Protein Electron Transfer Reorganization Energy Spectrum from Normal Mode Analysis. 2. Application to Ru-Modified Cytochrome *c*

Gautam Basu,*,† Akio Kitao,‡ Atsuo Kuki,§ and Nobuhiro Go*,‡

Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700054, India, Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan, and Alanex Corporation, 3550 General Atomics Court, San Diego, California 92121

Received: September 2, 1997; In Final Form: January 5, 1998

In an accompanying paper (part 1) we presented a model (NMRES) that describes the coupling of protein fluctuations to electron transfer. The NMRES model, employing normal mode analysis that incorporates Tanford−Kirkwood reaction field energies, relates each normal mode to a mode-specific reorganization energy (λ_k^{prot}), ultimately yielding a protein λ spectrum. In this paper we have successfully applied the NMRES model and analyzed intramolecular electron transfer in Ru-modified cytochrome *c* (at His33). The NMRES estimate for the total protein λ was found to be 15.6 kcal/mol, while the bulk solvent contribution was found to be 7.2 kcal/mol. Of this 15.6 kcal/mol, the high-frequency inner sphere protein modes contributed 3.2 kcal/mol ($\lambda_{\text{in}}^{\text{prot}}$), while the remaining 12.4 kcal/mol ($\lambda_{\text{out}}^{\text{prot}}$) arose from the low-frequency outer sphere protein modes, the focus of this paper. Out of about 600 "soft" low-frequency modes, 60% contributed very little, while the remaining 40% contributed more or less equally. There were no special soft modes in terms of contribution to $\lambda_{\text{out}}^{\text{prot}}$, structurally or energetically. In other words, although not all the soft modes contributed, those that did shared the coupling more or less equally, implying that minor changes in the dynamic structure will not alter the total λ significantly. This could be the reason that the experimental λ on Ru-modified (at various His sites) cytochrome *c* is found to be almost invariant.

1. Introduction

A quantitative understanding of biological function at the molecular level is the ultimate goal of any biophysical research. A vital part of these functions is energy transduction, whose mechanism and efficiency are controlled by a group of proteins that participate in a complicated series of inter- or intramolecular electron-transfer (ET) processes.[1] These ET reactions typically occur between suitable cofactors embedded within the protein matrix. The protein matrix plays an important but only a passive role of "controlling" the intrinsic inter-cofactor ET by surrounding and shielding the cofactors from an otherwise nonprotein surrounding. The mechanism of this passive control of ET is subtle, and a quantitative understanding of this control holds the vital clue for understanding the secrets of biological energy transduction.

Experimental studies of protein ET yield a single (average) number, the rate constant for the ET process, k_{ET}. To delineate various factors that give rise to k_{ET}, appropriate theoretical models have been formulated. When quantum mechanical effects along the nuclear degrees of freedom are negligible, a suitable expression for k_{ET} is given by[2]

$$k_{\text{ET}} = \left(\frac{2\pi}{\hbar}\right) H_{\text{DA}}{}^2 \frac{1}{\sqrt{4\pi\lambda k_{\text{B}}T}} \exp\left\{-\frac{(\Delta G^\circ + \lambda)^2}{4\lambda k_{\text{B}}T}\right\} \quad (1)$$

Here three factors that control the rate constant, H_{DA}, ΔG°, and λ, are in turn modulated by the protein matrix.

The first factor, H_{DA}, is related to the interaction of the protein-embedded donor−acceptor electronic wave functions. The reaction free energy, ΔG°, is related to the stability of the oxidized and reduced cofactors in the protein matrix. These two factors stand out from the third factor, λ. The factor λ originates from the fluctuation of the nuclear coordinates surrounding the red-ox pair. It is not that such fluctuations will have no effect on H_{DA} or ΔG°, but even when the medium is strictly "frozen", H_{DA} and ΔG° will have nonzero values. In contrast, λ will have no physical meaning in a frozen medium. It originates from the fluctuations of the medium, specifically from the ability of the medium to respond to ET. Appropriately, λ is known as the reorganization energy and defined as the free energy released as the medium nuclear coordinates reorient upon ET.

Experimentally one can directly measure k_{ET} and ΔG° of eq 1. The factors H_{DA} and λ, on the other hand, are determined only indirectly.[3] One can indirectly determine λ from variations of k_{ET} as a function of ΔG° followed by an analysis through eq 1. A knowledge of λ subsequently yields H_{DA} through eq 1. Alternatively, temperature dependence studies of k_{ET} followed by a nonlinear least-squares fitting of $k_{\text{ET}}(T)$ to eq 1 can yield H_{DA} and λ.

Another important class of ET experiments is to measure k_{ET} as a function of disposition of the red-ox pair in a protein matrix. The variation of k_{ET} in these experiments has typically been interpreted as arising solely from "distance" dependence of H_{DA}, assuming that λ remains unchanged, an assumption that may not be strictly valid.[3] Therefore, to properly ascribe different factors that are responsible for the observed variations of k_{ET} in this class of experiments theoretical modeling of both H_{DA} and λ are important.

† Bose Institute.
‡ Kyoto University.
§ Alanex Corp.

S1089-5647(97)02847-2 CCC: $15.00 © 1998 American Chemical Society
Published on Web 02/20/1998

A prerequisite for any theoretical estimate is the 3D structure of the protein obtained from X-ray or NMR studies. Given a protein 3D structure, H_{DA} can be calculated through established theoretical models of various degrees of sophistication. A popular way to model H_{DA} is by breaking it up as contributions from several "paths", emphasizing the complex architecture of the protein matrix.[4,5]

A single 3D structure, however, does not allow a theoretical estimate of λ. One way to circumvent this problem is to model the medium by a continuous[6,7] or piecewise continuous dielectric[8] and account for the relevant relaxation of the coordinates through appropriate factors containing the static and optical dielectric constants. Given two static structures, corresponding to the oxidized and reduced protein, one can also estimate λ either by constructing a reaction coordinate (the vector joining the two structures) or by converting the observed geometry changes into appropriate electrostatic energy differences.[9]

Often experimental structures for both the oxidized and the reduced proteins may not be available, and more importantly, an explicit connection of protein fluctuation to λ may be desirable, rather than a connection to static (equilibrium) conformational changes. Similar to expressing H_{DA} as arising from several paths, this is equivalent to expressing λ as arising from several orthogonal protein fluctuation modes. In this respect, a complete approach toward estimating λ would be to theoretically model explicit fluctuations of the protein (and bulk solvent) around an equilibrium conformation, followed by a systematic identification of those fluctuations that are affected by ET. A molecular dynamics simulation followed by an analysis of the energy difference (reactant−product) probability or the energy difference autocorrelation function has been found to be a satisfactory strategy toward this goal.[10−13] Although in terms of a detailed analysis of the rate constant, the latter approach is complete, a shortcoming of the approach is that it cannot structurally identify important (with respect to ET) fluctuations. In a preceding paper[14] we have proposed a model (NMRES) where the emphasis is on structural identification of collective fluctuations of proteins that couple to ET. In the NMRES model the total λ is first partitioned into the solvent and the protein contribution, followed by a complete breakup of the latter contribution among all the protein collective modes as the reorganization energy spectrum.

The connection between the observed rate constant k_{ET} and λ, as prescribed by eq 1, is strictly valid only for intramolecular ET reactions. In contrast, most biological ET reactions are intermolecular with few exceptions, like the photosynthetic reaction center. Intramolecular ET is also exhibited by a small and extremely well characterized protein, cytochrome *c* (Cytc), where intramolecular ET has been studied extensively by synthetically modifying His residues with ruthenium complexes by Gray and co-workers.[3] Besides being smaller and simpler than the reaction center, the Ru-modified Cytc system has the added advantage that ET has been studied in almost all possible directions from the central heme, wherever an accessible His group is available.

