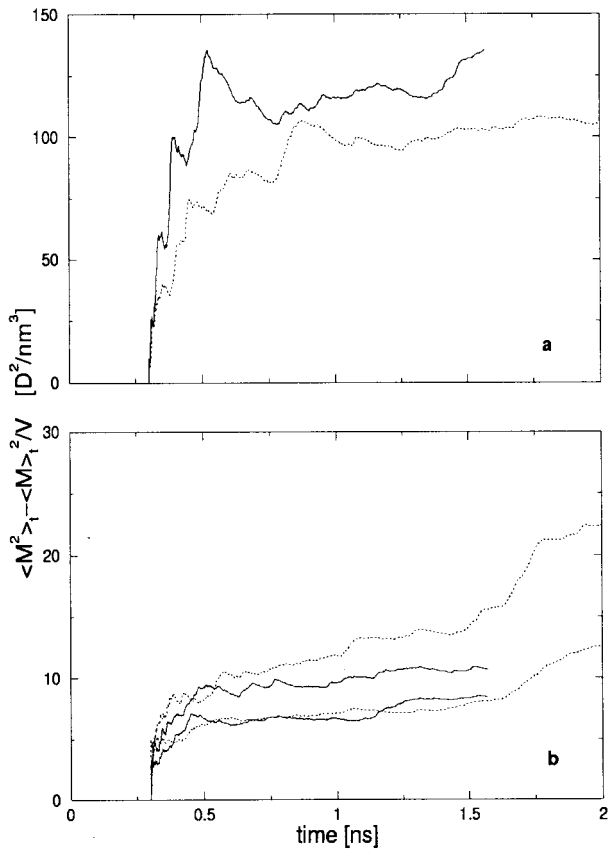

FIGURE 8 Total dipole moment fluctuation per unit volume $(\langle \mathbf{M^2} \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time for the simulations of α-lactalbumin (α-LAC) at pH 2.0 (···) and pH 8.0 (——). (*a*) Total dipole moment fluctuation per unit volume calculated over all protein atoms; (*b*) Total dipole moment fluctuation per unit volume calculated for all protein atoms excluding charged residues and the charged N and C termini (*upper two curves*) as well as the contribution of the protein backbone only (*lower two curves*).

are ~30% higher than in the low-pH simulation, yielding values of $\epsilon(0)_{\text{protein}}$ of ~6.2 (high pH) and 12.6 (low pH), respectively. In our analysis, we discovered that the simulation of the high-pH state had been carried out with an incorrect net charge of $-0.74\ e$ rather than $-1\ e$ for the terminal CO_2^- of the protein. We made use of the correct charge assignment, which yields a net charge of $-8\ e$, in our calculations. Although the net charge of the protein is lower in magnitude at pH 8.0, the number of charged residues is actually higher, with 34 side chains ionized versus 15 in the low-pH case. The results for the protein excluding charged residues and termini and for the protein backbone show a slight reversal of the results for the entire protein, with lower fluctuations in the high-pH state than in the low-pH state. This is compatible with the observations of Smith et al. (1999) who observed larger atom-positional fluctuations for the low pH simulation. These larger positional fluctuations are only evident once the overwhelming contribution of more than twice as many charged residues in the high-pH state is removed.

In contrast to changes in pH and in solvent environment, changes in the number or composition of ligands bound within the protein appear to have little effect on the calculated $\epsilon(0)$. This is shown in Fig. 9, where calculations on the *apo*- and *holo*-FABP simulations are reported. Results are displayed for the *apo*-FABP with and without its 25 bound water molecules included in each stage of the dipole moment calculation. Similarly, we have analyzed the *holo*-FABP trajectory with and without contributions from the buried palmitate ligand. Over these four calculations, the calculated values of $\epsilon(0)_{\text{protein}}$ vary by less than 14% of the smallest value, and the other permittivities are within the range (1.8–3.5) seen for the other proteins we have studied. As expected, the inclusion of additional dipolar or charged species (the waters or palmitate) in the calculation tends to

