Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics

Thomas Simonson*,† and Charles L. Brooks III*,‡

Contribution from the Institut de Génétique et Biologie Moléculaire et Cellulaire (C.N.R.S.), 1 rue Laurent Fries (C.U. de Strasbourg), 67404 Illkirch, France, and Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received March 18, 1996. Revised Manuscript Received June 19, 1996[⊗]

Abstract: The dielectric constants of myoglobin, apomyoglobin, the B fragment of staphylococcal protein A, and the immunoglobulin-binding domain of streptococcal protein G are calculated from 1−2 ns molecular dynamics simulations in water, using the Fröhlich−Kirkwood theory of dielectrics. This dielectric constant is a direct measure of the polarizability of the protein medium and is the appropriate macroscopic quantity to measure its relaxation properties in response to a charged perturbation, such as electron transfer, photoexcitation, or ion binding. In each case the dielectric constant is low (2−3) in the protein interior, then rises to 11−21 for the whole molecule. The large overall dielectric constant is almost entirely due to the charged protein side chains, located at the protein surface, which have significant flexibility. If these are viewed instead as part of the outer solvent medium, then the remainder of the protein has a low dielectric constant of 3−6 (depending on the protein), comparable to that of dry protein powders. Similar results were already observed for ferro- and ferricytochrome *c*, and are probably valid for many or most stable globular proteins in solution, leading to a rather comprehensive picture of charge screening and the dielectric constant of proteins. This picture suggests ways, and supports some ongoing efforts, to improve current Poisson−Boltzmann models. Indeed, treating a protein as a homogeneous, low dielectric medium is likely to underestimate the actual dielectric relaxation of the protein; this would affect calculations of the self-energy of titrating protons, or the reorganization energy of a redox electron.

1. Introduction

Electrostatic forces play a decisive role in protein structure and function, and they have been the object of intense experimental and theoretical scrutiny.[1−8] Charge screening by protein and solvent is very complex. Recent experimental studies of charge screening in proteins include measurements of pK_a and redox potential shifts due to point mutations,[9,10] of emission or absorption shifts of a spectroscopic probe,[11−13] and of gas-phase basicities of various protein ions, including cytochrome *c*.[14] Several features emerge from these and earlier studies. First, Coulomb interactions between protein groups are strongly screened by the surrounding solvent, often being reduced by a factor of 40−80 compared to the bare gas-phase interaction. Second, the protein itself contains many polar groups, which allow it, for example, to stabilize charged or polar reaction intermediates to about the same extent as bulk water. Third, the polarizability of the protein is low, at least in the interior. Thus even highly polar enzyme active sites are likely to have a moderate polarizability. The protein tends to use rather rigid, preorganized, polar groups to stabilize enzyme reaction intermediates, unlike bulk water, which is highly polarizable. This moderate internal polarizability is important in reducing the reorganization energy, and hence the activation barrier, for proton and electron transfer in enzymes.

To help quantify the polarizability of the protein interior, we have analyzed nanosecond molecular dynamics simulations of four proteins in water: myoglobin, apomyoglobin, the 60-residue B fragment of staphylococcal protein A, and the 56-residue immunoglobin-binding domain of streptococcal protein G. By comparing the magnitude of their dipole fluctuations to the predictions of dielectric theory, we can estimate the dielectric constants of the four proteins. The dielectric constant calculated in this way is a direct measure of the protein polarizability (rather than of its polarity). It should not be confused with effective "dielectric constants" sometimes quoted in the literature, which measure the reduction in the interaction energy between two protein sites, compared to the gas-phase interaction, and which normally include a large solvent contribution. We also analyze more complex, microscopic, effects, namely the spatial variation of the dielectric constant within each protein, and the relative contributions of charged and polar side chains. Early applications of this approach were limited to proteins in vacuum, because of insufficient computer resources,[15−17] leading to low estimates of the dielectric constant. More recent

* Address correspondence to either author.
† Institut de Génétique et Biologie Moléculaire et Cellulaire (C.N.R.S.).
‡ The Scripps Research Institute.
⊗ Abstract published in *Advance ACS Abstracts,* August 1, 1996.

(1) Perutz, M. *Science* **1978**, *201*, 1187−1191.
(2) Warshel, A.; Russell, S. *Q. Rev. Biophys.* **1984**, *17*, 283−342.
(3) Krishtalik, L. *J. Theor. Biol.* **1985**, *116*, 201−214.
(4) Rogers, N. *Prog. Biophys. Mol. Biol.* **1986**, *48*, 37−66.
(5) Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *19*, 301−332.
(6) Moser, C.; Keske, J.; Warncke, K.; Farid, R.; Dutton, P. *Nature* **1992**, *355*, 796−802.
(7) Warshel, A. *Computer modelling of chemical reactions in enzymes and solutions*; John Wiley: New York, 1991.
(8) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(9) Russell, A.; Fersht, A. *Nature* **1987**, *328*, 496−500.
(10) Varadarajan, R.; Zewert, T.; Gray, H.; Boxer, S. *Science* **1989**, *243*, 69−72.
(11) Lockhardt, D.; Kim, P. *Science* **1992**, *257*, 947−951.
(12) Pierce, D.; Boxer, S. *J. Phys. Chem.* **1992**, *96*, 5560−5566.
(13) Steffen, M.; Lao, K.; Boxer, S. *Science* **1994**, *264*, 810−816.
(14) Schnier, P.; Gross, D.; Williams, E. *J. Am. Chem. Soc.* **1995**, *117*, 6747−6757.
(15) Gilson, M.; Honig, B. *Biopolymers* **1986**, *25*, 2097−2119.
(16) Nakamura, H.; Sakamoto, T.; Wada, A. *Prot. Eng.* **1988**, *2*, 177−183.
(17) Simonson, T. Electrostatics and dynamics of proteins. Ph.D. Thesis, University of Paris, 1989.