In this paper we present simulation results, as prescribed in the NMRES model, on horse heart Cytc modified by $Ru^{II/III}(NH_3)_5$ at His33. The primary aim of this paper is to test the applicability of the NMRES model to a well-characterized protein system. The protein fluctuations are modeled by collective normal modes (NM) that incorporate reaction field energies (E_{RF}) to account for equilibrium solvation energy. This is the first time that an NM analysis (NMA) has been performed with the inclusion of E_{RF}. We find that low-frequency modes contribute significantly toward λ and are sensitive to how the solvent is modeled. A detailed analysis of the reorganization energy spectrum is presented, and implications for the observed results are discussed.

2. Materials and Methods

2.1. The System. To perform a set of actual simulations with a protein, we were constrained by three factors: (1) that the system should not be very large, (2) that it should have been experimentally well studied, exhibiting a clean intramolecular ET, and (3) that it should be as spherical as possible for the NMRES model[14] to be better applicable. Semisynthetic Ru-modified Cytc, studied extensively by Gray and co-workers,[3] satisfied all the three criteria.

Accordingly we chose the simplest member of the Ru-modified proteins: horse heart Cytc, modified by the attachment of $Ru(NH_3)_5$ at His33 (Ru-Cytc). The initial coordinates for the unmodified protein, along with six bound water molecules, were obtained from NMR studies by Xiurong et al.[15] The residue His33 was modified by the attachment of a Ru atom with five liganded NH_3 molecules using AMBER geometry and the molecular modeling package InsightII (Biosym Technologies, 1993). Previously published partial charges were used for the reduced[16] and the oxidized[9] heme group, and partial charges for the reduced and oxidized Ru complex [Im(His33)−Ru-$(NH_3)_3$] were used as reported elsewhere.[17] All other parameters were taken from AMBER−OPLS potential energy function.[18,19] The potential parameters employed here can be provided upon request by e-mail to go@qchem.kuchem.kyoto-u.ac.jp.

2.2. Energy Minimization. Two sets of energy minimizations were performed, differing from each other in how the dielectric constant/solvent screening was modeled. In the first set, a distance-dependent dielectric constant [$\epsilon = 2r$ (in Å)] was used (MIN_{2R}), and in the second set, Tanford−Kirkwood (TK) type[20] E_{RF} were incorporated (MIN_{TK}). Energy minimizations were performed with the package PRESTO[21] with a suitable combination of steepest descent and conjugate gradient techniques.

Both MIN_{2R} and MIN_{TK} minimizations were performed with the reduced (Fe^{II}/Ru^{III}) protein charge set. The starting NMR conformation, as modified by attaching $Ru(NH_3)_5$, was first energy-minimized by holding the backbone atoms of the proteins fixed. Subsequently unrestrained minimizations were performed. For MIN_{2R}, the latter was straightforward. For MIN_{TK}, the E_{RF} and the corresponding energy gradients were calculated, for a fixed cavity radius *b* of 21 Å, as described in the Appendix. The value of *b* arises from a sum of 18.5 Å (radial distance of the outermost atom, R_{max}) and 2.5 Å (solvent exclusion radius, R_{ex}).

Since Ru-Cytc is not strictly spherical, R_{max}, which in turn determines the depth of the buried charged atoms, crucial in E_{RF}, had to be chosen with care. In Figure 1 the radial distribution of the number and number density of atoms in Ru-Cytc are shown. R_{max} was found to be 23.2 Å for the starting conformation, and as can be seen from Figure 1, the radial density tails after about 18.5 Å. Further, atoms lying beyond 18.5 Å (53 atoms) mostly belong to the Lys side chains sticking out of the core protein, and the distances of these atoms from the Fe or Ru center (Fe/Ru and the attached N atoms) are large (>14 Å). Before a full MIN_{TK} calculation was attempted, these atoms were slowly packed inside a spherical cavity until the final R_{max} became 18.5 Å through several steps of constrained minimization with a harmonic penalty energy for atoms lying outside the desired R_{max}.

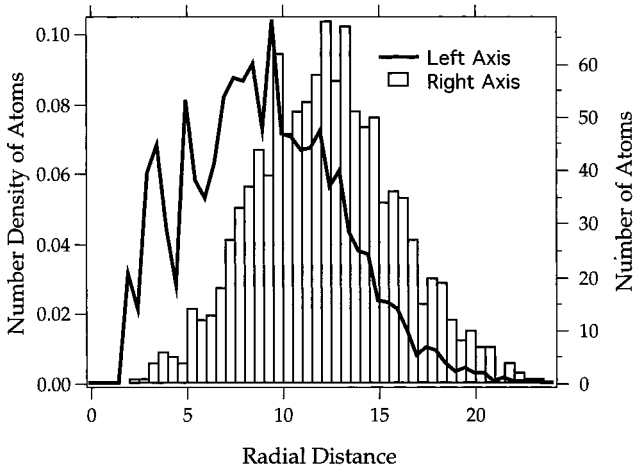

Figure 1. Radial distribution of the number of atoms (histogram) and number density of atoms (thick curve) in $Ru(NH_3)_5$−His33−Cytc at the beginning of energy minimization. The radial distance is measured from the center of gravity of the molecule.

2.3. Normal Mode Analysis. Within the harmonic approximation, NMA is an established method for studying fluctuations of a protein molecule around a minimum energy conformation (MEC). In the present work, NMA was performed in the mass-weighted Cartesian coordinate space ($X_i = \sqrt{m_i}x_i$) as described elsewhere.[22] Displacement of the mass-weighted Cartesian coordinates, ΔX_i, as a function of the NM variables, σ_k, is given by

$$\Delta X_i = \sum_k^{3N-6} \alpha_{ik}\sigma_k \tag{2}$$

where N is the total number of atoms and α_{ik} is the ikth member of the eigenvector matrix $\boldsymbol{\alpha}$, which transforms the Hessian $\mathbf{F}$ to a diagonal matrix $\boldsymbol{\omega}$ containing the NM frequency squared and orthonormalized as

$$\alpha^{\mathrm{T}}\alpha = I \tag{3}$$

In the present work, the Hessian of the conformational energy was calculated numerically from the analytically calculated gradients. In terms of σ_k, the potential energy is given by

$$E = \sum_{k=1}^{3N-6} \frac{1}{2}\omega_k{}^2\sigma_k{}^2 \tag{4}$$

where ω_k is the angular frequency of the kth mode.

2.4. Equilibrium Shift along Normal Modes. In the NMRES model, the total protein reorganization energy, λ^{prot} and the mode-specific reorganization energy, $\lambda_k^{\mathrm{prot}}$, is related to the equilibrium shift of σ_k, $\Delta\sigma_k^\circ$, as

$$\lambda^{\mathrm{prot}} = \sum_k^{3N-6} \lambda_k^{\mathrm{prot}} = \sum_k^{3N-6} \frac{1}{2}\omega_k{}^2(\Delta\sigma_k^\circ)^2 \tag{5}$$

The shift, $\Delta\sigma_k^\circ$, is given by the following derivative:

$$\Delta\sigma_k^\circ = \frac{1}{\omega_k{}^2}\left(\frac{\partial\Delta E}{\partial\sigma_k}\right) \tag{6}$$

where ΔE is the energy difference between the product and the reactant states. Although an analytical formula for the derivative in eq 6 can be derived, we determine the derivative numerically in this paper. One can, for convenience, rewrite E in the NMRES model in terms of the scaled NM variable, τ_k (= $(\omega_k/\sqrt{k_{\mathrm{B}}T})\sigma_k$) as

