Microscopic Dielectric Properties of Cytochrome *c* from Molecular Dynamics Simulations in Aqueous Solution

Thomas Simonson*,† and David Perahia‡

Contribution from the Laboratoire de Biologie Structurale (CNRS), IGBMC, BP 163, C.U. de Strasbourg, 67404 Illkirch, France, and Laboratoire d'Enzymologie Physico-Chimique et Moléculaire, Université de Paris-Sud, 91405 Orsay, France

Received January 31, 1995⊗

Abstract: The dielectric properties of proteins play an important role in biological charge transfer and catalysis. We investigate microscopic charge screening by the electron transfer protein cytochrome *c* in solution, using linear response and thermodynamic perturbation theory. Electronic relaxation is described using atomic point polarizabilities. Dipolar relaxation is described using a 1 ns molecular dynamics simulation of the protein in solution. Dielectric relaxation in response to perturbing charges is calculated from simulations of a single, unperturbed, reference state. This technique allows us to study the microscopic dielectric properties throughout the entire protein interior using a single simulation. We calculate relaxation free energies in response to a single test charge, located successively on each C_α of the protein backbone. For small test charges, these energies are given by the variance of the reference electrostatic potential at the test charge site. Protein and solvent contributions are nearly equal, and the total relaxation free energy is much smaller than either, due to protein−solvent coupling. The fluctuations of the reference electrostatic potential are approximately Gaussian, leading to a nearly linear dielectric response for perturbing charge magnitudes of $\leq e/4$ and relaxation free energies of ≤4 kcal/mol. The electronic contribution to the relaxation free energies is spatially homogeneous and can be fit by a continuum model, with a dielectric constant of 2. The dipolar contribution is 1.5−2 times larger, is less homogeneous, and is fit only moderately well by a continuum model, with a dielectric constant of 4. The relaxation free energies increase smoothly by a factor of 2 when the test charge is moved from the protein center to its surface. The heme center is in a region where the relaxation is minimal. This correlates directly with the biological requirement to reduce the reorganization free energy for electron transfer to and from the heme.

1. Introduction

Dielectric properties of proteins play a crucial role in their structure and activity.[1,2] The complexity of protein structures and their solution environment make the analysis of charged and polar interactions difficult, and they have been the subject of intense experimental and theoretical work. Dielectric relaxation of bulk protein powders has been studied for many years.[3] Measurements of pK_a's of ionizable groups,[4] and changes in stability[5] or redox potential[6−8] upon mutating charged or polar groups, have been abundantly used to probe charge interactions and screening. Energy transfer[9] and fluorescence spectra[10] are sensitive to the dielectric properties around the excitable group. Internal Stark effect measurements were used recently to measure the electric field of an α helical peptide[11] and to analyze microscopic dielectric properties of the photosynthetic reaction center.[12] Theoretical studies of protein electrostatics go back to Linderstrom−Lang[13] and currently fill a large number of books. They include many continuum and molecular dynamics studies,[14−17] as well as recent calculations aimed specifically at the macroscopic dielectric constant of several proteins.[18−24]

All these studies show that dielectric screening is an important part of protein energetics, influencing structure, stability, binding, and activity. The low polarizability of the protein interior for example excludes net charges[25] and drives polar

† IGBMC.
‡ Université de Paris-Sud.
* To whom correspondence should be addressed.
⊗ Abstract published in *Advance ACS Abstracts*, July 15, 1995.

(1) Perutz, M. *Science* **1978**, *201*, 1187−1191.
(2) Warshel, A.; Russell, S. *Q. Rev. Biophys.* **1984**, *17*, 283−342.
(3) Bone, S.; Pethig, R. *J. Mol. Biol.* **1985**, *181*, 323−326.
(4) Russell, A.; Fersht, A. *Nature* **1987**, *328*, 496−500.
(5) Nicholson, H.; Becktel, W.; Matthews, B. *Nature* **1988**, *336*, 651−656.
(6) Rogers, N.; Moore, G.; Sternberg, M. *J. Mol. Biol.* **1985**, *182*, 613−616.
(7) Moore, G.; Pettigrew, G.; Rogers, N. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4998−4999.
(8) Varadarajan, R.; Zewert, T.; Gray, H.; Boxer, S. *Science* **1989**, *243*, 69−72.
(9) Northrup, S.; Wensel, T.; Meares, C.; Wendolowski, J.; Matthew, J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9503−9507.
(10) Pierce, D.; Boxer, S. *J. Phys. Chem.* **1992**, *96*, 5560−5566.
(11) Lockhardt, D.; Kim, P. *Science* **1992**, *257*, 947−951.
(12) Steffen, M.; Lao, K.; Boxer, S. *Science* **1994**, *264*, 810−816.
(13) Linderstrøm-Lang, K. *C. R. Trav. Lab. Carlsberg* **1924**, *15*, 1−2.
(14) Rogers, N. *Prog. Biophys. Mol. Biol.* **1986**, *48*, 37−66.
(15) Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *19*, 301−332.
(16) Brooks, C.; Karplus, M.; Pettitt, M. *Adv. Chem. Phys.* **1987**, *71*, 1−259.
(17) Warshel, A. *Computer modelling of chemical reactions in enzymes and solutions*; John Wiley: New York, 1991.
(18) Gilson, M.; Honig, B. *Biopolymers* **1986**, *25*, 2097−2119.
(19) Nakamura, H.; Sakamoto, T.; Wada, A. *Protein Eng.* **1989**, *2*, 177−183.
(20) King, G.; Lee, F.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366−4377.
(21) Smith, P.; Brunne, R.; Mark, A.; van Gunsteren, W. *J. Phys. Chem.* **1993**, *97*, 2009−2014.
(22) Simonson, T.; Perahia, D.; Bricogne, G. *J. Mol. Biol.* **1991**, *218*, 859−886.
(23) Simonson, T.; Perahia, D.; Brünger, A. T. *Biophys. J.* **1991**, *59*, 670−90.
(24) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082−1086.
(25) Stites, W.; Gittis, A.; Lattman, E.; Shortle, D. *J. Mol. Biol.* **1991**, *221*, 7−14.

0002-7863/95/1517-7987$09.00/0

groups into specific hydrogen bonding networks.[26] Charge transfer or charge separation steps in enzyme reactions generally require a polar but relatively nonpolarizable environment for optimal kinetics.[27,28] Thus specific local dielectric properties are important. The sensitivity of redox potentials to point mutations depends on the degree of dielectric screening of the partial charges on the modified residues.[6] Finally the kinetics of electron and proton transfer are directly related to the structural relaxation in response to the transferred charge, through the relaxation, or reorganization, free energy.[29] Thus dielectric asymmetry in the photosynthetic reaction center appears to be functionally important.[12] While some of these properties can be interpreted qualitatively using simple continuum models,[28,30] the details of microscopic charge screening are obviously important.

Microscopic simulations have been used both to calculate average protein dielectric constants and to probe local dielectric properties, such as the structural reorganization around specific functional sites upon electron or proton transfer.[31-36] The former simulations give an average, macroscopic description of the dielectric behavior, while the latter are focused on a single site. These last have given detailed information on activation free energies and relaxation free energies for charge transfer and the importance of polar, relatively nonpolarizable groups in the active site. The present work, in contrast, is a comprehensive study of the microscopic dielectric properties of the redox protein cytochrome *c*, not only in the vicinity of its functional heme but throughout the entire protein interior. We analyze systematically the response of the protein and the surrounding solvent to perturbing test charges as a function of their position. The individual components of the dielectric relaxation are compared: electronic polarization (described using atomic point polarizabilities), atomic and dipolar polarization of the protein and solvent (described using molecular dynamics), as well as coupling between them. Dielectric saturation is also analyzed. We pay particular attention to the spatial variation of the dielectric properties. Indeed, the dielectric properties are calculated from simulations of an unperturbed, reference state, using free energy perturbation methods along with a linear response approximation. This technique enables us to study an arbitrary number of test charge locations using a single reference simulation. It does require a sufficiently long simulation and a careful analysis of the range of validity of the perturbation method. We shall see that the microscopic dielectric properties vary markedly within cytochrome *c*, in a way that may be important for the protein's function. A continuum model is studied in comparison to the microscopic models. This can help to clarify the nature of the protein dielectric constant.

Cytochrome *c* has served for many years as a model for redox and electron transfer processes in biology, and several simulation studies have already been carried out. Molecular dynamics simulations of moderate length (100 ps) in bulk water gave valuable structural and dynamic information but were too short to extract dielectric properties.[37] Other simulations specifically addressed reorganization associated with electron transfer to and from the heme[31,32] and showed that the reorganization energy is much lower in the protein than for a heme in bulk water. However, these studies were based on a simplified, purely electrostatic model (the protein dipole Langevin dipole model) which cannot give the same detailed microscopic information as a full molecular dynamics model. In the present work, we exploit a nanosecond molecular dynamics simulation in explicit water to study the dielectric properties not only in the heme region but throughout the entire protein. In earlier articles, Simonson et al. applied a similar analysis to cytochrome *c* in vacuo,[22,23] while more recently, we calculated the macroscopic dielectric constant of cytochrome *c* in solution, using Fröhlich–Kirkwood theory.[24]

The next section recalls some theoretical background. The third section describes computational details of the simulations and the analysis. The fourth section describes numerical results, based on a 1 ns molecular dynamics simulation of yeast ferricytochrome *c* in a spherical droplet of 1400 explicit waters. The fifth section is a discussion.

2. Relaxation Free Energy and Generalized Susceptibility in Response to a Static Perturbation

2.1. General Theory. Consider a set of "parent" particles such as a folded protein or a liquid and a set of fixed perturbing particles. We focus below on perturbing charges, which interact with the parent particles through a Coulomb potential. However, most of what follows can be applied to other types of interaction and, thus, has implications for other types of free energy calculations.