S0002-7863(96)00884-0 CCC: $12.00

applications to trypsin, lysozyme, pancreatic trypsin inhibitor (BPTI), and cytochrome *c* in water[18−20] have shown that the protein dielectric constant can vary from as little as 2 in the protein center to values of 25 or more in the polar outer region, with intermediate values in certain specific regions such as the trypsin active site. In contrast, measurements on very dry protein powders have given a static dielectric constant in the range of 2−5 for several proteins, including myoglobin and cytochrome *c*.[21−23] Because of this variability, data for additional proteins are important to clarify which types of behavior can be generalized, and whether there are specific differences between electron transfer proteins such as cytochrome *c*, enzymes such as lysozyme and trypsin, and other proteins. We shall see that for all the proteins analyzed, the dielectric constant varies from 2−3 in the interior of the molecule to 13−30 in the outer part. This large outer dielectric constant arises almost entirely from the flexible charged protein side chains located at the protein surface.

2. Methods

Dielectric Constant. To apply the Fröhlich−Kirkwood theory of dielectrics, we view the protein as a spherical dielectric medium with a dielectric constant ϵ_p and radius r_p, surrounded by an infinite solvent medium with a dielectric constant ϵ_w. We then have[24]

$$\langle \Delta M_p^2 \rangle / kTr_p^3 = (2\epsilon_w + 1)(\epsilon_p - 1)/(2\epsilon_w + \epsilon_p) \quad (1)$$

where ΔM_p is the deviation of the protein dipole moment from its mean and the brackets represent an ensemble average (taken as a time average over the trajectory in the current study). Notice that the deviation ΔM_p is not origin-dependent. The dimensionless term on the left side is conveniently referred to as the *G* factor, by analogy to the Kirkwood *g* factor of a homogeneous liquid.

In order to analyze the radial variation of the dielectric constant going from the center of the protein to its surface, we can view the protein as made up of two distinct, concentric, spherical regions (Figure 1): an inner region of radius r_1 and dielectric constant ϵ_1, and an outer region of radius $r_2 = r_p$ and dielectric constant ϵ_2. In this view, an analogous fluctuation formula is easily derived:[20]

$$\langle \Delta M_1^2 \rangle / kTr_1^3 = \frac{(\epsilon_1 - 1)[(1 + 2\epsilon_2)(2\epsilon_w + \epsilon_2) - 2(r_1/r_2)^3(\epsilon_w - \epsilon_2)(1 - \epsilon_2)]}{(\epsilon_1 + 2\epsilon_2)(2\epsilon_w + \epsilon_2) - 2(r_1/r_2)^3(\epsilon_w - \epsilon_2)(\epsilon_1 - \epsilon_2)} \quad (2)$$

where ΔM_1 corresponds to the dipole moment of the inner region 1.

The probability distribution of ΔM_p is predicted by continuum theory to have a simple form,[24] which has been observed in polar liquids:[25−27]

$$p(\Delta M_p) = 4\pi \Delta M_p^2 A \exp(-a\Delta M_p^2) \quad (3)$$

where *a* is a simple function of the dielectric constants ϵ_p and ϵ_w and the radius r_p. *A* is a normalization factor.

Computational Details. Our analysis of the dielectric response in the proteins listed above is based on simulations carried out by Brooks and co-workers.[28−31] For each system studied, molecular dynamics in explicit solvent was carried out for periods exceeding 1 ns of production data collection, following a minimum of 200−300 ps of equilibration. In Table 1 we summarize the basic conditions of these simulation studies. In general, the molecular dynamics simulations were performed with periodic boundary conditions, using the CHARMM/param19 force field[32] and the TIP3P water model.[33] Heavy atom−hydrogen covalent bond lengths and the internal geometry of the water molecules were kept fixed with the SHAKE algorithm.[34] Long-range forces were truncated using an atom-based force-shift method,[35] using a list-based algorithm with a 1 to 2 Å shell beyond the nonbonded cutoffs listed in Table 1. This list was updated every 20−25 dynamics steps. The time step for integration of the equations of motion was 2 fs. Temperature was maintained near the target 300 K by periodic reassignment of velocities from a Boltzmann distribution if the average temperature over the period of 1000−2000 steps was outside a window of 15 K around the target temperature.

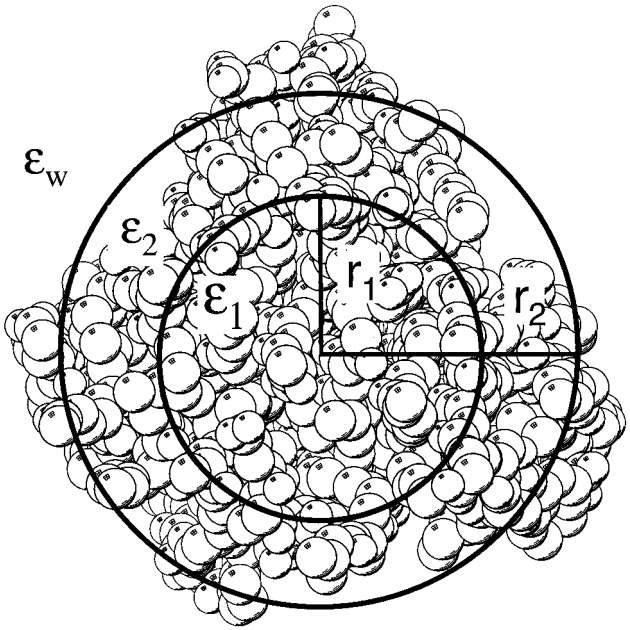

Figure 1. Myoglobin viewed as two concentric dielectric regions (CPK view, produced with the program Molscript (P. Kraulis)).

3. Results

Myoglobin. The first system we examine is myoglobin in its deoxy form. The spherical radius for the protein is taken to be $(5/3)^{1/2}$ its radius of gyration, giving $r_p = 20.4$ Å, while the axes of inertia have half lengths of 23.7, 22.0, and 14.1 Å, indicating a somewhat flattened shape. Excluding the charged portions of the charged side chains gives a radius of 19.9 Å.