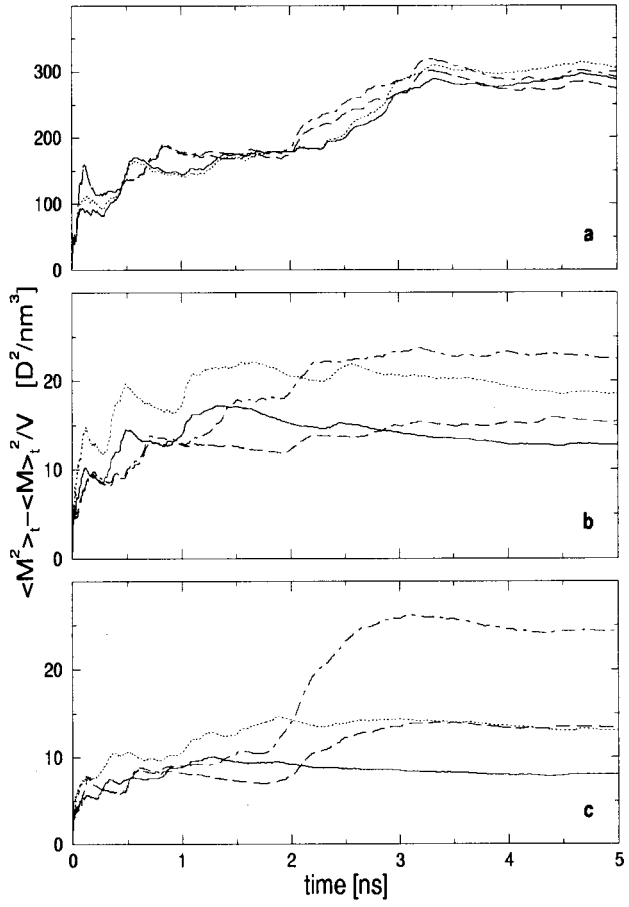

FIGURE 9 Total dipole moment fluctuation per unit volume ($\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time for the simulations of the *apo*- and *holo*-FABP. The solid line corresponds to the *apo*-FABP protein alone, whereas the dotted line shows results for the *apo*-FABP when 25 cavity waters are included in the dipole moment calculations. Similarly, the long-dashed line shows the results for the *holo*-FABP protein, and including the palmitate ligand in the calculation yields the dot-dashed curve. Again, the uppermost panel (*a*) shows the results for the analysis over the entire protein, whereas the contributions from charged residues and termini (though not the palmitate) are excluded in *b*. (*c*) Total dipole moment fluctuations arising from the protein backbone plus the water or palmitate, as appropriate.

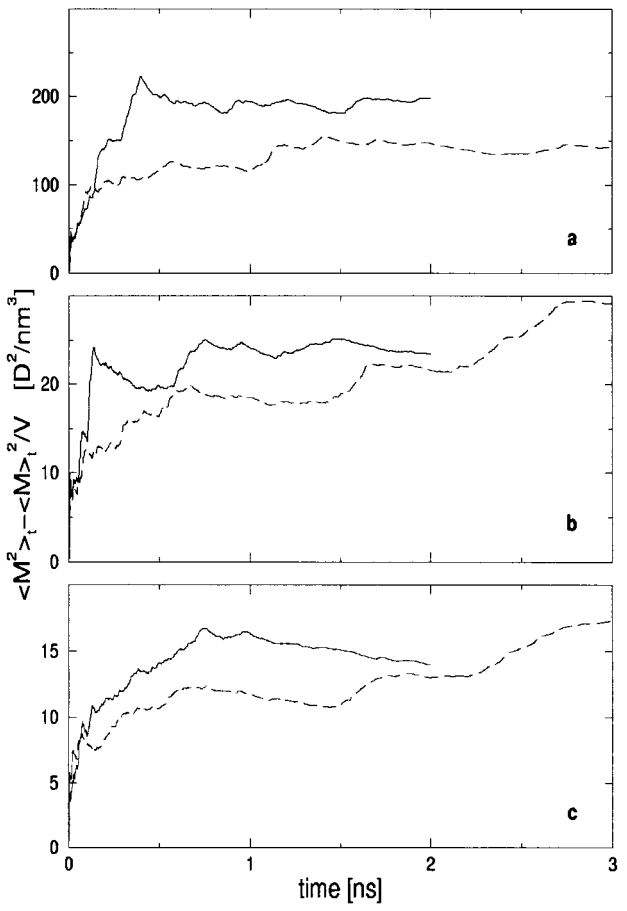

FIGURE 10 Total dipole moment fluctuation per unit volume ($\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 as a function of time for the simulations of the llama antibody heavy-chain variable domain (LLAMA) at 300 K (— — —) and 340 K (——). (*a*) Total dipole moment fluctuation per unit volume calculated over all protein atoms; (*b*) Total dipole moment fluctuation per unit volume calculated for all protein atoms except for charged residues and the charged N and C termini; (*c*) Restricts the calculation further to depict only the contribution of the protein backbone.

yield increased dipole moment fluctuations, but only slightly.