The perturbation free energy A_{tot} contains a "static" term A_{static} and a relaxation term A_{rlx}:

$$A_{tot} = A_{static} + A_{rlx} \tag{1}$$

The first term is the work to introduce the perturbation while constraining the parent system to retain its unperturbed structure:

$$A_{static} = \langle U_{tot} \rangle_0 \tag{2}$$

where U_{tot} is the (classical) perturbing Hamiltonian and the brackets represent an ensemble average in the absence of the perturbing particles. The second term is associated with the structural relaxation after the constraints are removed and can be obtained from thermodynamic perturbation theory:[38,39]

$$\begin{aligned} A_{rlx} &= -kT \ln\langle \exp(-U_{tot}/kT) \rangle_0 - \langle U_{tot} \rangle_0 \\ &= -kT \ln\langle \exp[-(U_{tot} - \langle U_{tot} \rangle_0)/kT] \rangle_0 \end{aligned} \tag{3}$$

where k is Boltzmann's constant and T is the temperature. In charge transfer theory this term is known as the reorganization energy.[40] It can always be expanded with respect to U_{tot}, giving

$$A_{rlx} = -\frac{1}{2kT}(\langle U_{tot}^2 \rangle_0 - \langle U_{tot} \rangle_0^2) + O([(U_{tot} - \langle U_{tot} \rangle_0)/kT]^3) \tag{4}$$

For small perturbations, *i.e.* in the linear response limit, A_{rlx} is given by the variance of U_{tot}. Terms of order 3 and higher

(26) Baker, E.; Hubbard, R. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97–117.
(27) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–224.
(28) Krishtalik, L. *J. Theor. Biol.* **1985**, *116*, 201–214.
(29) Moser, C.; Keske, J.; Warncke, K.; Farid, R.; Dutton, P. *Nature* **1992**, *355*, 796–802.
(30) Tanford, C.; Kirkwood, J. *J. Am. Chem. Soc.* **1957**, *79*, 5333.
(31) Churg, A.; Weiss, R.; Warshel, A.; Takano, T. *J. Phys. Chem.* **1983**, *87*, 1683–1694.
(32) Churg, A.; Warshel, A. *Biochemistry* **1986**, *25*, 1675–1681.
(33) Warshel, A.; Chu, Z.; Parson, W. *Science* **1989**, *246*, 112–116.
(34) Nonella, M.; Schulten, K. *J. Phys. Chem.* **1991**, *95*, 2059–2067.
(35) Warshel, A.; Hwang, J.; Aqvist, J. *Faraday Discuss.* **1992**, *93*, 225.
(36) Yadav, A.; Jackson, R.; Holbrook, J.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4800–4805.
(37) Wong, C.; Zheng, C.; Shen, J.; McCammon, J.; Wolynes, P. *J. Phys. Chem.* **1993**, *97*, 3100–3110.
(38) Landau, L.; Lifschitz, E. *Statistical Mechanics*; Pergamon Press: New York, 1980.
(39) McQuarrie, D. *Statistical Mechanics*; Harper and Row: New York, 1975.
(40) Marcus, R. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155.

correspond to a nonlinear response of the system and, therefore, measure dielectric saturation. If the perturbing potential U_{tot} follows a Gaussian distribution, all these higher order terms are identically zero, as pointed out by Levy et al.[41] (Indeed, for a normally distributed random variable X, the expectation value of $\exp(X)$ is $\exp(\mu + \sigma^2/2)$, where μ and σ are the mean and standard deviation of X. By applying this to U_{tot}/kT and inserting into (3), A_{rlx} reduces to the second-order expression.) Thus linear response, with its harmonic free energy surface, corresponds to Gaussian fluctuations of U_{tot}. Dielectric saturation is related to the deviation of U_{tot} from a Gaussian distribution. Physically, a large variance of U_{tot} reflects the electrostatic "softness" of the parent system at the perturbing charge locations. A Gaussian distribution of U_{tot} could presumably arise from the independent contributions of many parent polar groups. Deviations from the Gaussian distribution could arise from correlations between polar groups, which in turn would limit the ability of these groups to relax independently, introducing a degree of frustration.

In this article we will be focusing on a protein in solution, interacting with a single perturbing test charge q. For this case, $U_{tot} = qV_{tot}$, where V_{tot} is the electrostatic potential of the parent system on the test charge. From (4), A_{rlx} can be written as

$$A_{rlx} = -\frac{q^2}{2kT}[(\langle V_p^2\rangle_0 - \langle V_p\rangle_0^2) + (\langle V_w\rangle_0^2 - \langle V_w\rangle_0^2) + 2(\langle V_pV_w\rangle_0 - \langle V_p\rangle_0\langle V_w\rangle_0)] = A_p + A_w + A_{pw} \quad (5)$$

V_p and V_w are the protein and solvent contributions to V_{tot}. The complete relaxation free energy is thus a sum of a protein contribution A_p, a solvent contribution A_w, and a cross-term A_{pw} due to protein−solvent correlations.

An important special case arises when two conditions are fulfilled: (1) the parent particles are localized at definite mean positions, as in a folded protein *in vacuo* or a crystal of small molecules; (2) the motions of the parent particles are small compared to the parent particle−perturbing particle distances. We shall refer to this special case as the *nondiffusive* limit, since it corresponds to localized particles, in contrast to the diffusive motions occurring in liquids. In this limit the relaxation free energy takes a discrete form

$$A_{rlx} = -{}^1/_2{}^t f\alpha f \quad (6)$$

where f is the $3n$ vector

$$f = (f_1, f_2, ..., f_n) \quad (7)$$

f_i is the perturbing electrostatic field at the mean position of atom i, and n is the number of atoms in the system. The raised t represents vector transposition. α is a $3n \times 3n$, real-space, susceptibility operator.[38] It is a special case of the usual dielectric susceptibility functional,[42] taking into account the nondiffusive character of the system. We showed earlier that it has the approximate analytical form[23]

$$\alpha \approx \frac{1}{kT}\mathbf{M} \quad (8)$$

where $\mathbf{M}$ is the dipole−dipole correlation matrix of the unperturbed system

$$\mathbf{M}_{i\alpha,j\beta} = q_iq_j\langle\delta u_i^{\alpha}\delta u_j^{\beta}\rangle_0 \quad (9)$$

q_i is the partial charge on atom i, δu_i^{α} is its instantaneous displacement from its average position along the Cartesian axis α (= x, y, or z), and the brackets represent an ensemble average. The analytical form (8) of the susceptibility operator can be generalized easily to the case where the perturbing particles interact with the parent particles through a broader class of pair potentials, including shifted or truncated Coulomb potentials (T. Simonson, unpublished).

In the nondiffusive limit, rather than calculating the entire susceptibility matrix, a simpler characterization of the dielectric properties can be obtained by considering a single perturbing test charge q and calculating the relaxation free energy as a function of its position. Dividing the relaxation free energy by the square of the perturbing field gives a *scalar* susceptibility $\alpha(r_q)$:

$$A_{rlx} = -{}^1/_2\alpha(r_q)f^2 \quad (10)$$

The scalar susceptibility is a fuction of the position r_q of the test charge but not of its magnitude. It represents a one-dimensional contraction of the full susceptibility operator.

For a system such as a protein in solution, the nondiffusive limit does not apply. We cannot write the real-space susceptibility operator as a matrix; rather it takes the form of a linear functional. In this case it is difficult to define a physically meaningful scalar susceptibility, and we shall focus below on the relaxation free energies. It will be useful however to convert these to "susceptibility units", by dividing them by an average value of f^2, to be defined below.

In this paper, we focus on electrostatic perturbations and dielectric relaxation. However, the analysis developed in this section applies equally well to other pair potentials, including the Lennard-Jones potential. Structural relaxation in response to other perturbations, such as chemical changes, can be characterized by a relaxation free energy and a susceptibility operator. This type of analysis is related to the generalized Green's function formalism of Wong et al.[37]

2.2. Relation to Free Energy Perturbation Calculations and Charge Transfer Studies. Free energy calculations using the so-called thermodynamic perturbation or thermodynamic integration approaches[43] are based on extrapolation of a perturbation free energy from simulations of an unperturbed reference state. Either the free energy is calculated directly, by averaging $\exp(-U_{tot}/kT)$, or the derivative of the free energy is calculated with respect to the amplitude of the perturbation. The amplitude is measured by a coupling parameter λ, and the derivative is given by $\partial A/\partial\lambda = \langle\partial U_{tot}/\partial\lambda\rangle$. The derivative is then integrated to give the free energy change. For large perturbations, extrapolation from a single reference state can be inaccurate. The coupling parameter is then increased in several discrete steps, and a free energy increment is calculated at each step. Thus a multiple reference state method is used.

The linear response approximation can be viewed as a single reference state thermodynamic integration method, which uses the first and second free energy derivatives. The perturbation consists of introducing a set of point charges $\{q_i\}$ at fixed

(41) Levy, R.; Belhadj, M.; Kitchen, D. *J. Chem. Phys.* **1991**, *95*, 3627−3633.

(42) Madden, P.; Kivelson, D. *Adv. Chem. Phys.* **1984**, *54*, 467−530.

(43) Beveridge, D.; DiCapua, F. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.

locations. We define the coupling parameter λ by

$$U_{tot}(\lambda) = \lambda \sum_i q_i V_i = \lambda U_{tot}(1) \quad (11)$$

V_i being the parent electrostatic potential on the perturbing charge q_i. The perturbation is introduced by increasing λ from zero to one. The first two derivatives of the free energy are

$$\partial A/\partial\lambda(\lambda) = \langle U_{tot}(1)\rangle_\lambda \quad (12)$$

$$\partial^2 A/\partial\lambda^2(\lambda) = -(1/kT)(\langle U_{tot}(1)^2\rangle_\lambda - \langle U_{tot}(1)\rangle_\lambda^2) \quad (13)$$

If we start from the unperturbed system ($\lambda = 0$) and introduce the perturbing charges in a single step, then the first thermodynamic derivative $\partial A/\partial\lambda(\lambda = 0)$ is equal to the static part of the perturbation free energy. The second derivative is equal to the second-order, linear response, part of the dielectric relaxation free energy. Higher derivatives would give the deviation from linear response.

Our ability to extrapolate the dielectric response from a single unperturbed reference state will decrease as the magnitude of the perturbing charges increases. The higher order derivatives become important, and while they are easy to calculate analytically, they rapidly become very difficult to estimate numerically. Simulations of the unperturbed reference state do not provide sufficient sampling of all the conformations that contribute to these derivatives. This problem is analyzed in detail further on (Figure 6). For this reason, large perturbing charges are usually introduced in several discrete steps. The dielectric relaxation is thus explicitly simulated, including saturation effects, as the perturbing charges are introduced.

Free energy perturbation methods have been used recently to study electron and proton transfer in liquids[44-47] and in proteins.[33-35] The perturbation consists in the displacement of a charge from a donor to an acceptor site and is usually carried out in several steps. The charge can be either gradually moved from one site to the other or else deleted at the donor site and recreated at the acceptor site. In the classic Marcus theory,[40] the relaxation free energy A_{rlx} for charge transfer is known as the reorganization energy and is closely related to the activation free energy. The studies just mentioned found that the free energy as a function of charge transfer was approximately harmonic over a broad range, indicating that the linear response approximation is reasonably accurate. Thus a single reference state calculation would presumably have been feasible. However, a multistep calculation based on a series of short simulations is probably more efficient computationally, and more accurate, when a single charge transfer process is being studied. In the present article, in contrast, we investigate a large number of different perturbations (a test charge placed successively on each C_α), so that extrapolation of the free energy from a single reference state is very advantageous.

2.3. Application to a Set of Fixed, Polarizable Particles. We now consider the special case where the parent particles are fixed but bear atomic point polarizabilities. This model is commonly used in biomolecular simulations to represent the electronic polarizability of the system,[27] as we shall do below. When the perturbation is introduced, dielectric relaxation corresponds to a change δm_i in the induced dipoles on each particle i. The relaxation energy turns out to be[77]

$$\langle U_{rlx}\rangle = -\sum_i \delta m_i f_i \quad (14)$$

The relaxation free energy is $\langle U_{rlx}\rangle/2$;[23] thus

$$A_{rlx} = -{}^1/_2 \sum_i \delta m_i f_i \quad (15)$$

The quantity conjugate to the perturbing field is just the list of induced dipole shifts, $\{\delta m_i\}$.

If the parent particles are mobile, and simultaneously carry a point polarizability, then electronic polarization and atomic and dipolar polarization (polarization due to atomic motions) coexist. A coupling then arises between the two. An analytical expression is available for the coupling energy,[22] in the form of a series expansion in powers of the dipole−dipole tensor of the system. Because this tensor is small, the coupling represents a corrective effect compared to the electronic and dipolar contributions.