Convergence of the *G* factor is shown in Figure 2. It is useful to calculate the protein dipole moment with and without the contribution of charged portions of the charged side chains, Asp, Glu, Arg, and Lys. Indeed, the fluctuations of the charged side chains are found to be quite different from the protein bulk, as noted for cytochrome *c*.[20] We shall refer to the "overall" *G* (charged side chains included) and the "uncharged" *G* (charged portions of the charged side chains excluded). Convergence is good for the uncharged *G*, while the overall *G* appears to be still increasing slowly after 1.6 ns.

(18) King, G.; Lee, F.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366−4377.
(19) Smith, P.; Brunne, R.; Mark, A.; van Gunsteren, W. *J. Phys. Chem.* **1993**, *97*, 2009−2014.
(20) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082−1086.
(21) Rosen, D. *Trans. Faraday Soc.* **1963**, *59*, 2178−2191.
(22) Bone, S.; Pethig, R. *J. Mol. Biol.* **1982**, *157*, 571−575.
(23) Bone, S.; Pethig, R. *J. Mol. Biol.* **1985**, *181*, 323−326.
(24) Fröhlich, H. *Theory of Dielectrics*; Clarendon Press: Oxford, 1949.
(25) Kusalik, P. *Mol. Phys.* **1993**, *80*, 225−231.
(26) Kusalik, P.; Mandy, M.; Svishchev, I. *J. Chem. Phys.* **1994**, *100*, 7654−7664.
(27) Simonson, T. *Chem. Phys. Lett.* **1996**, *250*, 450.
(28) Boczko, E. M.; Brooks, C. L., III *Science* **1995**, *269*, 393−396.
(29) Brooks, C. L., III *Proteins* In preparation.
(30) Boczko, E. M.; Brooks, C. L., III *J. Mol. Biol.* In preparation.
(31) Sheinerman, F. B.; Brooks, C. L., III *J. Mol. Biol.* Submitted.
(32) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187−217.
(33) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(34) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comp. Phys.* **1977**, *23*, 327−341.
(35) Brooks, C., III; Pettitt, B. M.; Karplus, M. *J. Chem. Phys.* **1985**, *83*, 5897.

Table 1. Characteristics of Protein Simulations

protein	periodic volume	edge[a] (Å)	no. of water molecules	cutoff[b] (Å)	simulation time (ps)
myoglobin	parallelepiped	56.57/56.57/37.71	2788	9.25	1500
apomyoglobin	cube	56.57	4783	9.25	1500
protein G	truncated octahedron	62.21	3659	11.0	2000
protein A	truncated octahedron	62.21	3627	11.0	2000

[a] For the parallelepiped volume all three edge lengths are given. [b] The cutoff is the atom based cutoff for energy and force calculations. The list for processing the nonbonded energy and forces was built using a 1−2/Å shell beyond this value.

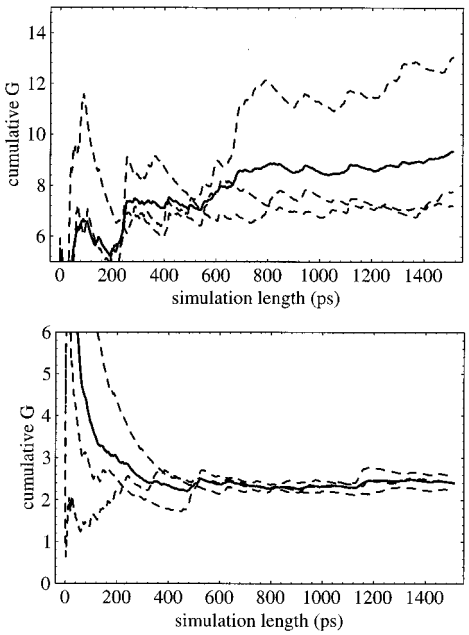

Figure 2. Convergence of the myoglobin *G* factor during the simulation. The cumulative average of *G* is shown (Cartesian components dashed) with (upper panel) and without (lower panel) charged side chains.

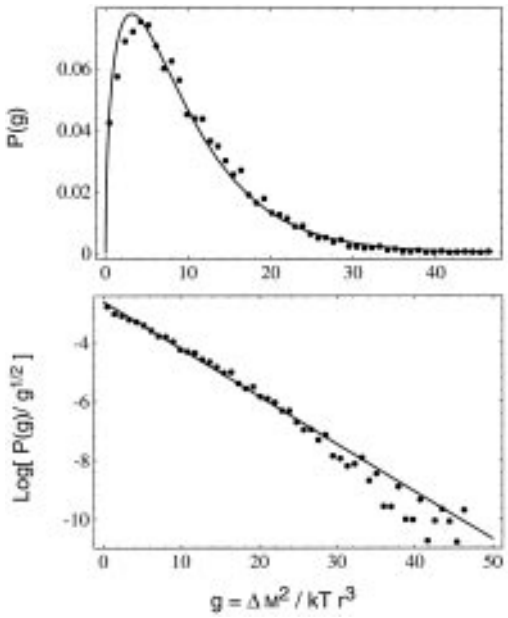

Figure 3. Probability distribution of myoglobin total dipole moment. The theoretical form of the probability distribution[25] is superimposed (solid lines).

The variance of the overall protein dipole moment is 141 (eÅ)^2; the variances of its three Cartesian components are 36, 66, and 39 (eÅ)^2. The uncertainty in $\langle \Delta M_p^2 \rangle$, computed as the dispersion of 300 ps batch averages, is ±30 (eÅ)^2. The dimensionless *G* factor is then 9.3 ± 2.0. If the charged portions of the charged protein side chains are omitted from the calculation of $\langle \Delta M_p^2 \rangle$, we obtain an "uncharged" result of 33.6 ± 5.2 (eÅ)^2, and $G = 2.4 \pm 0.4$. Thus a large portion of the dipole fluctuation arises from the charged side chains.

The probability distribution of the overall ΔM_p seen in the simulation is shown in Figure 3, along with the probability distribution predicted by continuum theory.[25] The agreement is rather good, consistent with the continuum view we wish to adopt in this analysis. The agreement is similar when the charged portions of the charged side chains are excluded (not shown). Similar agreement was seen for cytochrome *c*.[36] If $\langle \Delta M_p^2 \rangle$ is estimated using a fit to the continuum probability distribution, instead of simple averaging over the trajectory, a slightly lower overall *G* of 8.4 is obtained.