Unlike the first three sets of simulations, the simulations of Voordijk et al. (2000) on the llama antibody variable-domain heavy chain do not vary in terms of the composition of the simulated system. Instead, the same solvated protein fragment was simulated at two temperatures, 300 K and 340 K. The dipole moment fluctuations per unit volume from these two simulations are shown in Fig. 10. As expected, we see an increase in the fluctuations of the dipole moment with increasing temperature. When these results are translated into values for the static dielectric permittivity, we find values for $\epsilon(0)_{\text{protein}}$ of 17.2 (300 K) and 21.2 (340 K). In this analysis, we have not attempted to account for the fact that the solvent dielectric permittivity, ϵ_{RF}, should also change with changes in temperature. For the core of the protein ($\epsilon(0)_{\text{protein,e.c.r.}}$ and $\epsilon(0)_{\text{backbone}}$) we find slightly higher permittivities at 300 K, though this may be an artifact of the difference in simulation times (300 K, 3 ns; 340 K, 2 ns) at the two temperatures. Again, the dipole moment fluctuations calculated over the entire protein are roughly a factor of 10 larger than either value for the core of the protein.

In addition to the comparisons of different conditions acting on the same protein, our data also permit comparison of the total dipole moment fluctuations per unit volume for all five different proteins simulated in water solution at 300 K. These results are displayed in Fig. 11, in a format similar to the preceding figures. One clear result of this presentation is that there is no obvious influence of protein secondary structure on any of the calculated fluctuations. As noted,

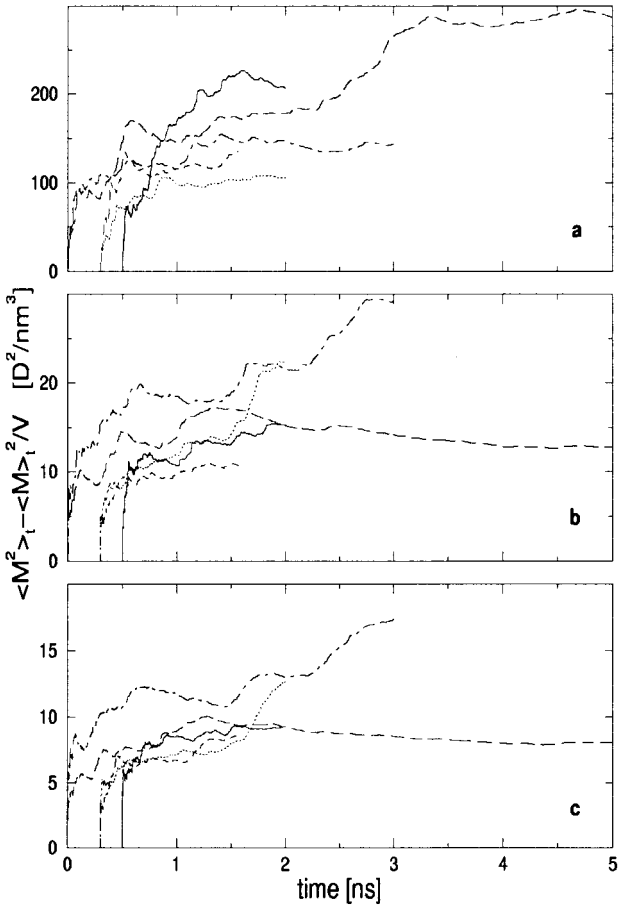

FIGURE 11 Summary of the results for simulations of the different proteins carried out in water at 300 K. The total dipole moment fluctuations per unit volume ($\langle \mathbf{M^2} \rangle - \langle \mathbf{M} \rangle^2)/V$ in units of Debye^2/nm^3 for the simulations of HEWL in water (*solid lines*), α-LAC at pH 2.0 (*dotted lines*), α-LAC at pH 8.0 (*dashed lines*), *apo*-FABP (*long-dashed lines*), and LLAMA (*dot-dashed lines*) are displayed. (*a*) Results of analysis where the entire protein is included; (*b*) Data when charged residues and the charged N and C termini are excluded from the calculation; (*c*) Results when only the protein backbone is considered.