2.4. Application to a Macroscopic Dielectric Continuum. It is instructive and practically important to consider the special case where the "parent" system is actually a dielectric continuum, with fixed partial charges embedded at discrete positions. This model is commonly used to model protein electrostatics.[14,15] By comparing it to the molecular dynamics results, we hope to clarify the nature of the protein dielectric constant.

When the perturbation is introduced, the parent particles remain fixed, but dielectric relaxation occurs by rearrangement of induced volume and surface charge.[22] If the perturbing charge is outside the parent medium, only induced surface charge contributes. Let $\delta\sigma_{ind}$ be the shift in the induced surface charge due to the perturbation. Then

$$A_{rlx} = {}^1/_2 \oint \delta\sigma_{ind}\, V_{pert}\, d^2r = {}^1/_2 \int \varrho(r)\, V(\delta\sigma_{ind})\, d^3r \quad (16)$$

The first integral is over the dielectric boundary, and V_{pert} is the perturbing electrostatic potential at the surface element d^2r.[22b] The second integral is over all space, ϱ is the perturbing charge density, and $V(\delta\sigma_{ind})$ is the potential produced by the induced surface charge shift at r. If the perturbing charge is inside the parent dielectric medium, there is an additional contribution—the Born free energy to transfer the perturbing charge from vacuum to the parent medium. From (16), A_{rlx} is identical to the free energy to introduce the perturbing charge with all the parent charges set to zero. Thus the dielectric relaxation properties are a function only of the shape and dielectric constant of the parent system, but not of its distribution of fixed partial charges. In biochemical applications, the free energy to insert a charge into a protein with the protein permanent charges removed is sometimes referred to as the "self-energy" of the charge and plays a role in titration calculations.[15] We see that the "self-energy" of a titrating proton in a protein is equal to the relaxation free energy for inserting the proton.

3. Computational Details

3.1. The Simulation. Yeast ferri-cytochrome *c* was simulated in a spherical droplet of 1400 TIP3P water molecules,[48] starting from the crystal structure.[49] Electrostatic interactions were approximated by a Coulomb term shifted to zero beyond 12 Å.[50] The Charmm/Param19 empirical force field was used.[50]

(44) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218−2224.
(45) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.
(46) Kuharski, R.; Bader, J.; Chandler, D.; Sprik, M.; Klein, M.; Impey, R. *J. Chem. Phys.* **1988**, *89*, 3248−3257.
(47) Carter, E.; Hynes, J. *J. Phys. Chem.* **1989**, *93*, 2184.
(48) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(49) Berghuis, A. M.; Brayer, G. D. *J. Mol. Biol.* **1992**, *223*, 959−976.
(50) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

Arg, Lys, Asp, Glu, and His side chains were fully charged (except for His 18, bound to the heme iron, and His26, thought to be neutral[51]). The crystallographic sulfate ion was included; otherwise no counterions were used. Bond lengths were constrained with the SHAKE algorithm.[52] A soft spherical boundary potential[53] of radius 24 Å was used to confine the system. The system was weakly coupled to a heat bath at 293 K.[54] A 1 ns simulation was run, and the last 900 ps used for analysis, sampled every 0.5 ps. Some portions of the analysis were repeated with a finer-grained sampling (every 0.1 ps), giving nearly identical results. Simulations were done with the program X-PLOR.[55]

3.2. Relaxation Free Energies and Susceptibilities. We consider a single perturbing test charge, located successively on each C_α of the protein backbone. The atomic and dipolar relaxation is modeled using the molecular dynamics simulation. The relaxation free energies are calculated from the simulation using (4). The protein, solvent, coupling, and total contributions are all calculated: A_p, A_w, A_{pw}, and A_{rlx}, as defined in (5). We report the relaxation free energies both in energy units (kcal/mol) and in susceptibility units (Å^3). We convert the relaxation free energies to susceptibility units by dividing them by the constant factor $-1/2\langle\sum_i f_i^2\rangle$, where f_i is the perturbing field on protein atom i, the sum is over all protein atoms, and the brackets represent an average over all test charge positions (all Cα's). In other words, for each test charge position we report both A_{rlx} (kcal/mol) and the quantity

$$A'_{rlx} = -2A_{rlx}/\langle\sum_i f_i^2\rangle \quad (\text{Å}^3) \tag{17}$$

For comparison, recall that the scalar susceptibility of a nondiffusive system (eq 10) was defined as

$$\alpha(r_q) = -2A_{rlx}/\sum_i f_i^2 \quad (\text{Å}^3)$$

This change of sign and units makes the relaxation free energies independent of the magnitude of the test charge and easier to compare to the scalar susceptibilities, calculated earlier for cytochrome *c* and other proteins in vacuo.[22,23,56]

Since the simulation was performed with a 12 Å cutoff, the relaxation free energies are calculated with the same cutoff for consistency. This means that the test charge is effectively inserted into a medium with no interactions beyond 12 Å. A simple correction to account for the remaining interactions is the Born correction:

$$\Delta A_{Born} = \frac{q^2}{2R_{cut}}(1/\epsilon - 1) \tag{18}$$

where R_{cut} is the cutoff distance (12 Å) and ϵ is the dielectric constant of the medium surrounding the test charge. The material beyond 12 Å from the test charge is partly water, partly protein. Thus ϵ lies between about 4 and 80, $1/\epsilon - 1$ lies between -1 and -0.75, and for $q = e$, ΔA_{Born} lies between -10.5 and -14 kcal/mol. The Born correction should be included in the relaxation part of the free energy. Indeed, in the continuum model upon which it is based, the relaxation free energy is the free energy to insert the test charge with all permanent charges removed. The Born contribution is a part of this free energy. Therefore the cutoff correction to the relaxation free energy, to this level of approximation, lies between about -10 and -14 kcal/mol for a unit test charge.

3.3. Susceptibilities of a Collection of Polarizable Particles. To model electronic polarizability, we shall view the protein as a collection of fixed, polarizable, particles. The shift in induced dipoles due to a perturbing test charge is calculated iteratively as before.[22] This procedure is equivalent to summing the matrix series $\sum_{i=0}^{\infty}(\alpha_{at}\boldsymbol{T})^i$, where α_{at} is the diagonal matrix of order $3n$ formed by the list of atomic polarizabilities (each repeated three times) and $\boldsymbol{T}$ is the dipole−dipole matrix of the system. This sum converges if $\alpha_{at}\boldsymbol{T}$ has no eigenvalues greater than 1 in absolute magnitude. The series will usually diverge if interactions between nearby atoms are included in $\boldsymbol{T}$ with no special treatment. To avoid this problem, we used a damped dipole−dipole interaction at short range,[57] and we completely eliminated interactions between bonded pairs of atoms (covalent first neighbors). Numerical values for the atomic polarizabilities were taken from ref 57. The X-ray coordinates of the protein were used. A total of 10−30 iterations ensured good convergence of the induced dipoles.

3.4. Microscopic Susceptibilities of a Macroscopic Dielectric Continuum. A different approach is to view the protein as a macroscopic dielectric continuum and to calculate the relaxation free energy and susceptibility in response to one or more test charges from (16). In practice, we calculate the free energy to introduce the test charge(s), with all the permanent charges of the protein removed. The protocol depends on whether the test charge is to be placed inside or outside the protein. For a test charge inside the protein, we use the thermodynamic cycle depicted in Figure 1. The relaxation free energy A_{rlx} (step 1→4) is obtained in three steps. Step 1→2 has a zero free energy. Step 2→3 is the transfer of the test charge from vacuum into a uniform ϵ_p medium. The free energy change is just the Born self-energy, for which an assumption about the test charge radius is needed. Finally step 3→4 requires solving the Poisson equation, first with a uniform dielectric constant ϵ_p, then with a dielectric constant of ϵ_p inside the protein and one outside. The Poisson equation was solved with the program Delphi[58] on a cubic grid using a finite-difference algorithm. A cubic grid with 65 nodes per side was used, and a multigrid calculation with two focusing steps was performed.[59] The final grid spacing was 0.75 Å. Results for several grid orientations were averaged. The dielectric boundary was defined by the solvent accessible surface of the protein, with a solvent probe radius of 1.4 Å. The radii of the protein atoms were taken from the Charmm19 force field,[50] and the X-ray coordinates were used.

It is useful to define a protein contribution (step 2→6) and a solvent contribution (step 2→5) to A_{rlx}. The sum of these contributions differs from the total relaxation free energy; this difference defines a protein−solvent coupling contribution.

4. Results

4.1. Electronic Susceptibilities. We first view the protein atoms as fixed polarizable particles. The point polarizabilities give an approximate description of the electronic contribution

(51) Hartshorn; Moore, G. *Biochem. J.* **1988**, *258*, 595−598.
(52) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327−341.
(53) Brooks, C.; Brünger, A. T.; Karplus, M. *Biopolymers* **1985**, *24*, 843−865.
(54) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *811*, 3684−3690.
(55) Brünger, A. T. X-PLOR *version 3.1, A System for X-ray crystallography and NMR*; Yale University Press: New Haven, CT, 1992.
(56) Simonson, T.; Perahia, D. *Biophys. J.* **1992**, *61*, 410−427.
(57) Thole, B. *Chem. Phys.* **1981**, *59*, 341−350.
(58) Sharp, K. *DelPhi, Version 3.0*; Columbia University: New York, 1988.
(59) Gilson, M.; Sharp, K.; Honig, B. *J. Comput. Chem.* **1988**, *9*, 327−335.

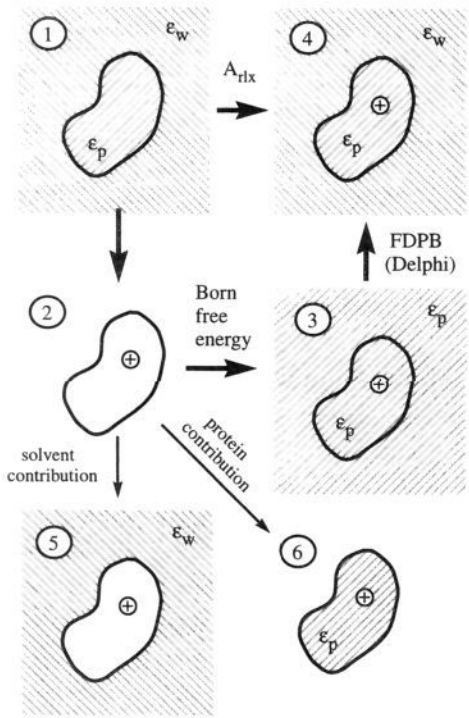

Figure 1. Thermodynamic cycle to calculate the relaxation free energy A_{rlx} from the continuum model. Test charge insertion is done in steps 1→2→3→4. The free energy $\Delta A(1 \rightarrow 2)$ is zero; $\Delta A(2 \rightarrow 3)$ is the Born free energy; $\Delta A(3 \rightarrow 4)$ is obtained from a finite-difference Poisson Boltzmann calculation; $\Delta A(2 \rightarrow 5)$ is defined as the solvent contribution to the relaxation free energy; $\Delta A(2 \rightarrow 6)$ is the protein contribution.