A protein dielectric constant of 11 is obtained using $G = 9.3$ in (1). This is significantly higher than the values of 2−5 measured for dry protein powders, and often assumed for protein interiors (e.g. in Poisson−Boltzmann calculations[5]), but lower than values calculated for cytochrome *c*,[20] BPTI, and lysozyme.[19] If the slightly lower estimate of $G = 8.4$ is used (based on a fit to the continuum probability distribution of ΔM_p^2, we obtain a dielectric constant of 10. When charged portions of the charged side chains are omitted from the analysis, a value of 3.5 is obtained, in the same range as for dry myoglobin powders.[21]

These theoretical estimates include an implicit contribution from electronic polarizability (in addition to the explicit contribution of dipolar reorientation), since the force field used contains such an implicit contribution. We can make an explicit estimate of the contribution of electronic polarizability to the dielectric constant as follows. We first estimate the high-frequency dielectric constant of the protein from the Clausius−Mossotti equation. This gives a value of about 2.[37] We then insert this value into the Fröhlich−Kirkwood fluctuation formula (in its most general form),[20] along with our observed *G* factor. This gives us a total dielectric constant ϵ_p that is increased by about 1, compared to our initial estimate. However, this procedure presumably overestimates the effect of electronic polarizability, since the observed *G* factor already contains part of its effect. Therefore the exact additional contribution of electronic polarizability to ϵ_p is probably between 0 and 1.

The dielectric constant is sensitive to statistical uncertainty in $\langle \Delta M_p^2 \rangle$, and to the values chosen for the parameters r_p and ϵ_w, which are not uniquely defined. We can estimate the uncertainty in ϵ_p by varying $\langle \Delta M_p^2 \rangle$, r_p, and ϵ_w over a reasonable range. The sensitivity to ϵ_w turns out to be very weak, indicating that our idealization of the protein surroundings as a single, homogeneous, solvent medium with $\epsilon_w = 80$ is relatively benign. The sensitivity to statistical error in $\langle \Delta M_p^2 \rangle$ is larger: a change of ±30 (eÅ)^2 (one standard error) changes ϵ_p by ±3. The sensitivity to the protein radius r_p is also significant: a change of ±1 Å changes ϵ_p by ∓2. Thus we estimate the overall uncertainty in ϵ_p to be 3−4. Since the overall *G* is still increasing slowly at the end of our simulation, the observed *G* and the resulting dielectric constant are probably too low. A

(36) Simonson, T.; Perahia, D. *Faraday Discuss.* **1996**, *103*, 000.

(37) Simonson, T.; Perahia, D.; Bricogne, G. *J. Mol. Biol.* **1991**, *218*, 859−886.

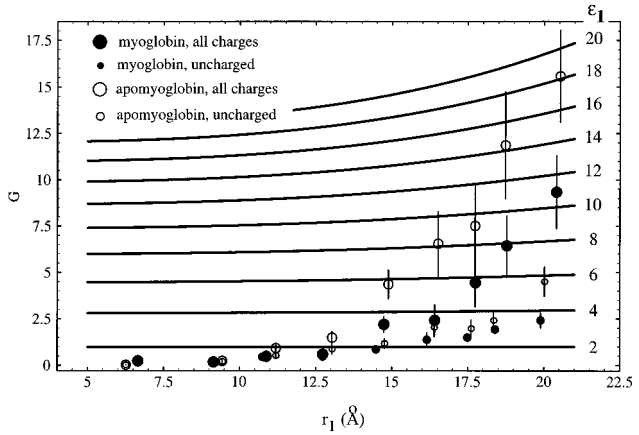

Figure 4. Radial variation of the G factor of inner region 1 for myoglobin and apomyoglobin. Theoretical curves (solid lines) are labeled on the right by the value of the dielectric constant ϵ_1.

longer simulation would then bring the dielectric constant closer to values obtained for cytochrome *c*, lysozyme, and BPTI. When charged side chains are excluded, the uncertainty in the dielectric constant is reduced to about ± 1.

The radial variation of the protein dielectric constant, going from its center to its surface, is analyzed in Figure 4. The protein is subdivided into two regions: we view the inner part as a microscopic cavity and the outer part as a continuum (Figure 1). Equation 2 gives the G factor of the inner region as a function of its radius r_1 and its dielectric constant ϵ_1. Before we can apply eq 2, however, we need to assume a value for ϵ_2, the dielectric constant of the outer portion of the protein. We could assume that ϵ_2 has the overall value $\epsilon_p = 11$ obtained above, or we could assume the overall value ϵ_p is a volume-weighted average of ϵ_1 and ϵ_2:

$$r_2^3\epsilon_p = r_1^3\epsilon_1 + (r_2^3 - r_1^3)\epsilon_2 \qquad (4)$$

Both of these assumptions give rise to values for G that are similar in all regions of the $G-r_1$ plane, except for regions of high ϵ_1 and low r_1. However, these are not physically interesting regions for proteins since, uniformly, the inner dielectric constant of the protein is rather low. In fact, it is nearly equivalent, and more convenient, to use a third assumption: $\epsilon_2 = 20$. This also gives similar values of G in the physically interesting regions of the $G-r_1$ plane. This last assumption obviously does not make use of the ϵ_p value calculated for the particular protein under consideration. Thus it can be used for all four proteins studied, allowing us to compare the simulation data from all four proteins to the same theoretical model.

