INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATICNAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/642 -14

College on Methods and Experimental Techniques in Biophysics

28 September - 23 October 1992

Electron Transfer in Proteins

J.N. ONUCHIC
University of California, U.S.A.

These are preliminary lecture notes, intended only for distribution to participants.

Annu. Rev. Biophys. Biomol. Struct. 1992. 21:349–77

PATHWAY ANALYSIS OF PROTEIN ELECTRON-TRANSFER REACTIONS

José Nelson Onuchic[1]

Department of Physics, University of California, San Diego, La Jolla, California 92093

David N. Beratan[2]

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, and Beckman Institute, California Institute of Technology, Pasadena, California 91125

Jay R. Winkler and Harry B. Gray

Beckman Institute, California Institute of Technology, Pasadena, California 91125

KEY WORDS: electron-tunneling pathways, electron coupling, ruthenium-modified proteins, cytochrome *c*, myoglobin

---

CONTENTS

[1] This review was begun when JNO was in residence at the Instituto de Física e Química de São Carlos, Universidade de São Paulo, 13560 São Carlos, S.P., Brazil.

[2] Present address: Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

1056–8700/92/0610–0349$02.00

PERSPECTIVES AND OVERVIEW

One of the central challenges in molecular biophysics is to understand how proteins control biochemical reactions in living organisms. In their folded states, proteins exhibit a variety of structural fluctuations. The question before us is; how do protein structure and dynamics control biological function? Our goal is to develop tools that allow us to simulate and understand those aspects of biomolecular structure and dynamics that establish the unique capabilities of these molecules. Our hope is to arrive at a deeper understanding of the mechanisms that control biochemical reactions and to establish design criteria for new proteins that will perform specific tasks.

With these issues in mind, this review focuses on electron transfer reactions. These reactions are extremely important in biology, particularly in bioenergetic reaction pathways (23, 32, 72). For example, in the early steps of photosynthesis, high efficiency solar-energy conversion is achieved with a complex of protein-bound electron donors and acceptors. Control of charge separation and recombination rates is required to insure that productive forward electron-transfer reactions within this complex occur rapidly, while wasteful back reactions occur orders of magnitude more slowly. Initial light-driven transfer steps are complete within 3 ps, and charge separation is subsequently stabilized for tens of milliseconds with a quantum efficiency near 100%. A comparable selective acceleration of electron transfer reactions has not been achieved in any artificial system.

Our goal in this paper is to present the results of a collaboration between theory and experiment aimed at developing a computational design capability for electron-transfer proteins. The theoretical methods we describe contain the minimal description that is needed to model adequately the fundamental mechanisms of protein-mediated electron tunneling. Although the description does not include every detail of the protein electronic structure, the model makes concrete, testable predictions about primary, secondary, tertiary, and quaternary structural effects on electron-transfer rates. Measurements of electron transfer rates in ruthenium-modified (ruthenated) proteins test the method's reliability. We find that

inclusion of protein features neglected in structureless barrier models is essential for understanding the observed transfer rates in these systems.

ELECTRON TUNNELING MATRIX ELEMENTS

This review focuses on the calculation of tunneling matrix elements (T_{DA}) and the comparison of these couplings with those derived from experiment (28, 38, 42, 51, 60). The tunneling matrix element is associated with the weak long-distance electronic coupling between donor (D) and acceptor (A) mediated by the protein. Electron-transfer rates in the proteins discussed here are in the nonadiabatic limit and are therefore proportional to T_{DA}^2. In this section, we present a short discussion of the dynamical limits associated with the nonadiabatic electron-transfer rate formulation. Special attention is given to the Hamiltonian that we use to describe our problem and why the nonadiabatic limit is appropriate for long-distance electron transfer in proteins.

We begin our discussion by presenting the Hamiltonian that has been used extensively for the generic electron transfer problem (12, 34, 52, 65, 67):

$$\mathscr{H}_{ET} = T_{DA}(\mathbf{Q})\sigma_x + \tfrac{1}{2}[\alpha_D^{eff}(\mathbf{Q}) + \alpha_A^{eff}(\mathbf{Q})] + \tfrac{1}{2}[\alpha_D^{eff}(\mathbf{Q}) - \alpha_A^{eff}(\mathbf{Q})]\sigma_z + \mathscr{H}_Q. \quad 1.$$

The terms σ_x and σ_z are the Pauli matrices, where the expression $\sigma_z = 1$ or -1 is associated with the donor- or acceptor-localized state, respectively. $\mathscr{H}_Q$ supplies the dynamics for the nuclear coordinates ($\mathbf{Q}$), and $\alpha_D^{eff}(\mathbf{Q})$ [$\alpha_A^{eff}(\mathbf{Q})$] is the instantaneous energy for the reactants (products) state.