HEWL and α-LAC are predominantly α-helical proteins, whereas the llama antibody and the rat FABP have mostly β-sheet secondary structure. The fluctuations for the three α-helical species are not clearly distinguishable from the results for the β-sheet proteins. However, there is one somewhat useful predictor of the magnitude of the total dipole moment fluctuations (Fig. 11 *a*). Specifically, of the five species compared, *apo*-FABP has the highest number of charged side chains (42) followed by α-LAC at pH 8.0 (34), then HEWL (27), LLAMA (21), and finally α-LAC at pH 2.0 (15). Their total dipole moment fluctuations per unit volume, and thus the values of $\epsilon(0)_{\text{protein}}$, also follow this order with the exception of α-LAC at pH 8.0. It must be noted that the α-LAC simulation at pH 8.0 is the shortest simulation we analyzed, so the observed fluctuations may not be converged in this case.

The results for each simulation we have analyzed are summarized in Table 3. The reported convergence time is based on a visual analysis of the curves displayed in Figs. 4–11. Cases where the calculated value of ϵ did not appear to converge within the available simulation time are indicated in the table with a tilde ($\sim$). Clearly, 1 ns is the absolute minimum of simulation time necessary for calculations of $\epsilon(0)$ with some values of $\epsilon(0)$ not converging within 5 ns of simulation time.

CONCLUSIONS

Our analysis yields values between 10 and 41 for the static dielectric permittivity of each protein, $\epsilon(0)_{\text{protein}}$. When only the protein backbone is considered, we find values between 1.8 and 3.5 ($\epsilon(0)_{\text{backbone}}$), whereas the contributions from all protein atoms except for charged residues yield values for $\epsilon(0)_{\text{protein,e.c.r.}}$ between 2 and 4. These values are entirely compatible with those previously reported in the literature (Simonson and Perahia, 1992; Smith et al., 1993; Simonson and Brooks, 1995; Simonson, 1998). In addition, from the relative magnitude of $\epsilon(0)_{\text{protein}}$ and $\epsilon(0)_{\text{protein,e.c.r.}}$, it is clear that charged residues are the primary determinant of the overall protein dielectric permittivity, as previously observed (Simonson and Perahia, 1992). In further agreement with prior results, a minimum simulation time between 1 and 5 ns is required to accurately converge the calculated values of $\epsilon(0)_{\text{protein}}$. This is certainly an underestimate, as many of our calculations did not show clear convergence within the available 1–5-ns simulation time.

Clearly, the surrounding environment exerts some effect on the behavior of the protein. The more polar water ($\epsilon_{\text{RF}} = 68$) leads to higher fluctuations of HEWL than chloroform ($\epsilon_{\text{RF}} = 5$). Of course, the mobility of atoms in a densely packed crystal is lower than in solution; thus, the HEWL crystal shows a lower dielectric permittivity than the identical molecule in solution. For HEWL, then, we find $\epsilon(0)_{\text{chloroform}} < \epsilon(0)_{\text{crystal}} < \epsilon(0)_{\text{water}}$. For the simulation of HEWL in water we find a slightly lower value for $\epsilon(0)_{\text{protein}}$ than in the prior study of Smith et al. (1993), namely, a value of 26 instead of 30. This is probably due to the use of a reaction-field force in the more recent simulations, which decreases artifacts in the simulation due to truncation of nonbonded interactions. Typically, such artifacts are increased fluctuations of charged atoms and species. A similar decrease in $\epsilon(0)$ has been observed in comparisons of simulations with simple cutoffs and with lattice-sum treatments of long-range electrostatic interactions (Simonson, 1998).

Again emphasizing the significance of charged species, our calculations on α-lactalbumin at two different pH values show that the effective protein dielectric permittivity is higher for the state with more charged side chains, i.e., the high-pH state of α-LAC. Though the results from the high-pH (8.0) simulation are not well converged, it systematically shows a higher total dipole moment fluctuation than