Using the $\epsilon_2 = 20$ assumption, Figure 4 plots the theoretical G vs r_1 for a series of values of ϵ_1, along with the G values observed in the simulation. For $r_1 \leq 10$, the observed G values fall below the curve $\epsilon_1 = 2$. Thus the dielectric constant is quite low in the inner part of the molecule, which is less polar and mobile than the exterior. For $r_1 > 10$, the observed G values increase rapidly, finally reaching the $\epsilon_1 = 11$ curve at the protein surface. Thus the dielectric constant in the outer part of the protein is much higher than in the interior, and dominated by the charged side chains, located at the protein surface. This is similar to the behavior seen in cytochrome *c*.[20]

Since $\epsilon_1 = 2$ and $\epsilon_p \approx 11$, eq 4 gives an estimate of $\epsilon_2 = 13$ for the dielectric constant of the outer half of the protein (atoms more than 12 Å from the center). The shortest half dimension of myoglobin is 14 Å, so that the rise in the dielectric constant corresponds approximately to the point where surface groups begin to be included in the inner region 1. This is also the point where charged residues begin to be included in region 1.

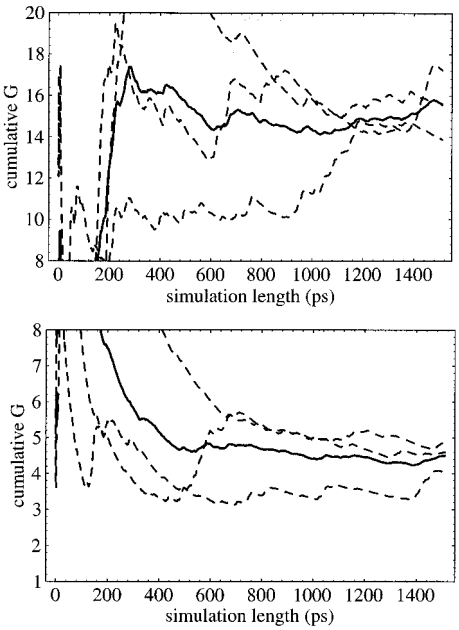

Figure 5. Convergence of the apomyoglobin G factor during the simulation. The cumulative average of G is shown (Cartesian components dashed) with (upper panel) and without (lower panel) charged side chains.

When the charged portions of the charged side chains are omitted from the analysis, the ("uncharged") G values remain much smaller as r_1 increases, rising gradually to reach the $\epsilon_1 = 3.5$ curve at the protein surface. Thus if we view the charged portions of the charged side chains as part of the outer, solvent medium, the remaining bulk of the protein has a low dielectric constant of 3.5, very close to that of dry myoglobin powders.[21] Again, this behavior is similar to cytochrome *c*.

Apomyoglobin. Depleted of its heme group, apomyoglobin has a radius of 20.5 Å, or 20.0 Å if the charged groups are excluded, only slightly larger than the intact holomyoglobin. The axes of inertia have half lengths 23.2, 22.5, and 14.6 Å, indicating again a flattened shape. The increased fluctuations of apomyoglobin, compared to holomyoglobin, are reflected in increased fluctuations of the protein density in each radial shell, by approximately a factor of 2.

The quality of convergence of the overall and the uncharged G, shown in Figure 5, is similar to that of myoglobin, although the z component of the overall G has not fully converged after 1.8 ns. The final overall G is 15.6 ± 2.5; its Cartesian components are 5.2, 4.6, and 5.7. Thus the anisotropy of the dipole fluctuations is moderate. The "uncharged" G is 4.5 ± 0.8; its Cartesian components are 1.5, 1.6, and 1.3. As in myoglobin and cytochrome *c*, the overall G is dominated by the charged portions of the charged side chains.

The probability distribution of the overall G (not shown) is in somewhat poorer agreement with the continuum prediction than in the myoglobin case. For the "uncharged" G the agreement is very good.

The radial variation of the G factor is shown in Figure 4 along with the holomyoglobin data and the Fröhlich−Kirkwood theoretical curves. While apomyoglobin has several internal water molecules, their contribution is not included in our calculation of G. Nevertheless, the looser apomyoglobin structure leads to a systematically higher G factor, compared to holomyoglobin, and a larger overall dielectric constant of 18 ± 7 with all charges included, or 6 ± 2 with charged groups excluded. The protein interior has a low dielectric constant of 1−4. Again, the overall dielectric constant is dominated by the charged side chains, located at the protein surface.

Table 2. Summary of Protein Properties

protein	radius (Å)	$\langle \Delta M^2_{x,y,z} \rangle^b$	$\langle \Delta M^2 \rangle$	*G* factor	ϵ
myoglobin, all	20.4	36, 66, 39	141(30)	9.3(2.0)	11(4)
myoglobin, u.c.[a]	19.9	12, 11, 10	33.3(5.2)	2.4(0.4)	3.5(1.0)
apomyoglobin, all	20.5	80, 71, 88	239(38)	15.6(2.5)	18(7)
apomyoglobin, u.c.	20.0	22, 23, 19	64.0(11.4)	4.5(0.8)	6(2)
protein G, all	13.8	30, 27, 25	82.0(22.1)	17.5(4.7)	18(8)
protein G, u.c.	13.3	2.4, 3.2, 5.5	11.1(3.0)	2.7(0.7)	4(1)
protein A, all	12.5			>16	>20[c]
protein A, u.c.	12.2			~3	~4
ferricytochrome *c*, all	16.5	36, 25, 71	132(6)	16.6(0.8)	25(10)
ferricytochrome *c*, u.c.	15.9	7, 5, 4	15.8(2.0)	2.3(0.3)	3.4(1.0)
ferrocytochrome *c*, all	16.5	37, 20, 70	127(7)	15.9(0.9)	24(10)
ferrocytochrome *c*, u.c.	15.9	9, 7, 8	23.9(3.0)	3.6(0.5)	4.7(1.0)

[a] u.c. = uncharged, i.e. excluding charged portions of charged side chains from the analysis. [b] Variances of the dipole components along each Cartesian axis. [c] Exact value not known due to incomplete convergence. [d] Dipole moments in eÅ.