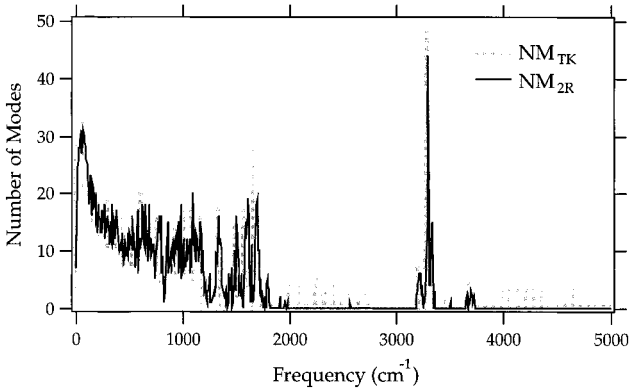

Figure 2. Histograms of NM-frequency number in each bin of 10 cm^{-1} from NMA_{2R} (thin black line) and NMA_{TK} (thick gray line).

$$E = \sum_{k=1}^{3N-6} \frac{1}{2}(k_{\mathrm{B}}T)\tau_k{}^2 \tag{7}$$

In terms of τ_k, the expressions for $\Delta\sigma_k^\circ$ and $\lambda_k^{\mathrm{prot}}$ of the NMRES model follow as

$$\Delta\sigma_k^\circ = \frac{1}{\sqrt{k_{\mathrm{B}}T}\omega_k}\left(\frac{\partial\Delta E}{\partial\tau_k}\right) \tag{8}$$

$$\lambda_k^{\mathrm{prot}} = \frac{1}{2}\left(\frac{1}{k_{\mathrm{B}}T}\right)\left(\frac{\partial\Delta E}{\partial\tau_k}\right)^2 \tag{9}$$

2.5. Bulk Solvent Reorganization Energy. Within the NMRES model the bulk solvent contribution to λ, λ^{blk}, is given by

$$\lambda^{\mathrm{blk}} = E_{\mathrm{RF}}(\Delta Q_i, D_{\mathrm{OP}}^{\mathrm{prot}}, D_{\mathrm{OP}}^{\mathrm{blk}}, \mathbf{r}_{\mathrm{MEC}}) - E_{\mathrm{RF}}(\Delta Q_i, D_{\mathrm{OP}}^{\mathrm{prot}}, D_{\mathrm{S}}^{\mathrm{blk}}, \mathbf{r}_{\mathrm{MEC}}) \tag{10}$$

where ΔQ_i is the change in charge distribution at the ith protein site (including the red-ox centers), D stands for the dielectric constant, static (S) or optical (OP), $\mathbf{r}_{\mathrm{MEC}}$ is the MEC protein coordinates, and E_{RF} is calculated according to eq A1 (Appendix). The values for $D_{\mathrm{OP}}^{\mathrm{prot}}$, $D_{\mathrm{OP}}^{\mathrm{blk}}$, and $D_{\mathrm{S}}^{\mathrm{blk}}$ used were 2.0, 1.8, and 78.0, respectively. We also calculated E_{RF} numerically by solving the Poisson−Boltzmann (PB) equation using the DelPhi (Biosym Technologies, 1993) package. In the latter calculation the actual protein shape is taken into account instead of a spherical protein shape.

3. Results

3.1. Energy Minimization and Normal Mode Analysis. Two sets of NMA, NMA_{2R} and NMA_{TK}, were performed corresponding to the two sets of energy minimizations, MIN_{2R} and MIN_{TK}, respectively. The softest five modes in MIN_{2R} and MIN_{TK} had frequencies of {5.32, 5.91, 6.70, 7.10, 7.61} and {4.50, 4.91, 5.44, 6.44, 6.63} cm^{-1}, respectively. Even though NMA_{TK} exhibited softer modes than NMA_{2R} in the lowest frequency range, the overall distributions of the number density of modes for both were very similar. This is shown in Figure 2. Because the AMBER/OPLS potential treats CH as a united atom, there are no CH bond stretching modes in Figure 2.

In Figure 3 deviations from ideal (harmonic) behavior for the 50 lowest frequency NMA_{TK} modes are shown. Within the

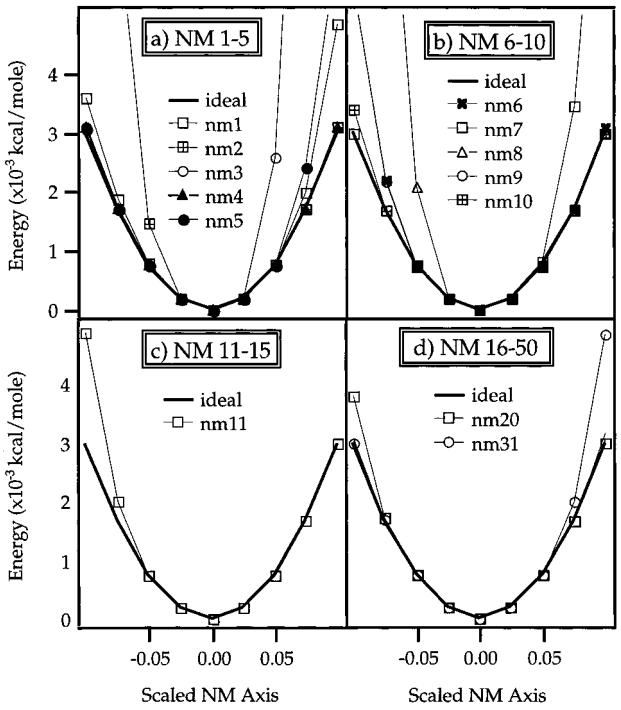

Figure 3. Ideal (thick) and actual (thin) potential energy curves for the 50 lowest frequency modes in NMA_{TK}.

range of τ_k shown (−0.1 to 0.1), 12 modes show appreciable deviations from the ideal behavior, of which 9 belong to the 10 lowest frequency modes. The actual potential energy surface along these modes is steeper than predicted by the harmonic approximation. Similar trends in deviation from ideal behavior have been observed for other systems as well.[23]

3.2. Energy Derivatives and Protein Reorganization Energy Spectrum. The mode-specific protein reorganization energy, λ_k^{prot}, is related to the derivative of the energy difference ΔE along the scaled NM variable τ_k as given in eq 9. In this paper we calculated the derivatives numerically, where ΔE is the difference of actual energy of the reactant and product states. While calculating the derivatives, we also checked the linearity assumption of $\Delta E(\tau_k)$ in the NMRES model.

In Figure 4a the energy difference ΔE is shown along the τ_k axis for some selected modes in NMA_{2R} (modes 22−31). The linearity of $\Delta E(\tau_k)$ holds quite well for this range of τ_k. The linearity assumption was checked up to $\tau_k = \pm 1.0$, and it was valid for other modes as well. Similarly in Figure 4b ΔE is plotted against τ_k for the first 50 modes in NMA_{TK}. Similar to the case of NMA_{2R}, again we observe that the linearity assumption holds quite well. The linearity assumption holds well even for modes that showed severe anharmonicity in Figure 2, and this was true over a substantial range of τ_k (not shown in Figure 4b).

The ΔE derivatives subsequently yielded a set of λ_k^{prot} through eq 9. The λ_k^{prot} spectrum is shown for NMA_{2R} in Figure 5; Figure 5a shows the spectrum along the mode frequencies, and in Figure 5b the mode numbers (starting from the lowest frequency modes) are shown. The cumulative λ_k^{prot} is also shown. The total protein reorganization energy from all modes $\lambda^{prot} = \sum\lambda_k^{prot}$) is about 3.4 kcal/mol.