TABLE 3 Summary of results

Protein	System specification	$\langle M_x \rangle$ protein (D)	$\langle M_y \rangle$ protein (D)	$\langle M_z \rangle$ protein (D)	$\lvert M \rvert$ protein (D)	$\langle M^2 \rangle$-$\langle M \rangle^2$ (D^2)	ε_{RF}	$\varepsilon(0)_{protein}$	$\varepsilon(0)_{protein,}$ e.c.r.	$\varepsilon(0)_{backbone}$	$\varepsilon(0)_{protein}$ convergence time (ns)	$\varepsilon(0)_{protein,e.c.r.}$ convergence time (ns)	$\varepsilon(0)_{backbone}$ convergence time (ns)
HEWL	Solution ($CHCl_3$)	−80.5	−26.9	82.0	121.0	1056	5	~15.4	~2.7	2.2			4.0
	Solution	147.1	−87.9	135.7	225.0	3543	68	25.7	~2.6	1.9	1.0	1.3	1.0
	Crystal (4 chains)	96.7	−132.5	−9.1	168.0	1688	68	11.7	2.3	1.9		1.2	1.1
		144.8	−114.4	1.0	189.3	2135		14.9	2.3	1.9			
		79.8	−108.3	84.2	162.7	1876		13.0	2.5	2.1			
		99.1	−173.7	−24.2	204.0	1491		10.4	2.3	1.8			
α-LAC	Low pH (2.0)	−74.9	−55.8	−53.7	114.0	1786	68	12.6	~3.3	~2.3	1.2	1.5	
	High pH (8.0)	52.2	39.5	34.9	84.4	2295	68	~16.2	2.1	1.9		0.5	0.3
FABP	*apo*	−157.1	−15.5	235.7	320.2	5219	68	~40.7	2.9	2.3	3.5	4.5	1.5
	(25 waters)	−147.7	−30.7	234.9	313.6	4921		~37.8	2.3	1.8			
	holo	22.2	140.4	280.0	290.5	4985	68	~38.4	3.3	3.5	4.0	3.0	3.5
	(palmitate)	24.1	131.7	277.0	285.6	4687		~35.6	2.6	2.4			
LLAMA	Room temperature (300 K)	−81.4	−159.2	11.7	163.7	2110	68	17.2	~4.0	~2.8	1.9	2.8	
	High temperature (340 K)	−62.4	−143.8	25.8	164.8	2911	68	21.2	3.1	2.3	0.8	0.7	1.5

the low-pH case. This yields an increased $\epsilon(0)_{protein}$ value of roughly 16.2 vs. 12.6 in the low-pH case. Interestingly, the protein actually has a smaller net charge (−8 vs. +16) at pH 8.0. However, at this higher pH α-LAC also has more than twice as many charged residues (34 vs. 15). A further interesting result is that at low pH α-LAC is known to exist in a molten globule state, with significant side-chain disorder and increased backbone fluctuations (Kuwajima, 1996). This is faintly evident in our results for $\epsilon(0)_{protein,e.c.r.}$ and $\epsilon(0)_{backbone}$; once the overwhelming contributions from charged residues have been excluded, the larger positional fluctuations of the α-LAC backbone at low pH yield slightly higher values for both $\epsilon(0)_{protein,e.c.r.}$ (3.3 vs. 2.1) and $\epsilon(0)_{backbone}$ (2.3 vs. 1.9)

Another environmental effect that changes the permittivity of a protein is the temperature. From our analysis of the llama antibody heavy-chain variable domain at two different temperatures, it is clear that $\epsilon(0)_{300K} < \epsilon(0)_{340K}$. As expected, the higher temperature yields larger fluctuations of the protein atoms and thus a larger fluctuation of the dipole moment. It should be noted that the increase in temperature should also alter the dielectric of the surrounding solvent, ϵ_{RF}, but our analysis has not taken this into account. Nonetheless, we are confident that an increase in the temperature yields a concomitant increase in the protein dielectric permittivity.

In contrast to the three effects mentioned previously, it is clear that buried ligands or buried water molecules have almost no effect on the effective permittivity of a protein. This conclusion arises from our analysis of two 5-ns simulations of the *apo*- and *holo*-FABP. FABP contains a large internal cavity that binds a palmitate ligand in the holoprotein complex. In the apoprotein, this cavity is filled by roughly 25 tightly bound water molecules. Interestingly, the cavity contents have little influence on the effective dielectric calculated for FABP, as we find roughly similar values of $\epsilon(0)_{protein}$ for the empty apoprotein, the apoprotein with water in the binding cavity, and the holoprotein with and without the palmitate ligand.

Our analysis shows that although environmental effects on the static dielectric permittivity of various proteins are non-negligible, they can be easily understood. To significantly change the overall static dielectric permittivity of a protein, it is necessary to alter the behavior of the functional groups that define the total dipole moment, namely, charged surface side chains. Buried water molecules or ligands exert little effect on $\epsilon(0)$, whereas appreciable effects are observed from changes in solvent, pH, or temperature, all of which clearly alter the dynamics of those crucial charged side chains.