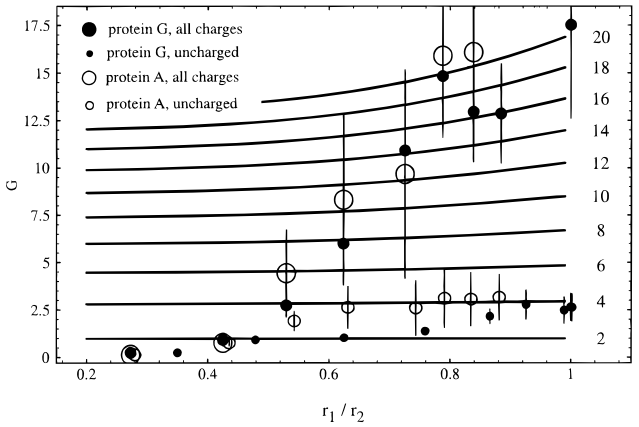

Figure 6. Radial variation of the *G* factor of inner region 1 for protein G and protein A. r_1 is measured as a fraction of the total protein radius r_2. Theoretical curves (solid lines) are labeled on the right by the value of the dielectric constant ϵ_1.

Protein G Fragment. The protein G fragment is the 56-residue, stable, immunoglobin-binding domain of streptococcal protein G. It is made up of an α-helix lying on a four-stranded β-sheet. Its effective radius is 13.8 Å including all atoms, or 13.3 Å excluding charged portions of charged side chains. The axes of inertia (all atoms) have half lengths of 18.2, 11.6, and 10.0 Å, so that the molecule is rather elongated.

Results for the dipole fluctuations and dielectric constant are qualitatively similar to myoglobin and apomyoglobin (Table 2). Convergence of the overall *G* appears reasonable (not shown), convergence of the uncharged *G* is quite good, and anisotropy of the dipole fluctuations is moderate with and without the charged groups. The probability distributions of the overall and uncharged *G* (not shown) follow closely the continuum prediction.

The overall dielectric constant is estimated to be 18 ± 8 with all charges, and 4 ± 1 with charged groups excluded. While the radial variation (Figure 6) is slightly noisier than for apomyoglobin and holomyoglobin, the qualitative behavior is the same. The overall dielectric constant is large and dominated by the charged side chains at the protein surface. Excluding these charged groups gives a low bulk dielectric constant, which increases gradually from 1−2 in the interior to 4 near the surface.

Protein A Fragment. The protein A fragment studied is the 60-residue B fragment of staphylococcal protein A. It is a three-helix bundle, with an effective radius of 16.1 Å (15.8 Å excluding charged groups), and half-axes of inertia 23.1, 12.2, and 9.9 Å. Its stability is low; it exhibits two folded forms, in equilibrium with one or more partially unfolded forms.[28] As a result, portions of the molecule undergo transient local unfolding at times, and the convergence of the *G* factor is incomplete on the nanosecond time scale sampled here. For regions at the surface that undergo transient local unfolding, the Fröhlich−Kirkwood analysis is not really suitable. The local response to a perturbing field would be nonlinear, and poorly described by a continuum model. We can still apply our analysis to the bulk of the protein, however. When charged groups are excluded, the *G* factor has converged except for the contributions of the outermost residues. When all charges are included, the error bars are large, but there is at least rough convergence for the bulk of the protein. The contribution of the outermost residues is still far from convergence, however. Figure 6 therefore shows the variation of *G* with the radius r_1, omitting the results for the outermost residues (the largest values of r_1). For the bulk of the protein, the general trend of a low inner dielectric constant and a rapid rise to a high overall value for the whole protein appears largely preserved. The overall dielectric constant appears to be over 20, or 4 excluding charged groups.

4. Discussion

Limitations of the Calculations. Our calculations have several obvious limitations. In addition to the finite conformational sampling, which leads to uncertainty in the calculated *G* factors, our calculations are limited by the use of a cutoff for the electrostatic interactions. To estimate the importance of this approximation on the *G* factors and dielectric constants, ideally we should repeat some of the simulations with a much larger cutoff; unfortunately this is very expensive to do. However, a range of different geometries and cutoffs were used for the proteins studied here (Table 1), for ferri- and ferro-cytochrome *c*,[20] and for BPTI and lysozyme.[19] The similar behavior of all eight of these proteins under different truncation conditions suggests that the main features seen here are rather robust, and independent of the exact electrostatic cutoff. The exact values of the dielectric constants, on the other hand, would certainly change under different cutoff conditions. Studies of liquid water droplets in vacuum have shown in addition that the use of a cutoff, in droplet simulations, does not change the general magnitude of the calculated dielectric constant.[20,36]

The continuum view itself has of course limitations due to its macroscopic basis. An average dielectric constant is a rough approximation to the exact microscopic dielectric response of a protein. Nevertheless, it does reproduce many microscopic properties fairly well, as discussed below. This is related in part to the large compensation between protein and solvent contributions to the dielectric relaxation: *e.g.* while the continuum model appears to predict poorly the individual protein and solvent contributions to dielectric relaxation free energies, it gives a reasonable estimate of their sum.[38] When surface

(38) Simonson, T.; Perahia, D. *J. Am. Chem. Soc.* **1995**, *117*, 7987−8000.

portions of a protein undergo very large conformational changes, as with protein A, one expects the continuum model to break down, however, and indeed, we observe here that the Fröhlich−Kirkwood theory fails to predict the magnitude of the dipole fluctuations (even though it may still give a reasonable description of the bulk of the protein).

Equilibrium and Relaxation Properties: The Dual Role of the Protein Dielectric Constant. The electrostatic and dielectric properties of proteins are very complex, and a characterization based on the macroscopic concept of a dielectric constant raises several questions. The usefulness of such a characterization is clear at the outset from the many successes of continuum models in interpreting important properties of proteins,[3−5,8] either in a qualitative or a semiquantitative way. However, great care must be used when discussing results and concepts from experiment, continuum models, and microscopic simulations.

In continuum electrostatics, the dielectric constant plays two fundamental roles: it helps to determine the equilibrium distribution of charge, field, and potential, and it determines the relaxation of the system in response to an external perturbation, such as a redox electron or an ion (see *e.g.* refs 39−41). This duality has often been a source of confusion. In practice, calculations of pure equilibrium properties and calculations of pure relaxation properties can require different values of the protein dielectric constant. Calculations of properties involving both the equilibrium and the relaxation behavior of the protein may require yet a third treatment.