As opposed to this low value of λ^{prot} in the NMA_{2R} model, similar calculations for the NMA_{TK} model yielded a much larger value of λ^{prot}, about 15.6 kcal/mol. In Figure 6 the protein reorganization energy spectrum for the NMA_{TK} model is shown along the frequency and the mode numbers. Fluctuations of

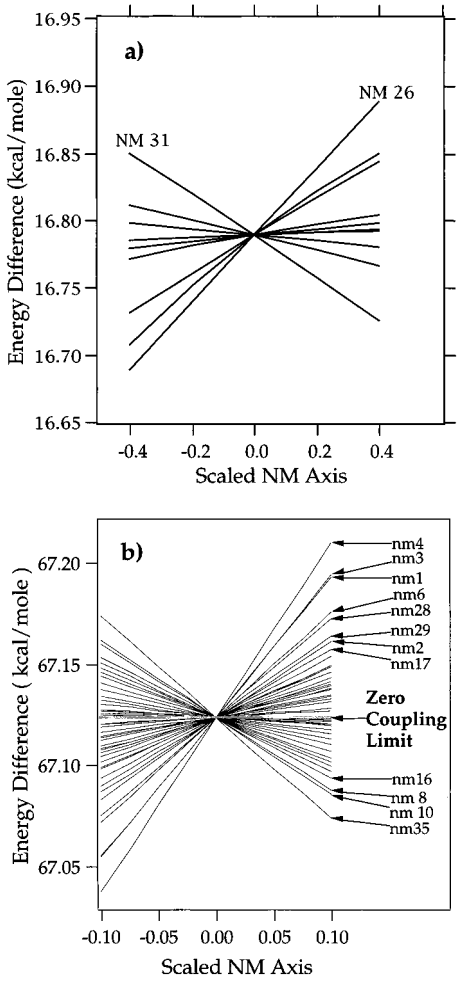

Figure 4. ΔE as a function of scaled NM axis τ_k for NMA_{2R} (a: modes 22−31) and NMA_{TK} (b: 50 lowest frequency modes).

the backbone atoms including the red-ox centers are shown in Figure 7 for three modes in the NMA_{TK} analysis: the most-coupled low-frequency mode (mode 4; $\nu = 6.44$ cm^{-1}; $\lambda_k^{prot} = 0.63$ kcal/mol), the most-coupled high-frequency mode (mode 2748, $\nu = 1510.13$ cm^{-1}; $\lambda_k^{prot} = 0.27$ kcal/mol), and a typical barely coupled mode (mode 5, $\nu = 6.63$ cm^{-1}; $\lambda_k^{prot} = 0.009$ kcal/mol).

3.3. Solvent Contribution to Total Reorganization Energy. The solvent contribution to the total reorganization energy, λ^{blk}, was calculated according to eq 10. Within the NMRES model, E_{RF} in eq 10 was calculated using the TK approximation for a range of cavity radius b including 21 Å (used for calculating λ^{prot}). With the appropriate radius of 21 Å, λ^{blk} was found to be 7.2 kcal/mol. As b was decreased by increments of 0.5 Å, λ^{blk} increased as 8.2(20.5), 9.5(20.0), and 11.2(19.5) kcal/mol, where the corresponding b values (in Å) are indicated in parentheses. The λ^{blk} contribution was also estimated from E_{RF} obtained by numerically solving the PB equation. This was found to be about 22.6 kcal/mol.

3.4. Conformational Change upon ET. From the set of shifts $\Delta\sigma_k°$, the corresponding change in the individual atomic coordinates, $\Delta X_i°$, can be calculated according to

$$\Delta X_i° = \sum_k^{3N-6} \alpha_{ik}\Delta\sigma_k° \qquad (11)$$

Equation 11 yielded a total root-mean-squared deviation (RMSD) between the reduced and oxidized Ru-Cytc of 0.31 Å. The

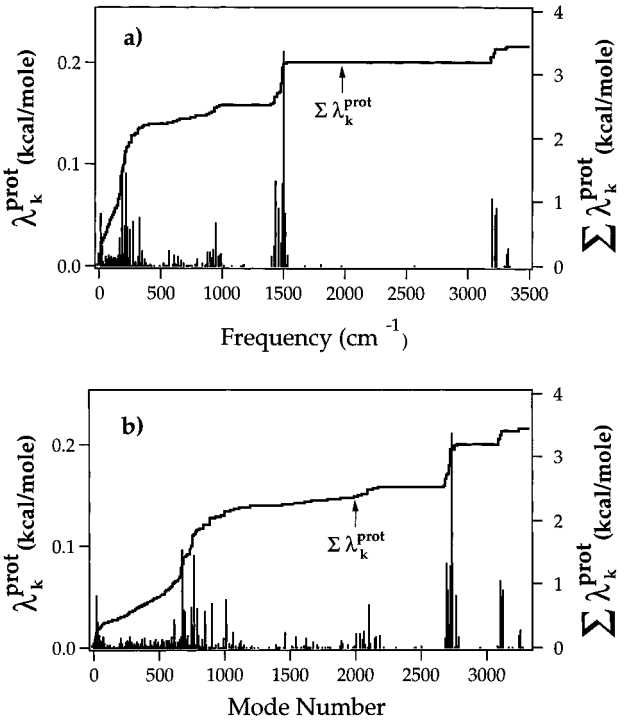

Figure 5. Protein reorganization energy spectrum from NMA_{2R} calculation, as a function of frequency (a) and mode number (b).

average shifts in atomic positions upon ET for all the residues are shown in Figure 8. For comparison, the average distances of the red-ox sites (Fe/Ru and the associated N atoms) from the residues are also shown in the figure.

4. Discussion

4.1. Energy Minimization and NMA. Energy minimization is not the central focus of this work; yet, since this is the first time that energy minimization was attempted with the inclusion of E_{RF}(TK), it deserves some attention. In a previous study Beglov and Roux[24] performed MD simulations with the inclusion of E_{RF}(TK) as well. In their case, the spherical cavity not only contained the solute (protein) but was also packed with explicit water molecules. Unfortunately when the cavity contains the solute and several explicit solvent molecules and its energy includes E_{RF}(TK), the quality of energy minimization becomes extremely poor. On the other hand, with only the solute and a few solvent molecules occupying the cavity, a significant amount of empty space within the cavity can underestimate solvent effects brought in through the E_{RF} term. Even with this limitation, we feel that provided the solute is fairly spherical, inclusion of E_{RF} in the potential energy function models solvent screening better than traditional models of the dielectric constant. We were able to energy-minimize Ru-Cytc in a fairly straightforward manner with appropriate centering of the molecule (see Appendix), up to a gradient of about 10^{-9} (kcal/mol)/Å, required for NMA. The RMSD of the backbone atoms were 2.37 Å (starting and MIN_{2R} conformations), 2.44 Å (starting and MIN_{TK} conformations), and 2.77 Å (MIN_{2R} and MIN_{TK} conformations).

The NMA performed for the two electrostatic models were also consistent with each other when judged from the distribution of the frequency densities (Figure 2). The anharmonicity test for the softest 50 modes in NMA_{TK} (Figure 3) is also consistent with NMA results obtained with other systems: the lowest few modes are anharmonic with slightly steeper energy surfaces.[23] Similar results were found with NMA_{2R}, not shown here. In other words, except for the softest modes shifting toward lower frequencies, the NMA_{TK} results were not significantly different (at least, apparently) from NMA with traditional models of the dielectric constant.