We thank all our collaborators who provided us with trajectories for analysis: Dr. Urs Stocker (HEWL in solution and HEWL in crystal), Dr. Lorna Smith (α-LAC), Dr. Dirk Bakowies (FABP), and Dr. Tomas Hansson (LLAMA). Without their contributions this work would not have been possible.

REFERENCES

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for water in relation to protein hydration. *In* Intermolecular Forces. B. Pullman, editor. Reidel, Dordrecht, The Netherlands. 331–342.

Creighton, T. E. 1984. Proteins. Freeman, New York.

Fröhlich, H. 1958. Theory of Dielectrics. Clarendon, Oxford, UK.

Gilson, M. K., and B. H. Honig. 1986. The dielectric constant of a folded protein. *Biopolymers.* 25:2097–2119.

Heinz, T. N., W. F. van Gunsteren, and P. H. Hünenberger. 2001. Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J. Chem. Phys.* In press.

Honig, B. H., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

King, G., F. S. Lee, and A. Warshel. 1991. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* 95:4366–4377.

Kirkwood, J. G. 1939. The dielectric polarizibility of polar liquids. *J. Chem. Phys.* 7:911–919.

Kuwajima, K. 1996. The molten globule state of α-lactalbumin. *FASEB J.* 10:102–109.

Lide, D. R., ed. 2000. Handbook of Chemistry and Physics, 81st ed. CRC Press, Boca Raton, FL.

Nakamura, H., T. Sakamoto, and A. Wada. 1988. A theoretical study of the dielectric constant of protein. *Protein Eng.* 2:177–183.

Neumann, M., O. Steinhauser, and G. S. Pawley. 1984. Consistent calculation of the static and frequency-dependent dielectric constant in computer simulations. *Mol. Phys.* 52:97–113.

Ramachandran, G. N., and V. Sasisekharan. 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23:283–437.

Sanner, M. F., J. C. Spehner, and A. J. Olson. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers.* 38:305–320.

Scaife, B. K. P. 1989. Principles of Dielectrics. Oxford University Press, New York.

Scott, W.R.P., P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter J. Fennen, A. E. Torda, T. Huber, P. Krüger and W. F. van Gunsteren. 1999. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A.* 103:3596–3607.

Simonson, T. 1998. The dielectric constant of cytochrome c from simulations in a water droplet including all electrostatic interactions. *J. Am. Chem. Soc.* 20:4875–4876.

Simonson, T., and C. L. Brooks. 1995. Charge screening and the dielectric constant of proteins: insights from molecular dynamics. *J. Am. Chem. Soc.* 118:8452–8458.

Simonson, T., and D. Perahia. 1992. Internal and interfacial dielectric properties of cytochrome c from molecular dynamics in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.* 4:1082–1086.

Smith, P. E., R. M. Brunne, A. E. Mark, and W. F. van Gunsteren. 1993. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. *J. Phys. Chem.* 97:2009–2014.

Smith, L. J., C. M. Dobson, and W. F. van Gunsteren. 1999. Molecular dynamics simulations of human α-lactalbumin: changes to the structural and dynamical properties of the protein at low pH. *Proteins.* 36:77–86.

Stocker, U., K. Spiegel, and W. F. van Gunsteren. 2000. On the similarity of properties in solution or crystalline state: a molecular dynamics study of hen lysozyme. *J. Biomol. NMR.* 18:1–12.

Stocker, U., and W. F. van Gunsteren. 2000. Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins.* 40:145–153.

Svensson, B., B. Jönsson, and C. Woodward. 1990. Electrostatic contributions to the binding of Ca^{2+} in calbindin mutants. *Biophys. Chem.* 38:179–183.

Tironi, I. G., R. Sperb, P. E. Smith, and W. F. van Gunsteren. 1995. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* 102:5451–5459.

Tironi, I. G., and W. F. van Gunsteren. 1994. A molecular dynamics simulation study of chloroform. *Mol. Phys.* 83:381–403.

van Gunsteren, W. F., and H. J. C. Berendsen. 1987. Groningen Molecular Simulation (GROMOS) Library Manual. Biomos, Groningen, The Netherlands.

van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Vdf Hochschulverlag AG an der ETH Zürich, Zürich.

Voordijk, S., T. Hansson, D. Hilvert, and W. F. van Gunsteren. 2000. Molecular dynamics simulations highlight mobile regions in proteins: a novel suggestion for converting a murine V_H domain into a more tractable species. *J. Mol. Biol.* 300:963–973.