The distinction between the equilibrium and relaxation properties of the protein is best discussed by considering a set of perturbing charges, such as a titrating proton or a redox electron, and decomposing the perturbation free energy A into a static and a relaxation part:[42,43,39]

$$A = A_{\text{static}} + A_{\text{rlx}} \tag{5}$$

The first term on the right is the free energy to introduce the perturbation, constraining the system so that it cannot relax; the second term is the free energy released when the constraints are removed. The first term is determined by the equilibrium electric field in and around the protein, in the absence of the perturbing charges. The second term is determined by the polarizability of the protein and surrounding solvent. (For explicit formulas see *e.g.* ref 38.) In practical applications, different values of the dielectric constant may be suitable for calculating each of these terms, as emphasized recently by Krishtalik et al.[41]

The Fröhlich−Kirkwood dielectric constant calculated here is a direct measure of the *polarizability* of the protein medium, and is indeed the appropriate macroscopic quantity to measure relaxation properties of the protein (*e.g.* A_{rlx}) in response to a charged perturbation, such as electron transfer, photoexcitation, or ion binding. It is, however, only indirectly related to the equilibrium *polarity* of the medium. Indeed, a polar but rigid medium can have a quite weak polarizability. Examples are the heme regions of both myoglobin and cytochrome *c*, which both contain several charged and polar groups, yet have a low Fröhlich−Kirkwood dielectric constant. Other theoretical studies have illustrated the high polarity, yet moderate polarizability of enzyme-active sites.[18,44,45]

The dual role of the dielectric constant has not always been clearly stated, despite many insightful discussions of the problem (*e.g.* refs 46, and 47). When a perturbing charge is introduced at a particular site in a protein, it may be stabilized by preorganized polar groups (large A_{static}), or by a large reorganization of the environment (large A_{rlx}), or both. In the second case, the environment is polarizable, and it is certainly reasonable to speak of a large local dielectric constant. In the first case, the environment is polar but not very polarizable; to speak of a large dielectric constant in this case is probably more confusing than helpful.

Statements that the definition of the protein dielectric constant depend on the context, or the application, can also be misleading. It is not the definition of the dielectric constant that changes; rather some applications are sensitive to equilibrium properties, and the equilibrium charge distribution; others are sensitive to relaxation properties, or a mixture of equilibrium and relaxation properties. Thus calculations of the equilibrium electric field around a protein, in the absence of any perturbation, are normally done with a protein dielectric constant of one or two.[4,5,8] This is because the equilibrium protein charge distribution is accurately described by the force field employed, along with the average experimental structure of the protein. On the other hand, calculation of the relaxation—or reorganization—free energy in response to a redox electron involves reorganization of the weakly polarizable protein interior, the mobile charged protein side chains, as well as the surrounding solvent. We have just seen that the polarizability of the protein as a whole appears to be best characterized by a rather large dielectric constant. The self-energy of a titrating proton is another example of a property that depends purely on the dielectric relaxation of the protein and solvent.

Finally, calculation of pK_a's, redox potentials, binding constants, and many other properties involves a mixture of the equilibrium and the relaxation properties of the protein. Thus Krishtalik et al. recently proposed that pK_a calculations should be done using two different protein dielectric constants: a low protein dielectric constant of 1−2 to describe the equilibrium field at the titrating site (*i.e.* to calculate A_{static}), and a higher protein dielectric constant to describe the relaxation of the protein in response to the perturbing proton (*i.e* to calculate A_{rlx}).[41] Our results give additional support of this idea, because we find that the protein relaxation properties are characterized by such a large dielectric constant: this makes it all the more important to give a specific treatment to this part of the perturbation. Our results thus suggest, as pointed out earlier,[20] that Poisson−Boltzmann models that treat the protein bulk as a single low-dielectric medium contain a systematic error when relaxation properties of the protein are calculated. This could be corrected at no extra computational cost by introducing a position-dependent dielectric constant within the protein.

The reasonable success of simple continuum models in predicting *e.g.* pK_a shifts due to point mutations, and other properties,[4,5,8] raises the question of how significant this systematic error is. Clearly, charged side chains at the protein surface are strongly shielded by solvent, even when they are modeled as part of a low-dielectric protein medium. Thus continuum models have a certain robustness with respect to the exact definition of the protein−solvent boundary. Nevertheless, simple analytical models, such as a pair of charges near a planar boundary, suggest that modeling the pair of charges as part of the solvent can have a significant effect on the self-energy and

(39) Marcus, R. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155.
(40) Simonson, T.; Perahia, D.; Brünger, A. T. *Biophys. J.* **1991**, *59*, 670−90.
(41) Krishtalik, L.; Kuznetsov, A.; Mertz, E. *Biophys. J.* **1996**, *70*, A225.
(42) Landau, L.; Lifschitz, E. *Statistical Mechanics*; Pergamon Press: New York, 1980.
(43) Marcus, R. *J. Chem. Phys.* **1956**, *24*, 979−989.
(44) Yadav, A.; Jackson, R.; Holbrook, J.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4800−4805.
(45) Warshel, A.; Hwang, J.; Aqvist, J. *Faraday Discuss.* **1992**, *93*, 225.
(46) Warshel, A. *Proc. Natl. Sci. U.S.A.* **1978**, *75*, 5250−5254.
(47) Warshel, A. *Nature* **1987**, *330*, 15−17.

the interaction energy of the pair. Furthermore, a detailed comparison was made recently between the dielectric relaxation properties of cytochrome *c* predicted by molecular dynamics and by simple continuum models.[38] When the protein was modeled as a single low-dielectric medium, the polarizability of the system in response to perturbing test charges near the protein surface was underestimated by about a factor of 2. In pK_a calculations, this would translate directly into a large error in the proton self-energy. Finally, while simple continuum models have had success reproducing experimental pK_a shifts,[4,5,8] it is still unclear whether the agreement is really statistically significant, *i.e* better than a "null" model of zero pK_a shifts.[48]