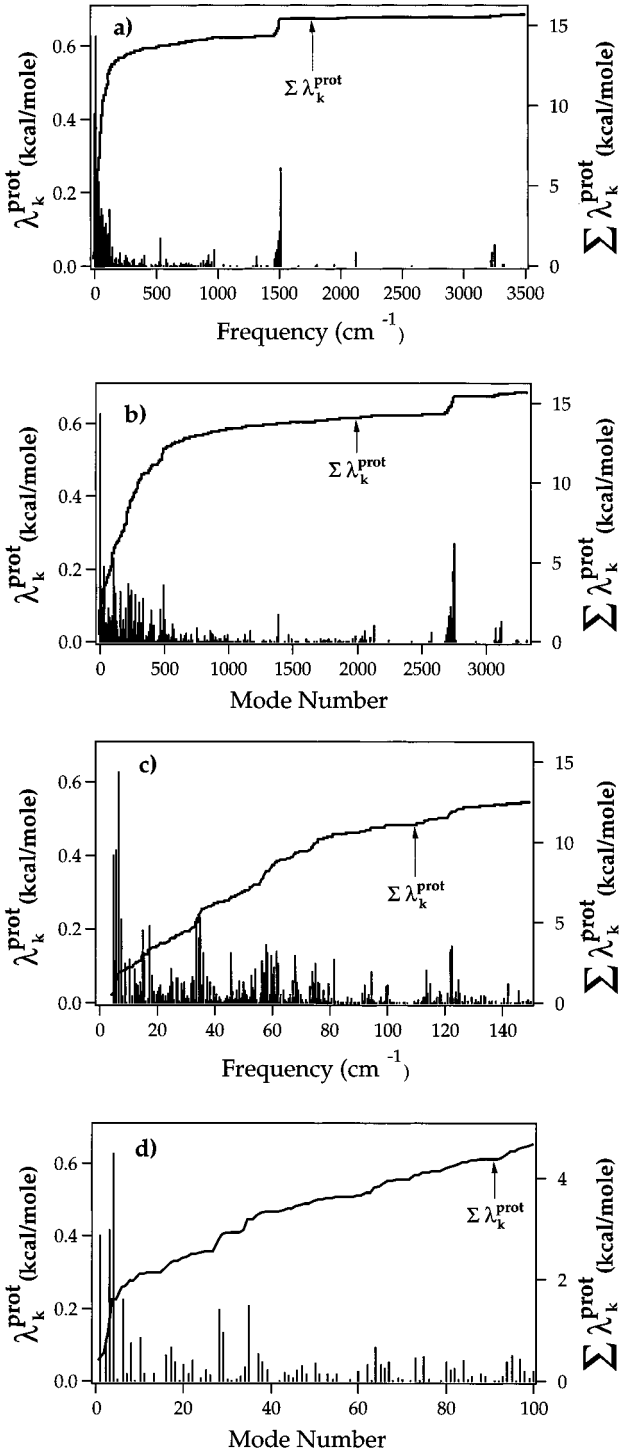

Figure 6. Protein reorganization energy spectrum from NMA_{TK} calculation, as a function of frequency (a, c) and mode number (b, d).

4.2. Inner and Outer Sphere λ^{prot}. The crucial difference between NMA_{2R} and NMA_{TK} arises when the mode-specific protein reorganization energies, λ_k^{prot}, are calculated. As seen in Figures 5 and 6, the total protein contribution, λ^{prot}, from the former is only 3.4 kcal/mol, whereas it is 15.6 kcal/mol from the latter calculation. If one considers 150 cm^{-1} to be the upper limit for low-frequency modes cumulatively contributing to the outer sphere λ^{prot}, λ_{out}^{prot}, and modes above this frequency

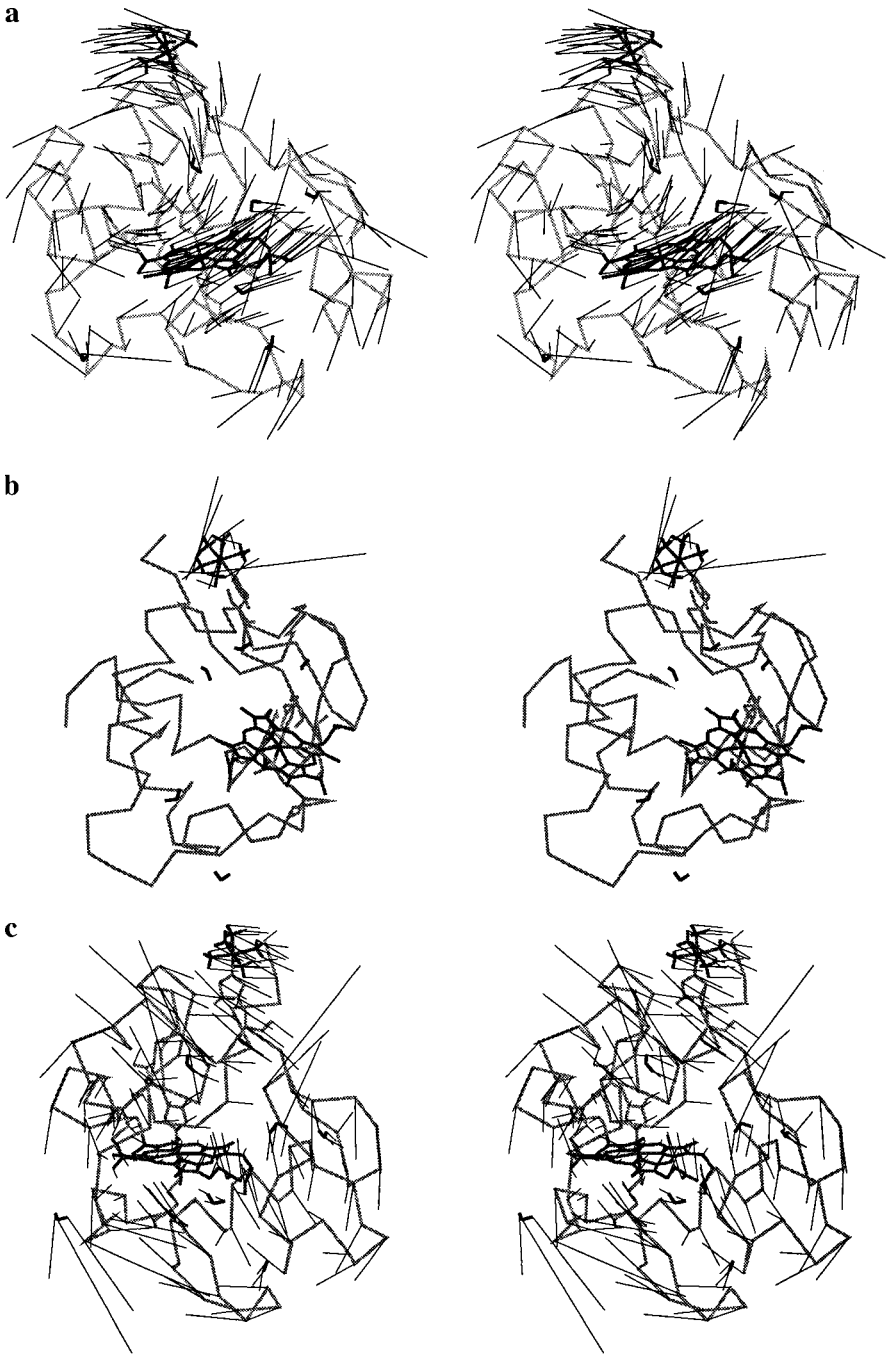

Figure 7. Three modes in NMA_{TK} calculation: (a) the most coupled low-frequency mode, (b) the most coupled high-frequency mode, and (c) a typical low-frequency uncoupled mode. Only the C^{α} atoms (thick shaded) and the red-ox sites (thick dark) are shown. The vectors are shown by thin dark lines.

contributing to the inner sphere λ^{prot}, λ_{in}^{prot}, the breakup of λ^{prot}-(NMA_{TK}) into its components yields $\lambda_{in}^{prot} \approx 3.2$ kcal/mol and $\lambda_{out}^{prot} \approx 12.4$ kcal/mol. A similar breakup of $\lambda^{prot}(NMA_{2R})$ yields $\lambda_{in}^{prot} \approx 2.5$ kcal/mol and $\lambda_{out}^{prot} \approx 0.9$ kcal/mol. The difference between $\lambda^{prot}(NMA_{2R})$ and $\lambda^{prot}(NMA_{TK})$ thus arises from the contributions from the low-frequency modes, λ_{out}^{prot}.

For both the electrostatic models, the contribution from the inner sphere high-frequency modes is very similar (2.5 and 3.2 kcal/mol). This value is consistent with an independent estimate reported earlier on the isolated heme moiety. This was estimated to be around 1 kcal/mol.[9] Adding a similar contribution from the $Ru(NH_3)_5$ moiety, which is closer to the surface and less constrained than the heme moiety, will have a slightly higher contribution. We find the predicted λ_{in}^{prot} from both the electrostatic models reasonable.