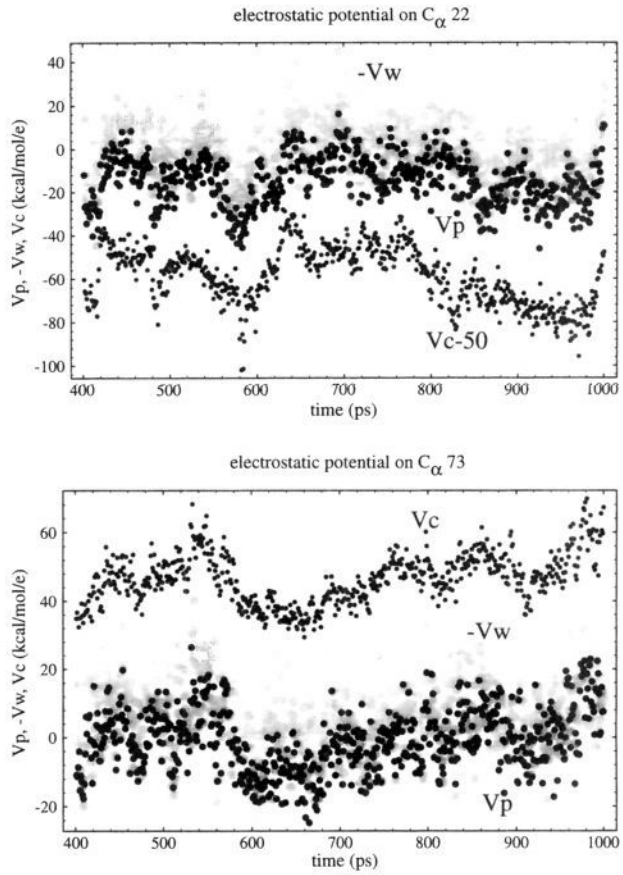

Figure 2. (top) Electrostatic potential on C_α of Lys-22 from molecular dynamics simulation as a function of time. Protein contribution V_p (black), solvent contribution V_w (grey), and contribution from charged protein side chains V_c (offset by −50 kcal/(mol•e) for clarity). (bottom) Idem for C_α of Lys-73.

to the dielectric properties. We consider a test charge $q = e$ located successively on each C_α. Since this model contains no solvent, the system is nondiffusive and we can calculate the scalar susceptibility (eq 10) as well as the relaxation free energies. The scalar susceptibility in response to the test charge is found to be virtually uniform throughout the protein. The mean value (averaged over the C_α's) is 0.97 Å^3; the standard deviation over the C_α chain is only 0.13 Å^3. The mean relaxation free energy is −25.4 kcal/mol; the standard deviation along the chain is 3.1 kcal/mol. In "susceptibility units" (eq 17), the mean relaxation free energy is $\langle A'_{rlx} \rangle = 0.94$ Å^3 (standard deviation of 0.11 Å^3). This is very close to the mean susceptibility. Recall that the susceptibility is given by $-2A_{rlx}/f^2$, whereas A'_{rlx} is given by $-2A_{rlx}/\langle f^2 \rangle$, *i.e.* we divide by the *average* squared field. Because f^2 depends weakly on the test charge position, the two are only slightly different.

Very similar results were obtained earlier[22] using a different set of polarizabilities and a slightly different treatment of short-range dipole interactions.

4.2. Electrostatic Potential from Molecular Dynamics Simulation: Protein−Solvent Complementarity. We now turn to the electrostatic and dielectric properties calculated from the molecular dynamics simulation. In this section we describe the electrostatic potential on the C_α's of the protein backbone.

Analysis of the potential on the C_α's yields several results. First, the protein contribution V_p and the solvent contribution V_w tend to be highly anti-correlated. Thus the solvent potential is roughly opposite to the protein potential, reflecting the strong solvent screening of the protein charges. The mean correlation coefficient of V_p and V_w, averaged over all 108 C_α's, is −0.72. Correlation coefficients for particular C_α positions vary from −0.50 to −0.94—with the exception of the C_α of Glu-4, for which the correlation is only −0.35. The stronger correlations are seen toward the protein surface. Inspection of the time series of V_p and V_w at typical positions (Figure 2) shows that V_w follows very closely the time variations of V_p, with a somewhat weaker amplitude (about 30% weaker on average). Time lag between the two potentials is on the scale of a few picoseconds at most. Thus the solvent does not merely screen V_p on average but does so on a picosecond time scale. The self-correlations of V_p and V_w, and the cross-correlations between the two, decay exponentially for short times (<40 ps), with a decay time of 20−60 ps depending on the C_α position.

The protein contribution V_p can be attributed mainly to the 33 charged side chains, Arg, Asp, Glu, Lys, and His. The correlation between the charged side chain contribution and V_p, averaged over time and over all C_α positions, is 0.66. Typical time series are shown in Figure 2. The correlation coefficients at individual C_α positions cover a wide range, from −0.1 to 0.9. Several surface residue C_α's show some of the weakest correlations.

On the other hand, the solvent contribution V_w to the total potential cannot be attributed to a simple subset of water molecules. For example there is no correlation between V_w and the potential produced by the first solvation shells of the charged protein residues.

In contrast to the strong protein−solvent correlations, correlations among the 33 charged protein side chain groups are mostly weak. The largest correlation coefficients between charged side chain motions are ±0.4, for just seven pairs of residues. Three of these pairs are oppositely charged residues

involved in salt bridges (Glu 88-Arg 91, Lys 89-Asp90, Lys 89-Asp 93). Two of the other pairs involve residues more than 30 Å apart. Inspection of the charged side chain motions shows that several of them undergo large, infrequent, reorientations, so that some correlation coefficients have probably not fully converged despite the nanosecond simulation length. Another measure of side chain correlations is the variance of the total dipole moment produced by the charged side chains, as discussed in ref 24. This quantity determines the average dielectric constant of the protein. It takes the form of a pairwise sum over self- and cross-correlations between the charged side chains. It is equal to 113 (e Å)^2, of which only 13 (e Å)^2 is due to correlations between different side chains. The remainder is due to self-correlations of each charged side chain with itself. A third way to analyze charged side chain correlations is to diagonalize the covariance matrix of the charged side chain displacements, and analyze the eigenvectors of this matrix. This is discussed in the next section. It appears that overall, except for transient salt bridges between a few oppositely charged pairs, the charged side chains oscillate and diffuse more or less independently of one another on a nanosecond time scale.

In summary, the total potential within the protein results from mutually compensating contributions from the charged protein side chains on the one hand and the solvent on the other.

4.3. Collective Motions of the Charged Side Chains. A simple way to analyze collective motions and correlations of the charged side chains is to diagonalize the covariance matrix of the charged side chain displacements and analyze the eigenvectors of this matrix. These eigenvectors represent quasimodes of vibration of the charged side chains.[60] The quasiharmonic approximation is usually applied to a protein as a whole. However since the mechanical coupling between the charged side chains and the rest of the protein is rather weak for our system, we can obtain insight into the charged side chain motions while neglecting the rest of the covariance matrix.

The covariance matrix was limited to one atom per charged residue, excluding histidines: Asp C_γ, Glu C_δ, Arg C_δ, and Lys N_ζ, giving 90 quasimodes in all. The projections of the instantaneous side chain motions onto these quasimodes can be thought of as normal coordinates. Analyses of the normal coordinates of proteins are known to give information about the shape of the dynamical attractor the system evolves on and the correlations that arise between collective degrees of freedom as the system moves over the attractor.[61,62] We therefore made scatter plots of all pairs of normal coordinates, in the hopes of detecting multiple attractor basins for some of the pairs. In all cases, the scatter plots only show one central cluster, corresponding to mutually independent motion along the two normal coordinates within a single free energy basin (Figure 3). Thus the collective motions of the charged side chains, like their individual motions, do not show any obvious, strong correlations. A more detailed quasiharmonic analysis of the protein dipolar fluctuations will be reported elsewhere.

4.4. Gaussian Character of the Electrostatic Potential. We now consider the probability distribution of the potential on the C_α positions. We saw above that the relaxation free energy in response to a test charge, in the linear response limit, is given by the variance of the reference electrostatic potential at the test charge site. Deviations from the linear response are given by the deviations of the potential from a Gaussian distribution.

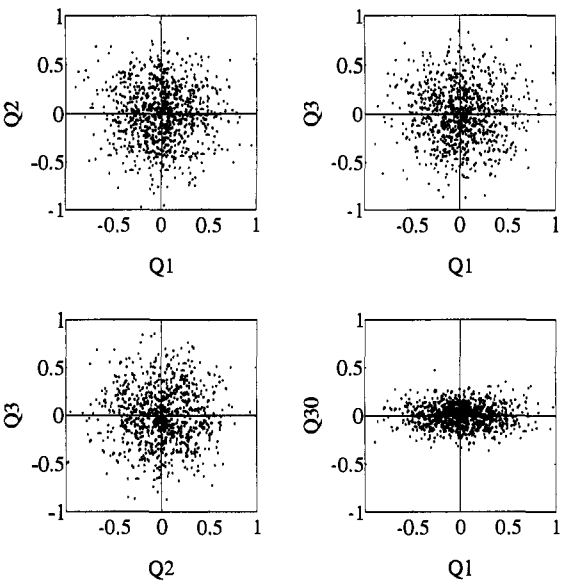

Figure 3. Scatter plots of pairs of quasinormal coordinates Q_i (Å) associated with charged side chain motions (one per side chain). The 90 quasinormal modes are ordered by decreasing eigenvalue, or mean amplitude.

We observe a remarkable property of the electrostatic potential: for all C_α positions, the total electrostatic potential is approximately Gaussian. This property had already been noted for various small solutes in water[41,44-47] and for a few sites in particular proteins.[33-35] The mean values of the potential at the different C_α positions range from −82 to 61 kcal/(mol·e). The standard deviations range from just 5 to 9 kcal/(mol·e). Figure 4 shows the probability distribution of the total potential V_{tot} at a few typical C_α positions, along with the Gaussian distribution having the same mean and standard deviation. A χ^2 test cannot identify statistically significant departures from a Gaussian distribution for any C_α positions, although a Kolmogorov−Smirnov test[63] does point to detectable departures for about 20 positions (including residues 5, 4, and 3, shown in the figure). A more physical measure of the departure from Gaussian distributions is given further on by the third- and fourth-order terms in the dielectric relaxation free energy in response to a perturbation.

For shorter time segments of ~100 ps, significant departures from a Gaussian distribution are seen at about half of the positions, showing that convergence of the probability distributions occurs on at least a 100−500 ps time scale.

The protein and solvent contributions taken separately, although somewhat noisier than V_{tot}, are also nearly Gaussian at about 80% of the C_α positions (Figure 4). On the other hand, the contribution of the charged protein side chains deviates strongly from a Gaussian distribution at about half of the C_α positions (Figure 4). This may reflect insufficient conformational averaging of some of the charged side chain motions.

4.5. Single Reference State Free Energy Calculations: Limits on Statistical Quality. Our ability to extrapolate relaxation free energies from a single unperturbed reference state will obviously decrease as the perturbing test charge increases. Before reporting relaxation free energies in the next section, we examine here the statistical quality that can be expected from a single reference state free energy calculation, as a function of the test charge magnitude.

(60) Karplus, M.; Kushick, J. *Macromolecules* **1981**, *14*, 325−332.
(61) Amadei, A.; Linssen, A.; Berendsen, H. *Proteins* **1993**, *17*, 412−425.
(62) Hayward, S.; Kitao, A.; Hirata, F.; Go, N. *J. Mol. Biol.* **1993**, *234*, 1207−1217.
(63) Press, W.; Flannery, B.; Teukolsky, S.; Vetterling, W. *Numerical Recipes*. Cambridge University Press: Cambridge, U.K., 1986.