The need to correctly account for large conformational flexibility of charged groups in Poisson−Boltzmann calculations has of course been recognized before, and various more sophisticated continuum models have been proposed. Some models have placed charged groups in the outer, high-dielectric medium,[49] or introduced explicit solvent molecules into the low-dielectric medium,[48,50] or introduced a position-dependent dielectric constant in the interfacial region.[51] Other models have included explicit averaging over multiple protein conformations, extracted from molecular dynamics simulations.[52] Very recently, methods have been developed to combine molecular dynamics and Poisson−Boltzmann calculations directly, thus allowing conformational averaging over charged side chain conformations to be included naturally into the model.[53−56] If the flexibility of protein charged and polar groups is explicitly included in the calculation through molecular dynamics, then the role of the Poisson−Boltzmann calculation would be limited to the calculation of the solvent polarization. In this case, a low protein dielectric constant of 1−2 would presumably be appropriate for calculating both the equilibrium electric field and the relaxation of the system in response to a perturbation.

A Comprehensive View of the Dielectric Constant of Proteins. In summary, Figure 7 compiles the *G* values for the four proteins studied here, as well as data for ferro- and ferricytochrome *c*, studied by Simonson and Perahia.[20] Results are also summarized in Table 2. The published cytochrome *c* data cannot actually be compared directly to that of the other proteins, since cytochrome *c* was not simulated in bulk water, but rather in a 24 Å water droplet, surrounded by vacuum. However, we can predict the fluctuations in bulk water from the droplet data, using the appropriate Fröhlich−Kirkwood fluctuation formula (analogous to eq 2), and given that we already know the protein dielectric constant. Thus the cytochrome *c* data in Figure 7 are not the raw data from the droplet simulations, but have a bulk-water correction applied to them. In bulk water the protein mean square fluctuations are predicted to increase by about 20% compared to the 24-Å droplet.

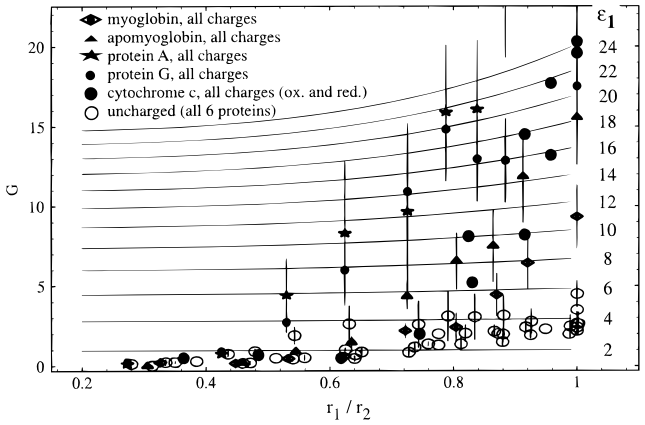

Figure 7. Radial variation of the *G* factor of inner region 1 for myoglobin, apomyoglobin, protein G, protein A, ferricytochrome *c*, and ferrocytochrome *c*. r_1 is measured as a fraction of the total protein radius r_2. Charged groups excluded: circles. All charges: symbols given in legend. The large dots are the ferri- and ferrocytochrome *c* all-charges results, to which a bulk-water correction has been applied (see text). Theoretical curves (solid lines) are labeled on the right by the value of the dielectric constant ϵ_1.

The six proteins present a rather unified behavior, and the salient features seen here are probably general features of many globular proteins. High overall dielectric constants were also calculated for BPTI and lysozyme, compatible with this (although their radial variation and the contribution of charged side chains were not analyzed).[19] The overall dielectric constants found here range from about 11 ± 4 (myoglobin) to 25 ± 10 (cytochrome *c*). The transition between the inner and outer regions is fairly abrupt, and occurs at about the radius where charged surface groups begin to be encountered. The dielectric constant is generally larger for the more flexible structures, as seen for apomyoglobin compared to holomyoglobin. The stable but highly charged cytochrome *c* has a somewhat larger dielectric constant than the other proteins. As pointed out before,[20] the dominance of a few charged side chains concentrated at the protein surface is inconsistent with a homogeneous continuum model for treating protein dielectric relaxation. A more consistent picture is obtained if one views the charged portions of the charged side chains as belonging to the outer, solvent, medium and calculates the dielectric constant of the remainder of the protein. In this picture the calculated dielectric constants are low (3−6 depending on the protein), anisotropy is moderate, and sensitivity to noise and model parameters is reduced. The low dielectric constants obtained here agree with the known low polarizability of the protein bulk. They are close to those of dry powders,[21−23] suggesting that in these powders the polar surface groups have limited mobility and that many or all of the ionizable protein groups are neutral. They are also very close to the local polarizabilities estimated for the interior of the photosynthetic reaction center, based on absorption shifts of probe chromophores.[13]

Acknowledgment. C.L.B. wishes to acknowledge partial support for this work from the National Institutes of Health (GM48807). T.S. acknowledges support from the Centre National de la Recherche Scientifique.

JA960884F

(48) Antosiewicz, J.; McCammon, J.; Gilson, M. *J. Mol. Biol.* **1994**, *238*, 415−436.
(49) Delepierre, M.; Dobson, C.; Karplus, M.; Poulsen, F.; States, D.; Wedin, R. *J. Mol. Biol.* **1987**, *197*, 111−130.
(50) Chan, S.; Lim, C. *J. Phys. Chem.* **1994**, *98*, 692−695.
(51) Pack, G.; Garrett, G.; Wong, L.; Lamm, G. *Prot. Eng.* **1993**, *2*, 177−183.
(52) Northrup, S.; Wensel, T.; Meares, C.; Wendolowski, J.; Matthew, J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9503−9507.
(53) Sharp, K. *J. Comp. Chem.* **1991**, *12*, 454−468.
(54) Gilson, M.; Davis, M.; Luty, B.; McCammon, J. *J. Phys. Chem.* **1993**, *97*, 3591−3600.
(55) Juffer, A.; Berendsen, H. *Mol. Phys.* In Press.
(56) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.