4.3. Solvent Contribution. Before commenting on the λ_{out}^{prot} contributions from the two models, let us first consider the calculated λ^{blk} values. The λ^{blk} was calculated through E_{RF}, the latter calculated both by numerically solving the PB equation, λ^{blk}(PB), and by using the TK expression as used in the NMRES model, λ^{blk}(TK). Both of these are pure continuum estimates, where explicit protein or solvent dynamics do not play any direct role. Rather, the static protein conformation used for the calculation matters. The crucial terms in determining λ^{blk} are the actual depth of the red-ox site atoms (that change charge upon ET) from the surface and the amount of charge change on each atom. In this respect, the NMRES estimate employing the TK model is only approximately correct, while a more realistic λ^{blk} estimate is given by λ^{blk}(PB). The estimated λ^{blk}(PB) is about 22.6 kcal/mol. In a previous study, Churg et

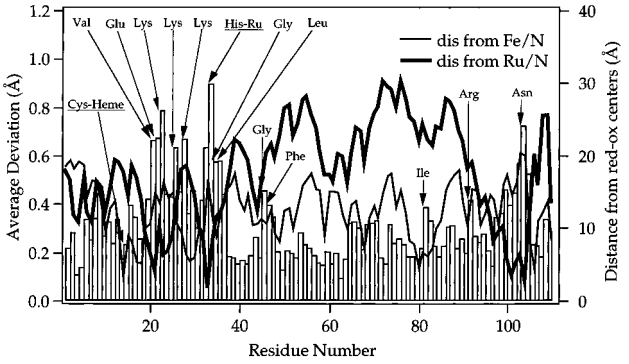

Figure 8. Average deviation of the residues upon ET as predicted by the NMRES model. Distances of the residues from the red-ox centers are also superimposed.

al. determined $\lambda^{blk} \approx 3$ kcal/mol for the self-exchange reaction of Cytc (without the $Ru(NH_3)_5$ moiety) using the microscopic FCLD model.[9] In Ru-Cytc, the extra contribution from the $Ru(NH_3)_5$ moiety, being severely exposed to the solvent as opposed to the fairly shielded heme group, will increase λ^{blk} substantially. In addition, the fact that ΔQ_j is much more distributed over the heme moiety than $Ru(NH_3)_5$ will render the Ru-Cytc λ^{blk}(PB) value much larger than that in native Cytc. When λ^{blk}(PB) was recalculated with no change in the charge distribution on the $Ru(NH_3)_5$ group, it turned out to be 7.0 kcal/mol, higher than the Churg et al. estimate. The discrepancy could arise due to different static protein conformations used by us in addition to using a different method.

Along with λ^{blk}(NMRES), λ^{blk}(PB) yields a total λ of 38.2 kcal/mol, which is about 10 kcal/mol higher than λ_{exp}. Strictly speaking, λ^{prot}(NMRES) and λ^{blk}(PB) are incompatible quantities—one calculated by numerically solving the PB equation and the other obtained by using the TK approximation—and should not be combined. However, the point of this exercise was to establish the reliability of the calculated λ^{prot}(NMRES). If λ^{blk}(PB) is to be taken to be realistic, one notes that the NMRES model overestimates λ^{prot}.

On the other hand, the calculated value of λ^{blk}(TK), 7.2 kcal/mol (for $b = 21$ Å), when combined with λ^{prot}(NMRES) yields a total λ that is about 5 kcal/mol lower than λ_{exp}, suggesting that the NMRES model underestimates λ^{prot}. As pointed out earlier, the spherical cavity assumption in λ^{blk}(TK) models the depths of the crucial red-ox atoms poorly. When λ^{blk}(TK) was calculated for a range of b values, it monotonically increased with decreasing b. The "effective" b at which λ^{blk}(TK) matches λ_{exp} is about 19.5 Å. It is difficult, if not impossible, to unambiguously assign a "correct" value for λ^{blk} since it cannot be accessed directly by experiments.

In light of the above discussion, it is fair to conclude that the NMRES model underestimates λ^{blk} by about 5 kcal/mol. In fact it is known that the TK energies slightly overestimate weak interactions,[25] which might render λ^{prot}(NMRES) a little higher than the actual value. In that case, the underestimation of λ^{blk}(TK) may be even more, as suggested by λ^{blk}(PB). In any case, the focus of this work is not to exactly predict a value of λ^{prot} or λ^{blk}; rather, we want to decompose the λ^{prot} into a frequency spectrum, as done in the next section.

4.4. Contribution from Low-Frequency Modes. Before commenting on the λ_k^{prot}(NMRES) values, it is important to point out the methodological differences that distinguish λ^{blk}(TK) from λ_k^{prot}(NMRES), also calculated using the TK approximation. The solvent relaxation in λ^{blk}(TK) is expressed entirely through the appropriate static and optical dielectric constants of the solvent for a fixed protein conformation through a purely continuum expression. On the other hand, the complementary λ_k^{prot}(NMRES) incorporates protein relaxation through explicit fluctuations of the protein atoms. The individual λ_k^{prot} terms are related to the variation of the energy difference $\Delta E(\sigma_k)$ (which includes contribution from E_{RF}). Even if the estimated $E(\sigma_k)$ is poor due to the spherical cavity shape assumption, the difference, $\Delta E(\sigma_k)$, may not be so.

In the NMA_{TK} model, up to a frequency of 150 cm^{-1}, there are 587 modes. It is interesting to note that these low-frequency modes do not contribute uniformly to λ_{out}^{prot}. Of these modes, 5% of the modes ($\lambda_k^{prot} \geq 0.1$ kcal/mol) contribute 41%, 62% ($\lambda_k^{prot} \leq 0.01$ kcal/mol) of the modes contribute 7%, and the remaining 33% of the modes contribute 52% to the total λ_{out}^{prot}. In other words, a large number of low-frequency modes (62%) do not contribute significantly toward λ_{out}^{prot}. This is clear from Figure 6c,d. Of the modes that contribute significantly, the first, third, and fourth stand out ($\lambda_k^{prot} \geq 0.4$ kcal/mol). In terms of characteristic structural fluctuations, one notes that the relative amplitudes of the heme and the Ru moieties (see Figure 7a) in these modes are high and the direction of motion have a significant radial component. In any case, even the most-coupled mode contributes only 2% of the total λ, indicating that a number of modes, rather than a handful, collectively contribute toward λ. Therefore, structural changes (like mutation) that can affect a few modes will not change λ significantly. Extending this argument, one can also predict that for a given protein system, λ will be invariant if the disposition of the red-ox pair is changed within the protein matrix. This will hold true only when the solvent accessibility of the red-ox sites does not change much. Indeed when Gray and co-workers did experiments with three $Ru(NH_3)_5$-modified Cytc (at His33, 39, and 62), they found λ to be invariant ($\lambda \approx 1.2$ eV).[3]

The contributions of the low-frequency modes (λ_{out}^{prot}) from the NMA_{2R} calculation are barely 1 kcal/mol. This low value almost rules out any role played by the low-frequency protein modes. Besides, having no explicit reaction field energy incorporated in the NMA_{2R} calculation, $\lambda_{out}^{prot}(NMA_{2R})$ lacks any clear physical meaning. It is to be noted that even though the NMA_{2R} model yielded an unacceptably low value of λ_{out}^{prot}, the $\lambda_{in}^{prot}(NMA_{2R})$ values are compatible with $\lambda_{in}^{prot}(NMA_{TK})$. This shows the importance of the reaction field terms in affecting the low-frequency modes preferentially over the high-frequency ones and establishes an important role of E_{RF} in the estimation of λ_{out}^{prot}.