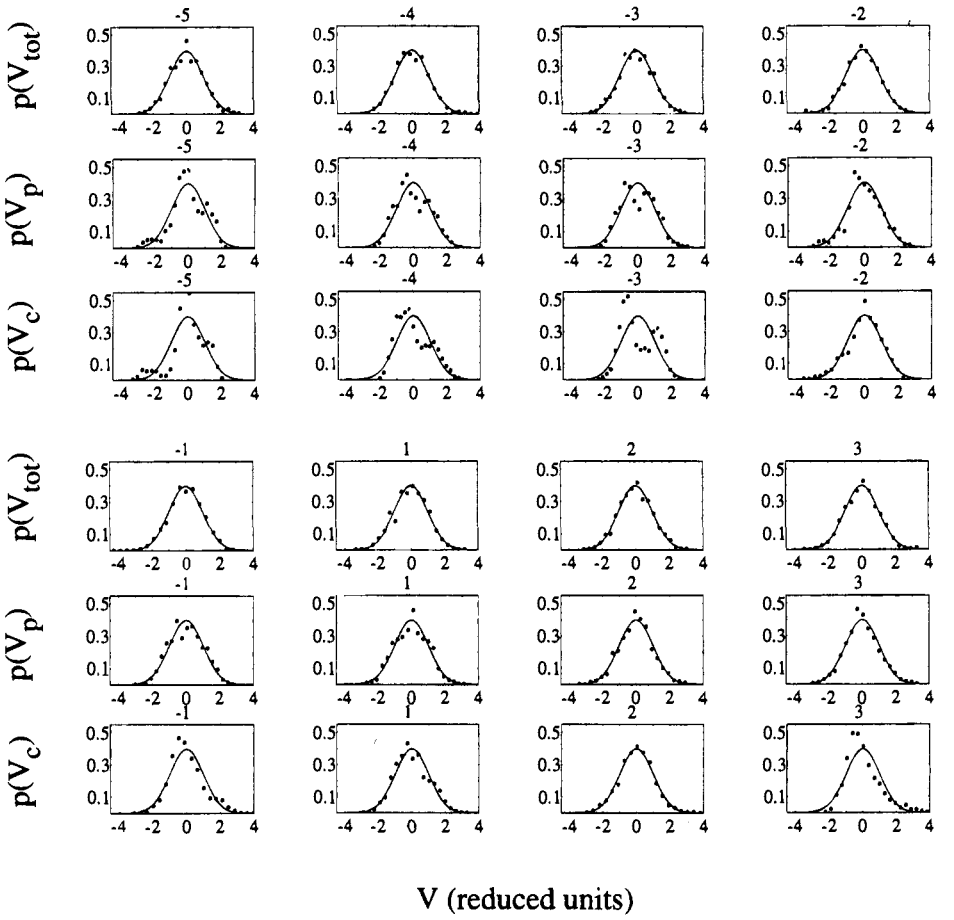

Figure 4. Probability distribution of electrostatic potential on the first eight C_α's from the molecular dynamics simulation. The total potential V_{tot}, the protein contribution V_p, and the contribution of charged side chains V_c are compared. Black dots: results from simulation. Solid curves: Gaussian with the same mean and variance. Each panel is labeled by the C_α residue number. The unit of potential is one standard deviation.

We first consider the first- and second-order terms in the perturbation free energy

$$A_{static} = q\langle V_{tot}\rangle_0$$

$$A_{rlx}^{(2)} = -\frac{q^2}{2kT}\langle \delta V_{tot}^2\rangle_0$$

where $\delta V_{tot} = V_{tot} - \langle V_{tot}\rangle_0$. We denote the probability distribution of the (reference) potential $p(V_{tot})$. To reliably average V_{tot}, the simulation must sample the conformations where $V_{tot}p(V_{tot})$ is large. We just saw that p is approximately Gaussian. In Figure 5 we show the distribution $p(V_{tot})$ actually observed at a typical test charge position, the C_α of residue 5, and its fit to a Gaussian g. The function $V_{tot}g(V_{tot})$ is also shown. We see that the two large lobes of $V_{tot}g(V_{tot})$ are close to the center of the distribution g and extend out to less than three standard deviations of V_{tot}. The simulation covers this range adequately, so that our averaging is satisfactory. Turning to the second-order free energy term (lower panel in the figure), we see that the averaging is not as good as for the first-order term, but still satisfactory. The unsampled positive and negative tails of the probability distribution represent only a 3% correction (shaded in the figure) to $\langle \delta V_{tot}^2\rangle_0$. For higher order terms the quality of the averaging will rapidly decrease. Notice that, while the magnitude of the test charge does not affect the averaging of the individual free energy terms, it affects the weight of each term in the total free energy.

We next consider the full relaxation free energy, including the effects of dielectric saturation

$$A_{rlx} = -kT\ln\langle \exp(-q\delta V_{tot}/kT)\rangle_0$$

To estimate A_{rlx} we need the expectation value of $\exp(-q\delta V_{tot}/kT)$. Unfortunately this function grows so quickly for negative arguments that even states with a minute Boltzmann weight contribute to its expectation value. Figure 6 illustrates the difficulty for a test charge on C_α-5 and several test charge magnitudes. The probability distribution $p(q\delta V_{tot})$ from the simulation is shown and fit to a Gaussian g as before. The product function

$$h(q\delta V_{tot}) = g(q\delta V_{tot})\exp(-q\delta V_{tot}/kT)$$

is also shown. We see that the weak negative tail of the probability distribution g makes a large contribution to the mean value of h. For $q \geq e/4$, the values of $q\delta V_{tot}$ actually sampled in the simulation do not even extend down to the maximum in the function h. Thus the simulation cannot give reliable averaging of h. The undersampling of the distribution's tail leads to an *under*estimate of the mean exponential, and an *over*estimate of the relaxation free energy. Thus the error is systematic and could easily be mistaken for dielectric saturation. For $q = e/10$, on the other hand, the values of $q\delta V_{tot}$ sampled in the simulation cover most of the peak, and we can estimate the mean exponential reasonably well. The unsampled negative tail of the probability distribution represents a 10% correction (shaded in the figure) to $\langle \exp(-q\delta V_{tot}/kT)\rangle_0$. However, since

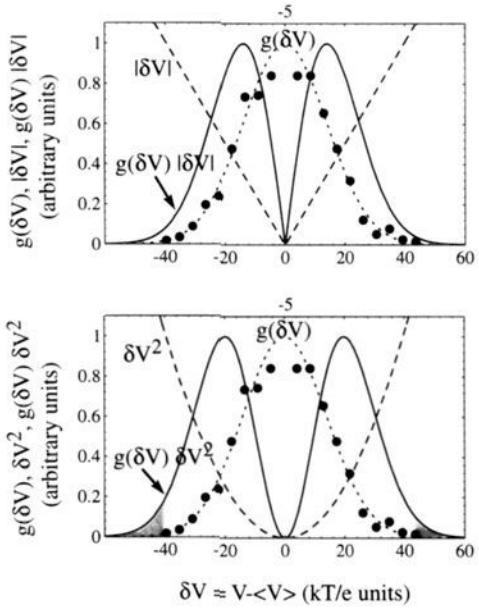

Figure 5. Single reference state free energy calculations: quality of the statistical averaging; first- and second-order free energy terms. Calculation of the expectation values $\langle V_{tot}\rangle_0$ (upper panel) and $\langle \delta V_{tot}^2\rangle_0$ (lower panel). V_{tot} is the potential at a typical test charge site (the C_α of residue 5). Observed probability distribution of V_{tot} (dots); Gaussian with the same mean and variance g (dashes). Long dashes: $|V_{tot}|$ (upper panel) or δV_{tot}^2 (lower panel). Solid: $g|V_{tot}|$ (upper panel) or $g\delta V_{tot}^2$ (lower panel). All functions are arbitrarily scaled.

there is little or no dielectric saturation for such a small charge, we are better off calculating the relaxation free energy directly from the second-order term, where the contribution of the unsampled tails is smaller.

4.6. Relaxation Free Energies in Response to a Test Charge. We next consider a test charge q, placed on each C_α in succession. The total relaxation free energy was calculated for each test charge position. For the moment, we focus on the second-order, linear response, part of the relaxation free energy, which varies as q^2. Our ability to extrapolate free energies from a single unperturbed reference state decreases as q increases, as we just discussed. Exact limits for single state extrapolation (with or without a linear response approximation) have only been established for a few systems, including small solutes in water.[41,64] For small solutes in water, the single state linear response extrapolation holds up to very large perturbation free energies—of over 50 kcal/mol in some cases—and perturbing charges of $\sim e/2 - e$. For oxidation/reduction of the heme group in cytochrome c (T. Simonson, unpublished), the linear response appears to apply moderately well for a relaxation free energy of about 4−5 kcal/mol. The heme environment is not typical of all protein environments, since it evolved specifically for redox purposes. For the present analysis, however, it is probably a reasonable rule of thumb that, for relaxation free energies up to 4−5 kcal/mol, linear response extrapolation is meaningful, if not always quantitatively accurate. Analysis of the third- and fourth-order terms in the perturbation free energies (below) will confirm this.

Because the second-order free energy term scales as q^2, we can report free energies for one particular value of q and derive free energies for other values as required. Instead of reporting free energies for $q = e/10$, or some similar small value, it is more convenient to report values with $q = e$ below and simply keep in mind the actual range of validity of the single state extrapolation.

(64) Smith, P.; van Gunsteren, W. *J. Chem. Phys.* **1994**, *100*, 577−584.

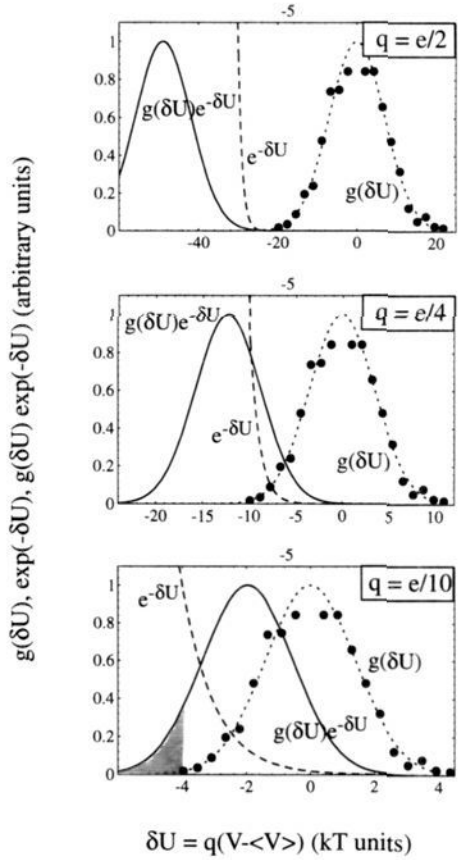

Figure 6. Single reference state free energy calculations: quality of the statistical averaging. Calculation of the expectation value $\langle \exp(-q(V_{tot} - \langle V_{tot}\rangle)/kT)\rangle$. V_{tot} is the potential at a typical test charge site (the C_α of residue 5). Observed probability distribution of the electrostatic energy $q(V_{tot} - \langle V_{tot}\rangle)$ (dots); Gaussian with the same mean and variance g (dashes); $\exp(-q(V_{tot} - \langle V_{tot}\rangle)/kT)$ (long dashes); and the product $g\ \exp(-q(V_{tot} - \langle V_{tot}\rangle)/kT)$ (solid). All functions are arbitrarily scaled.

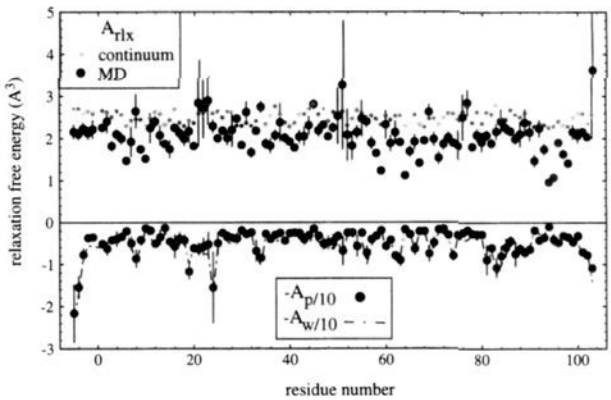

Figure 7. Relaxation free energy (susceptibility units) for a test charge on the C_α's vs residue number. Upper panel: total A_{rlx}, compared to continuum model ($\epsilon_p = 4$, $\epsilon_w = 80$). Lower panel: protein and solvent contributions, scaled by −1/10 for clarity. Error bars are shown for A_{rlx} and A_p (the latter scaled by 1/10). Errors for A_w and A_p are very similar.

The variations of the relaxation free energies as we move the test charge along the C_α chain are shown in Figure 7. The total relaxation free energy varies moderately throughout the protein. For $q = e$, the mean (standard deviation) along the

Table 1. Average Relaxation Free Energies from Different Models[a]

	total[b]		protein[c]		solvent[d]	
model	kcal/mol	Å^3 [e]	kcal/mol	Å^3	kcal/mol	Å^3
polarizable[f]	−25 (3)[g]	0.9 (0.1)	−25 (3)	0.9 (0.1)		
continuum ($\epsilon_p = 2$, $\epsilon_w = 1$)	−26 (4)	1.0 (0.1)	−26 (4)	1.0 (0.1)		
MD[h]	−68 (11)	2.5 (0.4)	−131 (85)	4.9 (3.2)	−151 (91)	5.6 (3.4)
continuum ($\epsilon_p = 4$, $\epsilon_w = 80$)	−67 (4)	2.5 (0.1)	−42 (4)	1.6 (0.2)	−46 (17)	1.7 (0.6)

[a] Relaxation free energies are averaged over all test charge positions (all C_α's). [b] Total relaxation free energy A_{rlx}. [c] Protein contribution A_p. [d] Solvent contribution A_w. [e] Free energies are converted to Å^3 simply by dividing them by a constant factor with the dimensions of a square electric field (see text). [f] Atomic point polarizable model. [g] Standard deviation over all C_α's in parentheses. [h] Molecular dynamics model, including Born cutoff correction.

chain is −56 (11) kcal/mol ≡ 2.1 (0.4) Å^3 (in "susceptibility units"). Results are summarized in Table 1. The total relaxation results from a remarkable cancellation of the three much larger contributions, A_p, A_w, and A_{pw}. Indeed, the direct protein and solvent contributions are nearly equal throughout the protein, while the coupling contribution is nearly equal to $-A_p - A_w$. This cancellation follows directly from the strong correlations between the protein and solvent electrostatic potentials seen above. The mean correlation between the relaxation free energies A_p and A_w, averaged along the C_α chain, is 0.97 (compared to the mean correlation between V_p and V_w of −0.72). The means (standard deviations) along the C_α chain are $\langle A_p \rangle$ = −131 (85) kcal/mol ≡ 4.9 (3.2) Å^3, $\langle A_w \rangle$ = −151 (91) kcal/mol ≡ 5.6 (3.4) Å^3, and $\langle A_{pw} \rangle$ = +226 (170) kcal/mol ≡ −8.4 (6.3) Å^3. The solvent contribution is 16% larger than the protein contribution, even though the average solvent potential is 30% smaller. The charged protein residues account for two-thirds of the protein contribution A_p. The mean correlation of the charged side chain contribution with the total protein contribution is 0.91.

Since the simulation is performed with a 12 Å cutoff, the relaxation free energies were calculated with the same cutoff for consistency. This means that the test charges are effectively inserted into a medium with no interactions beyond 12 Å. A simple correction for this is the Born correction, which we estimated above to be about −12 ± 2 kcal/mol (see the Computational Details). This additional contribution to A_{rlx} is mainly a solvent contribution and is expected to increase going toward the surface of the protein. Adding this term to the above terms gives a mean value for A_{rlx} of −68 kcal/mol ≡ 2.5 Å^3.

For a test charge of $e/4$, A_{rlx} is reduced by 16 and falls in the range where the linear response extrapolation should be reasonable. This sets a rough limit on the magnitude of charges to be considered in applications, which appears somewhat smaller than in applications to small solutes in water.[41,64]

The statistical uncertainties of the relaxation free energies, shown as error bars in Figure 7, were calculated from the deviation of block averages over subsegments of the trajectory. The mean relative uncertainty is 20% for A_p and A_w and 10% for A_{rlx}. Most of the uncertainties are small, with a few large values for test charges on the C_α's of surface residues. Very large uncertainties for A_p occur at Thr-5 (±6.8 Å^3) and Gly-24 (±8.4 Å^3).

Figure 8 illustrates the radial variation of the relaxation free energy going from the center of the protein to its surface. The microscopic results are compared to a continuum model (see section 4.8) which explicitly assumes uniform dielectric properties throughout the protein. The relaxation free energies from the continuum model are nearly uniform throughout the protein, increasing very weakly from the protein center to its surface. On the other hand, the microscopic A_{rlx} increases by about a factor of 2 going toward the outer surface. At points within 10 Å of the center, values are between 1 and 2 Å^3; at the surface, values range from 2 to 3.6 Å^3. The same two regions were

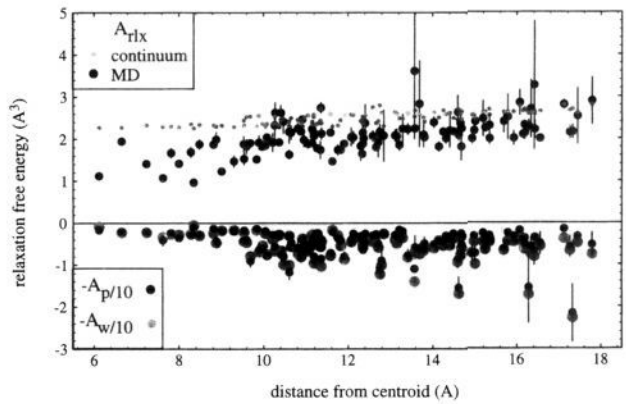

Figure 8. Relaxation free energy (susceptibility units) for a test charge on the C_α's, vs distance from protein centroid. Upper panel: total A_{rlx}, compared to a continuum model ($\epsilon_p = 4$, $\epsilon_w = 80$). Lower panel: protein and solvent contributions, scaled by −1/10 for clarity. Error bars are shown for A_{rlx} and A_p (the latter scaled by 1/10). Errors for A_w and A_p are very similar.

identified earlier from a detailed analysis of the dielectric constant of cytochrome c.[24] The inner region was seen to have a uniform dielectric constant of 2, while the outer portion has a dielectric constant of about 4. This ratio is similar to the ratio of the typical relaxation free energies in the two regions. The low dielectric constant and low relaxation free energy of the inner region are consistent with the redox function of cytochrome c and the requirement of a low reorganization free energy for charge transfer to the from the heme.[22] Earlier theoretical work on cytochrome c also predicted a low reorganization free energy.[31,32]

4.7. Departure from the Linear Response and Dielectric Saturation. As the test charge magnitude increases, dielectric saturation is expected to occur. To measure saturation effects, in principle we simply need to calculate the complete relaxation free energy

$$A_{rlx} = -kT \ln\langle \exp[-q(V_{tot} - \langle V_{tot} \rangle_0)/kT] \rangle_0$$

and subtract out the second-order term. Unfortunately we saw above that it is impossible to make an accurate estimate of $\langle \exp(-q(V_{tot} - \langle V_{tot} \rangle_0)/kT) \rangle_0$ from a single, unperturbed, reference state, for all but the smallest perturbations. A multiple reference state method can of course be used, but this involves at least one explicit simulation for each test charge location.

A simpler measure of the departure from the linear response limit can be obtained from our unperturbed reference state by calculating the third- and fourth-order terms in the perturbation free energy[39]

$$A_{rlx}^{(3)} = \frac{q^3}{6(kT)^2}(\langle V_{tot}^3 \rangle_0 - 3\langle V_{tot}^2 \rangle_0 \langle V_{tot} \rangle_0 + 2\langle V_{tot} \rangle_0^3) \quad (19)$$

$$A_{rlx}^{(4)} = -\frac{q^4}{24(kT)^3}(\langle V_{tot}{}^4\rangle_0 - 4\langle V_{tot}{}^3\rangle_0\langle V_{tot}\rangle_0 - 3\langle V_{tot}{}^2\rangle_0{}^2 + 12\langle V_{tot}{}^2\rangle_0\langle V_{tot}\rangle_0{}^2 - 6\langle V_{tot}\rangle_0{}^4) \quad (20)$$

These terms can be positive or negative. For $q = e$, the averages over the C_α positions are found to be $\langle A_{rlx}^{(2)}\rangle = -53$ kcal/mol, $\langle A_{rlx}^{(3)}\rangle = 0.9$ kcal/mol, $\langle |A_{rlx}^{(3)}|\rangle = 18$ kcal/mol, $\langle A_{rlx}^{(4)}\rangle = -42$ kcal/mol, and $\langle |A_{rlx}^{(4)}|\rangle = 115$ kcal/mol.

For $q = e/4$, these terms are reduced by 16, 64, and 256, respectively: $\langle A_{rlx}^{(2)}\rangle = -3.3$ kcal/mol, $\langle |A_{rlx}^{(3)}|\rangle = 0.3$ kcal/mol, and $\langle |A_{rlx}^{(4)}|\rangle = 0.4$ kcal/mol. The third-order term is 8% of the linear response term, and the fourth-order term is 14%. The third- and fourth-order terms together are 15% of the linear response term. This appears to confirm our rough earlier estimate of the range of q where linear response would apply, $q \leq e/4$. However the higher order terms are not uniform throughout the protein. They are very small in the inner part of the molecule (less than 10 Å from the center). For the outer part, they are fairly scattered and reach 50% of the second-order term at a few positions. Positive and negative values are fairly evenly distributed. Thus dielectric saturation, which implies a positive contribution to the relaxation free energy, must be associated with still higher terms, beyond the third and fourth order.

4.8. Relaxation Free Energies from the Continuum Model. It is instructive to compare the electronic, atomic, and dipolar relaxation just analyzed to a suitable continuum model, both as a consistency check and to gain insight into the nature of the protein dielectric constant. We therefore view the protein as a continuum with a uniform dielectric constant ϵ_p, surrounded by either vacuum or solvent. We again calculate the relaxation free energies and the scalar susceptibilities in response to a test charge located successively on each C_α. The dielectric relaxation free energy has two components: interaction of the test charge with induced volume charge and interaction with induced surface charge at the protein boundary. The first component (step 2→3 in Figure 1) is given by the change in the Born self-energy of the test charge, when the surrounding dielectric constant is changed from 1 to ϵ_p:

$$\Delta A_{Born} = q^2(1/\epsilon_p - 1)/2a$$

where a is the radius of the test charge and q its magnitude. It is always negative. Microscopically, it corresponds to local screening of the test charge by nearby protein groups. The second component (step 3→4 in Figure 1) corresponds to introducing the protein-vacuum dielectric boundary. It is always positive, as the surrounding vacuum repels the test charge. Microscopically, it corresponds to the cost of polarizing the protein in response to the test charge, without any solvent groups to screen the protein polarization.