4.5. Structural Changes upon ET. Experimental data on structural changes upon ET are best obtained from X-ray diffraction studies,[26,27] although recently NMR studies have also provided valuable information.[15] For Cytc, conformational changes upon ET have been found to be minor and subtle. A recent high-resolution structural analysis[27] shows retention of all the main-chain to main-chain H-bond interactions in both the oxidation states. There are only few alterations of side-chain−side-chain and side-chain−water H-bond interactions, while the heme is slightly more distorted upon oxidation. Although Ru-Cytc differs from native Cytc by an additional red-ox center, the structural changes, at least around the heme moiety, are expected to be minor.

The coordinates generated by eq 11 yielded an RMSD of 0.31 Å between the reduced and oxidized Ru-Cytc (backbone atoms). The average deviations of the backbone atoms and the heme moiety are 0.28 and 0.24 Å, respectively. These numbers

compare well with X-ray data of yeast iso-1-cytochrome *c* (0.31 and 0.24 Å). On the other hand, the average deviation of the Ru moiety was much higher (0.94 Å). Among all the residues, the dominant deviation comes from the Ru moiety. This is clearly indicated in Figure 8.

In Figure 8 the average deviations of the residues, including the six water molecules, are shown (residues 105−110) along with the distances of the residues from the Fe/N and the Ru/N centers. Apart from the large movement of the Ru-His group, residues that show higher deviation than the average value are proximal to the red-ox moats or charged. The correlation coefficients between the average deviation and residue−red-ox site distances are −0.13 (Fe/N) and −0.54 (Ru/N), respectively. Thus the charge change on the Ru moiety brings about a larger effect on the equilibrium atomic positions than that on the Fe moiety. Qualitatively this can be justified from the fact that atoms around the Fe/N center, being more in the interior of the protein, are packed and are in general not charged. On the other hand, residues close to the Ru/N center are more exposed and so tend to be charged and packed less well. In addition, the changed charge distribution is more distributed over the heme moiety than the Ru/N center, making the latter more susceptible to exert conformational perturbation. Apart from such qualitative observations, we did not observe any subtle structural changes that were conspicuous and showed any significant mechanistic importance.

Although a comparison of the results on Ru-Cytc obtained from the NMRES model and X-ray or NMR data (on native Cytc) is not strictly valid, the general trend of structural changes observed in the experimental data is expected to be present in Ru-Cytc, too. They constitute subtle changes, such as the reorientation of a water molecule or a side chain to form or break an H-bond. Capture of any such change in a simulation can be highly sensitive to the potential energy function used and the presence or absence of explicit solvent molecules. Moreover, it is beyond the scope of NMA, valid only for small fluctuations around a MEC, to reflect these changes accurately. Our observation is that although the NMRES model can predict reasonable values of λ, it fails to capture subtle and probably important structural changes. A better way to model such subtle conformational changes will probably come from MD simulations as mentioned in the next section.

5. Summary

In a preceding paper the NMRES model[14] was proposed that connected harmonic protein fluctuations and their coupling to ET through the protein reorganization energy spectrum. In this paper the NMRES model has been applied to Ru-modified (at His33) cytochrome *c*, a semisynthetic protein, studied extensively by Gray and co-workers ($\lambda_{exp} \approx 28$ kcal/mol). The present work was mainly intended to address two issues. By performing an actual simulation on a system with known λ, we wanted to check the reliability of the NMRES model in predicting the experimental λ. The other aim was to analyze the breakup of this total λ into the protein (λ^{prot}) and bulk solvent contribution (λ^{blk}) and further look into the role played by the individual protein modes through the mode-specific terms λ_k^{prot}.

The entire analysis was performed for two different electrostatic models, one with (NMA_{TK}) and the other without (NMA_{2R}) the inclusion of E_{RF}. In terms of the frequency spectrum, the two electrostatic models yielded very similar results, with the former exhibiting a slight shift toward softer modes. On the other hand, when λ^{prot} was estimated, NMA_{TK} and NMA_{2R} yielded significantly different values.

The cumulative λ_k^{prot} contribution from all modes yielded $\lambda^{prot}(NMA_{2R}) \approx 3.4$ kcal/mol and $\lambda^{prot}(NMA_{TK}) \approx 15.6$ kcal/mol. The accompanying λ^{blk}(TK), an underestimate due to poor modeling of charge depths, was 7.2 kcal/mol. When λ^{blk}(PB) was calculated by numerically solving the PB equation, we obtained 22.6 kcal/mol. Assuming the real λ^{blk} to be between these two limits, it was concluded that the estimated λ^{prot}-(NMA_{TK}) value was reasonable.

The inner sphere and outer sphere contributions to λ^{prot} were separated by considering modes above 150 cm^{-1} to be high-frequency modes. While the NMA_{TK} model yielded $\lambda_{out}^{prot} \approx 12$ kcal/mol, the NMR_{2R} model predicted about 1 kcal/mol for λ_{out}^{prot}. Surprisingly, the λ_{in}^{prot} contributions from both the electrostatic models yielded reasonable numbers (≈3 kcal/mol). We conclude that the solvent effect through the E_{RF} terms is crucial in determining λ_{out}^{prot}, whereas its role in λ_{in}^{prot} is not so vital. It should be pointed out that our calculated λ_{in}^{prot} considers all the high-frequency modes—originating from the red-ox sites as well as the protein matrix and it includes E_{RF} as well—as opposed to vacuum calculations performed on isolated red-ox moats. This way of partitioning λ^{prot} is more natural.

Among the low-frequency modes, about 60% did not contribute significantly to λ_{out}^{prot}. The rest of the modes contributed more or less evenly to λ_{out}^{prot}, although there were a subgroup that contributed relatively more than the rest. Extrapolating this to be a general trend in ET proteins, one can conclude that minor changes affecting the low-frequency modes will not affect λ_{out}^{prot}, indirect proof for which comes from the observation of λ_{exp} invariance in Ru-Cytc where the protein was modified at various His sites by Ru.[3] The magnitude of structural changes estimated from the NMRES model was compatible with experimental data on native Cytc. However, subtle structural changes that could be mechanistically important and have been observed experimentally[15,27] were not observed in this work.

In summary, we successfully demonstrated the applicability of the NMRES model to an ET protein in this paper. The underlying philosophy behind the NMRES model was to deconstruct the protein matrix coupling to ET into orthogonal collective mode components. Use of NMA was most straightforward toward achieving this goal, as has been done in this paper. However, it is clear that certain aspects of the protein/solvent involvement cannot be addressed reliably if one is restricted to NMA in modeling the collective modes. Specifically, with respect to capturing mechanistically important structural changes and estimating λ^{blk}, NMRES was not reliable. Conformational changes can often involve multiminima energy surfaces, inherently absent in NMA treatment. The spherical protein cavity assumption inherent in the TK model is responsible for poor estimation of λ^{blk}. One way to circumvent the first problem is to perform an MD simulation followed by a principal component analysis,[22] a more realistic approach to capture collective motion of proteins. Within the continuum model, the best estimate of λ^{blk} will come only from numerically solving the PB equation. Computationally this is still a major problem when incorporated with MD simulation of proteins. The other alternative is to use a spherical cavity that contains the protein and several solute molecules surrounded by a continuum with an appropriate boundary potential. This strategy has been demonstrated to work well by Beglov and Roux.[24] In the next phase of our work on protein ET, we plan to alter the NMRES model along these lines.

Acknowledgment. This work was supported from funds provided by the Ministry of Education and Culture, Japan, the Japanese Society for Promotion of Science (JSPS), and the Human Frontier Science Program Organization (HFSPO). A part of the computation was done at computer centers of the Institute for Molecular Science and of the Japan Atomic Energy Research Institute. G.B. was the recipient of PD fellowships from HFSPO and JSPS and acknowledges continued support from CSIR, India. We thank Dr. J. Wand for providing the coordinates of Cytc and B. Roux for providing the FORTRAN routine SSBP.