We begin by considering the protein in vacuum. The test charge is placed successively on each C_α along the protein backbone. We first assume $\epsilon_p = 2$. Indeed, the atomic polarizabilities used above lead, from the Clausius–Mossotti equation,[65,22] to an expected high-frequency dielectric constant of 2. Calculations with this dielectric constant should therefore approximate the electronic relaxation properties of the system.

We initially had set the test charge radius a to 1.9 Å, the van der Waals radius of the C_α in the Charmm19 force field. However, a slightly better agreement with the point polarizable model is obtained with a radius of 2.2 Å, and this is the value used below. The mean scalar susceptibility is then 0.98 Å^3; the standard deviation along the chain is 0.16 Å^3. The susceptibility varies weakly along the C_α chain. Notice that a uniform protein dielectric constant does not imply a completely uniform scalar susceptibility. The mean relaxation free energy is −24 kcal/mol (standard deviation of 4 kcal/mol). In "susceptibility units" the mean relaxation free energy is 0.89 Å^3 (standard deviation of 0.15 Å^3). The Born self-energy contribution is −44 kcal/mol ≡ 1.6 Å^3 (in "susceptibility units"). The mean interaction free energy with the induced surface charge (vacuum repulsion) is +12 kcal/mol ≡ −0.4 Å^3. Results are summarized in Table 2. The sensitivity of the results to the grid orientation in the finite-difference Poisson calculation is negligible.

The results are close to the results from the point polarizability model discussed above. This is consistent with the idea that the electronic polarization can be reasonably well described by a continuum model.[65] The rms deviation between the relaxation free energies in the two models is 2.9 kcal/mol ≡ 0.11 Å^3; the mean correlation is 0.68. Similar agreement was seen earlier for the decaalanine helical peptide[22] and for ions in a polarizable lattice.[2]

Values of ϵ_p up to 16 were also examined. The Born self-energy component of the relaxation free energy decreases as $1/\epsilon_p - 1$, while the vacuum repulsion component is found to increase very slowly with ϵ_p. For $\epsilon_p = 16$, the mean total relaxation free energy is −70 kcal/mol ≡ 2.6 Å^3. The vacuum repulsion component is virtually identical to the $\epsilon_p = 2$ result.

We next turn to the protein in solution. The outer dielectric constant is set to 80, while the inner dielectric constant is varied from 1 to 16. For $\epsilon_p = 1$, the mean relaxation free energy (due entirely to the solvent) is −46 kcal/mol ≡ 1.7 Å^3. The standard deviation along the chain is 17 kcal/mol ≡ 0.6 Å^3. For $\epsilon_p = 16$, the mean relaxation free energy is −84 kcal/mol ≡ 3.1 Å^3, and almost all the relaxation comes from the Born self-energy term. The variation along the chain is negligible as a result.

It is natural to define the protein, solvent, and coupling contributions to the relaxation free energy as in Figure 1. The protein contribution corresponds to step 2→6, and the solvent contribution to step 2→5. The protein and solvent contributions do not add up to the total relaxation free energy: the difference defines a protein–solvent coupling contribution. For $\epsilon_p = 2$–16, the coupling contribution is found to be nonzero, so that the protein and solvent contributions are not additive. However, the coupling contribution to A_{rlx} is only 20–30% of the total (Table 2). This is quite modest compared to the microscopic, molecular dynamics, model considered above. The polarization of the protein around the test charge, and the polarization of the solvent around the protein, conflict only slightly in the macroscopic model. There is little frustration in this purely linear theory.

The mean relaxation free energies can be compared to the molecular dynamics model discussed above. The microscopic relaxation free energy without the Born correction, $\langle A_{rlx}\rangle = -56$ kcal/mol ≡ 2.1 Å^3, is close to the results obtained for $\epsilon_p = 2$ (with $\epsilon_w = 80$ outside the protein). The rms deviation between the macroscopic and microscopic relaxation free energies is 9 kcal/mol ≡ 0.33 Å^3. When the microscopic results are corrected with the Born term, the mean value is −68 kcal/mol, which is close to the macroscopic results obtained with $\epsilon_p = 4$. This value of ϵ_p is very close to the experimental dielectric constant of dry cytochrome c powders[3] and is also very close to the dielectric constant calculated from the present molecular dynamics simulation.[24] While the macroscopic and microscopic models agree approximately, the agreement is significantly

(65) Fröhlich, H. *Theory of Dielectrics*; Clarendon Press: Oxford, U.K., 1949.

Table 2. Average Relaxation Free Energies from the Macroscopic Model[a]

dielectric[b]		total[c]		protein[d]		solvent[e]		Born[f]	
ϵ_p	ϵ_w	kcal/mol	Å^3 [g]	kcal/mol	Å^3	kcal/mol	Å^3	kcal/mol	Å^3
1	80	−46 (17)[h]	1.7 (0.6)	0	0	−46 (17)	1.7 (0.6)	0	0
2	80	−60 (8)	2.2 (0.3)	−26 (4)	1.0 (0.2)	−46 (17)	1.7 (0.6)	−38	1.4
4	80	−67 (4)	2.5 (0.1)	−42 (4)	1.6 (0.2)	46 (17)	1.7 (0.6)	−57	2.1
8	80	−71 (2)	2.6 (0.1)	−52 (3)	1.9 (0.1)	−46 (17)	1.7 (0.6)	−66	2.5
16	80	−73 (1)	2.7 (0.0)	−58 (2)	2.2 (0.1)	−46 (17)	1.7 (0.6)	−71	2.6
2	1	−26 (4)	1.0 (0.1)	−26 (4)	1.0 (0.1)	0	0	−38	1.4
polarizable[i]		−25 (3)	0.9 (0.1)	−25 (3)	0.9 (0.1)				

[a] Relaxation free energies are averaged over all test charge positions (all C_α's). The test charge radius is 2.2 Å. [b] Dielectric constant of protein (ϵ_p) and solvent (ϵ_w). [c] Total relaxation free energy A_{rlx} (step 1→4 in Figure 1), in kcal/mol and "susceptibility units". [d] Protein contribution (step 2→6), which includes both the Born contribution and the vacuum repulsion term (see text). [e] Solvent contribution (step 2→5). [f] Born contribution (step 2→3). [g] Free energies are converted to Å^3 simply by dividing them by a constant factor with the dimensions of a square electric field (see text). [h] Standard deviation over all C_α's in parentheses. [i] Atomic point polarizable model of preceding section.

poorer than for the point polarizability model. The macroscopic and microscopic models correspond to two very different physical pictures. In the microscopic model, the separate protein and solvent contributions are large but the total relaxation free energy is dramatically reduced by protein−solvent correlations. The spatial variation along the C_α chain is significant. In the macroscopic model there is little competition between the protein and solvent contributions, which are roughly additive. Spatial variation along the chain is very weak. Note that the microscopic results are limited to the second-order, linear response, part of the relaxation. This assumes that the test charge magnitude is small, on the order of $e/4$. For larger perturbations, the microscopic relaxation free energies will be reduced by saturation.

5. Discussion

5.1. Limitations of the Calculations. The treatment of electrostatics in the present molecular dynamics simulation suffers from a number of severe approximations, which are commonly used in this field. The use of a cutoff distance (12 Å), the modest number of explicit waters (1400), the absence of counterions, the absence of explicit electronic polarizability, and the still moderate simulation length (1 ns) are the most important limitations. All have a significant effect on calculated electrostatic and dielectric properties. We will argue however that they do not affect the qualitative features of our results: the orders of magnitude, the relative importance of electronic and dipolar relaxation, and the rough degree of spatial variation of the dielectric properties.

In liquid water, use of a cutoff distance has a dramatic effect on orientational correlations and calculations of the dielectric constant.[66] Reaction field[67] and Ewald sum methods[68] give much improved results. Multipole treatments[69] are another promising alternative. Finite cluster calculations provide a way to avoid dealing with long-range forces altogether.[70,24] These methods are difficult to apply to proteins, however, and are still somewhat experimental. In the present work, we apply a long-range correction to the perturbation free energies *a posteriori*, in the form of a continuum Born term. A continuum treatment of interactions beyond 12 Å should be qualitatively correct, and the magnitude of the correction is only 20% of the short-range contribution.

Our simulation system has a 24 Å radius and includes 1400 explicit waters. These choices were largely dictated by available computer resources. The mean protein radius is about 16 Å. Some of the charged side chains extend far out into the solvent and are only a few angstroms from the surrounding vacuum. Nevertheless, the effect of the solvent on the average protein dipolar fluctuations is about the same as that of bulk solvent. Indeed, the variance of the protein dipole moment can be calculated from statistical mechanics, treating the solvent as a continuum.[24] When we assume that the solvent extends to infinity, the result is almost the same as with the finite solvent shell. The variance increases by 18% in the bulk case, indicating a 9% increase in the root mean square atomic displacements if bulk solvent were included. This increase would mainly affect the charged protein side chains. Displacements of atoms in the uncharged protein interior are predicted to increase by only 2%. Thus our solvent layer should provide a reasonable overall model, although local properties are probably distorted in some parts of the structure. Notice that, even though the mobility of the charged side chains would presumably increase if more solvent were included in the model, the effect on dielectric relaxation free energies would be cancelled to a large extent by protein−solvent coupling.

Our molecular dynamics simulation does not include electronic polarization explicitly. However polarization is included in the force field implicitly, in an average way. This average treatment is not too different from a continuum treatment, where the electronic polarizability of protein and solvent reduces the partial charges throughout the system by a constant scaling factor. We showed earlier that coupling between dipolar polarization and electronic polarization is a corrective effect, giving a small contribution to the relaxation free energies.[22] Therefore we expect that adding electronic polarizability to our simulation would essentially lead to an additional, spatially uniform, contribution to the relaxation free energies. This effect should be adequately modeled by the effective scaling of the partial charges in the force field.

We felt that, with a nanosecond simulation length, we would not be able to achieve accurate sampling of counterion distributions, and therefore, we investigated the zero ionic strength limit. We investigated the effect of this assumption on the relaxation free energies using the continuum model and found it to be negligible (results not shown). The absence of any strong correlations between charged side chain motions, despite the lack of electroneutrality, may also be an indication that the solvent provides sufficient screening, so that the lack of counterions does not modify the dipolar fluctuations too strongly.

The nanosecond simulation length appears insufficient to fully sample all the large-scale reorientations of charged side chains

(66) Neumann, M. *J. Chem. Phys.* **1986**, *85*, 1567−1580.
(67) Alper, H.; Levy, R. *J. Chem. Phys.* **1989**, *91*, 1242−1251.
(68) Neumann, M.; Steinhauser, O. *Chem. Phys. Lett.* **1983**, *95*, 417−422.
(69) Board, J. A., Jr.; Causey, J. W.; Leathrum, A. W., Jr.; Schulten, K. *Chem. Phys. Lett.* **1992**, *198*, 89.
(70) Powles, J.; Fowler, R.; Evans, W. *Chem. Phys. Lett.* **1984**, *107*, 280−283.

at the protein surface. Some of the correlation coefficients between distant side chain motions were unrealistically large, and noise is apparent in the contributions of individual side chains to the overall electrostatic potential. There is no obvious remedy to this problem except to run the simulation for a significantly longer time. Nevertheless, since the average dielectric properties appear to have converged fairly well,[24] and since the uncertainties of the relaxation free energies estimated from block averages are mostly small, it seems unlikely that longer simulations would have a large qualitative effect.