Appendix. Energies and Gradients of E_{RF}(TK)

The reaction field energy term E_{RF} in the TK[20] model is given by

$$E_{RF} = -\frac{1}{b}\sum_{i>j} e_i e_j B_{ij} - \frac{1}{2b}\sum_i e_i^2 B_{ii} \quad \text{(A1)}$$

where

$$B_{ij} = \frac{1}{D_{in}}\sum_{n=0}^{\infty}\frac{(n+1)(D_{out}-D_{in})}{(n+1)D_{out}+nD_{in}}\left(\frac{r_i r_j}{b^2}\right)^n P_n(\cos\theta_{ij}) \quad \text{(A2)}$$

where b is the radius of the spherical cavity, r_i and r_j are the radial distances of atoms i and j, $\cos\theta_{ij}$ is the cosine of the angle between atoms i and j with respect to the center of the cavity, and $P_n(\cos\theta_{ij})$ are the Legendre polynomials of order n. Provided that no atom lies very close to the boundary, the infinite sum in B_{ij} in eq A2 converges very rapidly and the reaction field electrostatic energies can be calculated with the inclusion of the first few terms in the infinite sum of B_{ij}.

To perform the energy minimization with the inclusion of B_{ij} in the usual force field, one needs the derivatives arising from

$$V_{ij} = (r_i r_j)^n P_n(\cos\theta_{ij}) \quad \text{(A3)}$$

Here we note that mass center of the molecule, C_x, C_y, C_z, is constrained always at the center of the cavity. Therefore, the radial distance of atom i, r_i, and the cosine of the angle between atoms i and j, $\cos\theta_{ij}$, are given by

$$r_i = \{(x_i - C_x)^2 + (y_i - C_y)^2 + (z_i - C_z)^2\}^{1/2} \quad \text{(A4)}$$

and

$$\cos\theta_{ij} = \{(x_i - C_x)(x_j - C_x) + (y_i - C_y)(y_j - C_y) + (z_i - C_z)(z_j - C_z)\}/r_i r_j \quad \text{(A5)}$$

where

$$C_x = \sum_i m_i x_i / M, \quad M = \sum_i m_i \quad \text{(A6)}$$

Here m_i is the mass of atom i and M is the total mass of the molecule.

Calculation of derivatives of r_i and $\cos\theta_{ij}$ with respect to the Cartesian coordinate of atoms is done in two steps. First, derivatives are calculated without constraining the mass center at the center of the cavity. This means that C_x, C_y, C_z in eqs A3 and A4 are treated as constants for the derivative calculation. Derivatives calculated in this treatment will be denoted with subscript u (for unconstrained). Thus, writing only the x-component, we have

$$\left(\frac{\partial r_i}{\partial x_i}\right)_u = \frac{1}{r_i}(x_i - C_x) \quad \text{(A7)}$$

$$\left(\frac{\partial \cos\theta_{ij}}{\partial x_i}\right)_u = \frac{1}{r_i r_j}(x_j - C_x) - \frac{\cos\theta_{ij}}{r_i^2}(x_i - C_x) \quad \text{(A8)}$$

At the next step, derivatives are calculated with constraining the mass center at the center of the cavity. Such derivatives will be denoted with subscript c (for constrained). Thus, we have

$$\left(\frac{\partial r_i}{\partial x_k}\right)_c = \left(\delta_{ik} - \frac{m_k}{M}\right)\left(\frac{\partial r_i}{\partial x_i}\right)_u \quad \text{(A9)}$$

$$\left(\frac{\partial \cos\theta_{ij}}{\partial x_k}\right)_c = \left(\delta_{ik} - \frac{m_k}{M}\right)\left(\frac{\partial \cos\theta_{ij}}{\partial x_i}\right)_u + \left(\delta_{jk} - \frac{m_k}{M}\right)\left(\frac{\partial \cos\theta_{ij}}{\partial x_j}\right)_u \quad \text{(A10)}$$

By using eqs A9 and A10, the derivative of V_{ij} of eq A3 is obtained as follows:

$$\left(\frac{\partial V_{ij}}{\partial x_k}\right)_c = \left(\delta_{ik} - \frac{m_k}{M}\right)\left(\frac{\partial V_{ij}}{\partial x_i}\right)_u + \left(\delta_{jk} - \frac{m_k}{M}\right)\left(\frac{\partial V_{ij}}{\partial x_j}\right)_u \quad \text{(A11)}$$

where

$$\left(\frac{\partial V_{ij}}{\partial x_i}\right)_u = n(r_i r_j)^n\left[P_n(\cos\theta_{ij})\frac{1}{r_i^2}(x_i - C_x) + Q_n(\cos\theta_{ij})\left\{\frac{1}{r_i r_j}(x_j - C_x) - \frac{\cos\theta_{ij}}{r_i^2}(x_i - C_x)\right\}\right] \quad \text{(A12)}$$

with

$$Q_n(z) \equiv P_n'(z) = \frac{n\{zP_n(z) - P_{n-1}(z)\}}{z^2 - 1} \quad \text{(A13)}$$

The derivative of eq A12 is calculated in the SSBP routine supplied by Roux.[24]

References and Notes

(1) *Protein Electron Transfer*; Bendall, D. S., Ed.; BIOS Scientific Publishers: Oxford, 1996.
(2) DeVault, D. *Quantum-Mechanical Tunneling in Biological Systems*; Cambridge University Press: London, 1981.
(3) Winkler, J. R.; Gray, H. B. *Chem. Rev.* **1992**, *92*, 369−379.
(4) Onuchic, J. N.; Beraton, D. N.; Winkler, J. R.; Gray, H. B. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 349−377.
(5) Kuki, A. Electron Tunneling Paths in Proteins. In *Long-Range Electron Transfer in Biology*; Palmer, G., Ed.; Springer-Verlag: Berlin, 1991; Vol. 75; pp 49−83.
(6) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 966−978.
(7) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 979−989.
(8) Brunschwig, B. S.; Ehrenson, S.; Sutin, N. *J. Phys. Chem.* **1986**, *90*, 3657−3668.
(9) Churg, A. K.; Weiss, R. M.; Warshel, A.; Takano, T. *J. Phys. Chem.* **1983**, *87*, 1683−1694.
(10) Warshell, A.; Parson, W. W. *Annu. Rev. Phys. Chem.* **1991**, *42*, 279−309.
(11) Marchi, M.; Gehlen, J. N.; Chandler, D.; Newton, M. *J. Am. Chem. Soc.* **1993**, *115*, 4178−4190.
(12) Schulten, K.; Tesch, M. *Chem. Phys.* **1991**, *158*, 421−446.
(13) Xu, D.; Schulten, K. *Chem. Phys.* **1994**, *182*, 91−117.
(14) Basu, G.; Kitao, A.; Kuki, A.; Go, N. *J. Phys. Chem. B* **1998**, *102*, 2076.
(15) Xiurong, P.; Jeffrey, Q.; Urbauer, J. F.; Fuentes, E. J.; Leopold, M. F.; Wand, J. A. *Nature Struct. Biol.* **1994**, *1*, 378−382.
(16) Northrup, S. H.; Pear, M. R.; Morgan, J. D.; McCammon, A. J. *J. Mol. Biol.* **1981**, *153*, 1087−1109.

(17) Gruschus, J. Cornell University, 1993.
(18) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(19) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(20) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.
(21) Morikami, K.; Nakai, T.; Kidera, A.; Saito, A.; Nakamura, A. *Comput. Chem.* **1992**, *16*, 243−248.
(22) Kitao, A.; Hirata, H.; Go, N. *Chem. Phys.* **1991**, *158*, 447−472.
(23) Horiuchi, T.; Go, N. *Proteins* **1991**, *10*, 106−116.
(24) Beglov, D.; Roux, B. *J. Phys. Chem.* **1994**, *100*, 9050−9063.
(25) Gilson, M. K.; Honig, B. H. *Proteins* **1988**, *3*, 32−52.
(26) Takano, T.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6371−6375.
(27) Berghuis, A. M.; Brayer, G. D. *J. Mol. Biol.* **1992**, *223*, 959−976.