Our relaxation free energies are based on a single reference state, linear response approximation.[22,23,41] The accuracy of single reference state free energy calculations was discussed above. The method used here, a thermodynamic integration method including the first and second free energy derivatives, should be more efficient, and allow larger perturbation steps, than the thermodynamic perturbation method. If thermodynamic perturbation is used, the shape of the exponential function leads us to systematically underestimate the magnitude of the dielectric relaxation (although averaging forward and backward simulations will reduce this particular error in a predictable way). For test charges located on the backbone C_α's, the full relaxation free energy can only be calculated for test charge magnitudes up to about $e/10$. Dielectric saturation does not set in until larger magnitudes. Corrections to the linear response appear for test charges of $\approx e/4$ and relaxation free energies of ≈ 5 kcal/mol. For perturbing charges more exposed to solvent, such as a titrating proton, the single reference state method may be applicable to larger perturbations.[71,72] The linear response approximation applies to perturbing charges in water with magnitudes of up to $e/2 - e$.[41,64] When the potential of the unperturbed reference system is nearly Gaussian, as here, a Gaussian fit (Figure 4) can be used to estimate the importance of unsampled regions of conformation space.

5.2. Electrostatic Features of Cytochrome *c*. Despite the approximations made, the molecular dynamics simulation led to an estimate of the average protein dielectric constant in good agreement with experimental data on dry cytochrome *c* powders.[24] The simulation gives a wealth of other information on the electrostatic interactions in cytochrome *c*, summarized below. Thanks to the single reference state, linear response approximation, we can analyze the dielectric response to test charges throughout the protein using a single simulation. Properties that hold throughout most or all of the protein interior (all C_α positions) are likely to be rather general and extend to many or most other proteins.

We found that the electrostatic potential is approximately Gaussian at all C_α positions and probably throughout the protein interior. The Gaussian statistics presumably arise from the independent contributions of many polar groups. The charged protein side chain motions, in particular, are only weakly correlated among themselves. This is consistent with studies on specific redox or active sites in several other proteins[33-35] where the free energy as a function of charge transfer from a donor to an acceptor site was found to be harmonic over a broad range, indicating that the electrostatic potential is nearly Gaussian and the dielectric response of the system nearly linear.

The total potential within cytochrome *c* is found to result from mutually compensating contributions from the charged protein side chains on the one hand and the solvent on the other. It appears that the solvent adapts on a short time scale (a few picoseconds) to the fluctuating protein charges. The solvent screening of the protein charges strongly reduces the electrostatic field within the protein, as basic thermodynamics requires. In continuum theory, the free energy density is proportional to the square of the electric field and the solvent is expected to mirror nearly exactly the permanent charges at the protein surface. In microscopic theories of polar liquids, free energy density functionals have been proposed,[73] which also contain the square of the electric field (as well as additional terms related to the spatial variation of the polarization through the system). The macroscopic solute−solvent complementarity we see here for an entire protein is analogous to microscopic solvation effects at a very local level, where a broken hydrogen bond between two solute groups is immediately replaced by one or more hydrogen bonds to water molecules. Because of this complementarity, protein and solvent respond as a single group to a perturbing charge and the total relaxation is much smaller than either contribution taken separately.

The dielectric relaxation in response to our test charge includes electronic, atomic, and dipolar polarization. Electronic polarization was described with an atomic point polarizability model, commonly used in biomolecular simulations.[27] The relaxation free energies in this model are moderate (~ 1 Å^3) and spatially uniform within the protein. Note that specific anisotropy may exist in the heme region however, due to anisotropic polarizability of the heme π electrons, not included in the model.

The relaxation free energies from molecular dynamics correspond mainly to atomic and dipolar relaxation, although they also include an implicit electronic contribution, since some electron polarization is included in the force field implicitly. The dipolar relaxation (~ 2.5 Å^3) is larger than the electronic relaxation, particularly near the protein surface. The relaxation free energies include almost equal, negative, contributions from protein and solvent, and a large positive contribution from protein−solvent coupling, so that the total is much smaller than the individual contributions. The relaxation free energies can be compared to those from vacuum calculations on tuna cytochrome *c*,[23] which were smaller (1.6 Å^3), due to the very limited mobility of charged groups.

The microscopic free energies increase smoothly from 1−2 Å^3, for a test charge near the protein center, to 2−3.6 Å^3, for a test charge near the protein surface. This is the first observation of such a spatial variation of the microscopic dielectric properties in a solvated protein. Variations are much smoother than for the protein in vacuum.[23] A similar variation was seen recently for the dielectric constant of cytochrome *c*, which increases from 1.5−2 in the inner half of the protein to about 4 in the outer half. The low relaxation in the protein interior is obviously correlated with the limited polarity and mobility of this region. The center of the heme is in a region where the dielectric relaxation free energies are the lowest. This is consistent with earlier work,[31,32] which explicitly calculated the reorganization energy for heme oxidation, using a different model. The low polarizability of the heme region correlates directly with the biological function of cytochrome *c*, since a low reorganization free energy is necessary for efficient electron transfer kinetics. Our simulations thus support the hypothesis of Simonson et al.[22] that the spatial variation of the dielectric properties in proteins can be important for their function.

Departure from the linear response could only be investigated qualitatively because of the limitations of a single reference state calculation. For test charges of $e/4$, the third- and fourth-order free energy terms represent, on average, 15% of the linear response term (of order 2). These higher terms are not uniform

(71) Del Buono, G.; Figueirido, F.; Levy, R. *Proteins* **1994**, *20*, 85−97.
(72) Aqvist, J.; Medina, C.; Samuelsson, J. *Protein Eng.* **1994**, *7*, 385−391.
(73) Chandra, A.; Bagchi, B. *J. Chem. Phys.* **1991**, *94*, 2258−2261.

throughout the protein, so that some regions have a more linear response than others. For the heme region, nonlinear corrections become important for perturbing charges smaller than $e/2$ (T. Simonson, unpublished). Studies of specific redox sites in the photosynthetic reaction center[33,34] and at least one enzyme active site[35] show that specific protein sites can have a linear response to larger perturbations in some cases. It remains to be seen whether this is a specific property of functional charge transfer sites, arising from evolutionary tuning.

We are currently using the methods described here to estimate the redox potential shifts due to mutating several charged side chains of cytochrome *c*, which have been measured experimentally[6] and to estimate the reorganization free energy for intramolecular electron transfer of ruthenium-modified cytochrome *c*'s.[74]

Time-dependent dielectric properties were not treated here; they will be discussed elsewhere.

5.3. What Is the Dielectric Constant of a Protein? Many aspects of protein structure and function can be understood qualitatively (and sometimes calculated quantitatively) with the help of continuum electrostatics.[14,15,28,30] Yet the concept of a protein dielectric constant remains somewhat unclear. Since continuum electrostatics is very approximate for a microscopic system, the definition of the dielectric constant is not unique. Several operational definitions are possible, none ideal, as discussed by Warshel and Aqvist.[75] Probably the most natural definition is based on the relaxation properties of the protein, *i.e.* on its induced polarization shift upon introduction of an external field, or on its relaxation in response to perturbing charges. This is the definition implicit, for example, when one applies the Fröhlich–Kirkwood theory of dielectrics to calculate the dielectric constant of a protein.[20–24] These calculations have given low values (2–5) of the dielectric constant for the inner portion of cytochrome *c*, intermediate values (~10) for the active site of trypsin,[20] and much larger values (20–35) for surface regions containing charged, mobile, side chains. Measurements of the polarization induced in very dry protein powders by an external field also give low values, presumably because ionizable groups are neutral and mobility is reduced in the powders.

With this "relaxation-based" definition, a low dielectric constant does not imply that a protein region is nonpolar, merely that it is nonpolarizable. For example, the heme region of cytochrome *c* contains several polar groups but was tuned by evolution to relax only weakly upon oxidoreduction. Enzyme active sites in general are highly polar but only moderately polarizable. Many previous free energy studies have probed the energetics of enzymes and demonstrated the importance of polar groups in the active site for efficient catalysis.[17] Some of these studies have explicitly calculated the reorganization energy of the reaction, both for proton and electron transfer,[31–36] showing the importance of having relatively rigid polar groups, which lead to low reorganization energies. The high polarity of enzyme active sites has led some authors to speak of a high local dielectric constant.[32] This alternative definition of the dielectric constant is based on the equilibrium charge distribution, both in the protein and in the solvent, in the absence of any perturbing charge, in contrast to the previous "relaxation-based" definition. Finally, notice that charge–charge interactions in the active site of cytochrome *c* and other proteins are strongly screened by the surrounding solvent. This screening has often been measured in the literature by an energy reduction factor referred to as an effective dielectric constant,[76] which is normally large, despite the moderate polarizability of the protein interior itself.

In this work, we compared both the electronic relaxation and the atomic and dipolar relaxation of cytochrome *c* to suitable continuum models. Since continuum models explain many aspects of protein behavior qualitatively, this comparison is not merely a methodological issue but also a basis for intuitive understanding of protein electrostatics. The relaxation free energies associated with electronic polarizability could be accurately fit by a continuum model with $\epsilon_p = 2$ inside the protein. This is the dielectric constant predicted from the Clausius–Mossotti equation for cytochrome *c*.[22] The dielectric constant used in continuum Poisson–Boltzmann calculations normally plays two roles at once: it helps determine the equilibrium charge distribution, and it determines the response to perturbations. However since the continuum calculation of the relaxation free energies is done with all protein permanent charges removed, the dielectric constant here is purely a measure of the relaxation properties, not the equilibrium charge distribution.

The relaxation free energies from molecular dynamics are approximately fit by a continuum model with $\epsilon_p = 4–8$ inside the protein (and $\epsilon_w = 80$ outside). This is consistent with the dielectric constant measured for dry cytochrome *c* powders, and predicted for the cytochrome *c* interior by Fröhlich–Kirkwood theory, using the present simulation.[24] Again, this dielectric constant is purely a measure of the relaxation properties of the system, not the equilibrium charge distribution. Thus for cytochrome *c* in solution, the average relaxation in response to a perturbing point charge is approximately that of a homogeneous, low-to-moderate, dielectric medium. This is found to be especially true in the heme region, consistent with other simulations of cytochrome *c*;[31,32] in surface regions, the relaxation is much stronger.

Acknowledgment. Simulations were done on the Cray C98 of the IDRIS supercomputing center of the Centre National de la Recherche Scientifique. Discussions with Dr. A. Varnek are gratefully acknowledged.

JA950333+

(74) Wuutke, D.; Bjerrum, M.; Winkler, J.; Gray, H. *Science* **1992**, *256*, 1007–1009.

(75) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 268–298.

(76) Warshel, A. *Nature* **1987**, *330*, 15–17.

(77) Reference 22 gave the relaxation energy (p 871), but with the incorrect sign due to a typographical error.