Two major aspects of this Hamiltonian should be considered. First, it is necessary to describe why a multisite many-electron Hamiltonian can be reduced [renormalized to an effective two-level one-electron system (Equation 1)]. Second, if this renormalization is valid, we must present the conditions for the electron transfer rate to fall in the nonadiabatic limit (28, 51), i.e.

$$k_{ET} = \frac{2\pi}{\hbar} T_{DA}^2 (FC), \quad 2.$$

where (FC) is the nuclear (or Franck–Condon) factor. The analysis of experiments presented in this review relies on the separability of the rate expression.

Ideally, we would describe the molecular system from first principles including the motion of all the electrons and nuclei. Because this task is impossible, our strategy is to break the problem into pieces that can be

understood. To be successful, such a simplification relies on the identification of the relevant energy scales of the problem.

Before addressing the details of the molecular electronic structure, we assume that the Born-Oppenheimer approximation is valid. This assumption is appropriate if the energies for nuclear excitations are much smaller than those for electronic excitations. We comment later about the important excitations in the electron-tunneling problem and why we believe this assumption is valid. The electronic energies of chemical bonds are much smaller than the electronic excitation energies of core electrons. We can therefore describe our problem as valence electrons moving in a pseudopotential provided by the core electrons and nuclei. Actually, we can expand this picture by assuming that the energy associated with electronic coupling between atoms (or bonds) is small compared to the energy of excited states on isolated atoms, leading to a tight-binding or extended-Hückel picture (for details, see 6, 10, 12, 58).

The initial tight-binding electronic Hamiltonian for D, A, and their bridge is written (4, 6, 12, 27, 47, 48, 58, 59, 63, 64, 66, 70, 71):

$$\mathscr{H}_{el} = \alpha_D a_D^\dagger a_D + \alpha_A a_A^\dagger a_A + \sum_{i_D} v_{D,i_D}(a_D^\dagger a_{i_D} + a_{i_D}^\dagger a_D) + \sum_{i_A} v_{A,i_A}(a_A^\dagger a_{i_A} + a_{i_A}^\dagger a_A) + \sum_i \alpha_i a_i^\dagger a_i + \sum_{i,j>i} v_{ij}(a_i^\dagger a_j + a_j^\dagger a_i), \quad 3.$$

where the $a_\mu^\dagger$ (a_μ) creates (destroys) an electron on the μth orbital. The first two terms in the Hamiltonian represent the donor and acceptor sites. The third and fourth terms contain the coupling between the donor and acceptor, respectively, and the bridge. Bridge orbitals coupled to the donor and acceptor are labeled i_D and i_A, respectively. The last two terms are the bridge Hamiltonian. Because we are using the Born-Oppenheimer approximation, all the electronic energies (α and v) are a function of the nuclear configuration **Q**.

How do we reduce the above Hamiltonian (Equation 3) to a two-level system (reactants and products)? One way is to use the Löwdin partitioning technique (47, 59). With this method, one maps an eigenvalue problem of high dimension onto an equivalent problem of lower dimension. The Hamiltonian or Equation 3 in matrix notation is:

$$\begin{pmatrix} \mathscr{H}_{DA} & \mathscr{H}_{DA,B} \\ \mathscr{H}_{B,DA} & \mathscr{H}_{bridge} \end{pmatrix}, \quad 4.$$

where $\mathscr{H}_{DA}$ is the matrix Hamiltonian that only includes the donor and acceptor sites. The direct coupling between D and A in the case of long-distance transfer is negligible. $\mathscr{H}_{bridge}$ is the Hamiltonian matrix for the bridge, and $\mathscr{H}_{B,DA}$ is the matrix that couples the donor and acceptor to the bridge. Löwdin diagonalization yields a reduced 2 × 2 matrix

$$\tilde{\mathscr{H}}_{DA} = \mathscr{H}_{DA} - \mathscr{H}_{DA,B}\mathscr{H}_{bridge}^{-1}\mathscr{H}_{B,DA}. \quad 5.$$

The effective matrix one obtains is

$$\tilde{\mathscr{H}}_{DA} = \begin{bmatrix} \alpha_D^{eff}(E) & T_{DA}(E) \\ T_{AD}(E) & \alpha_A^{eff}(E) \end{bmatrix}, \quad 6a.$$

where

$$\alpha_{D(A)}^{eff}(E) = \alpha_{D(A)} + \Delta_{D(A)}(E), \quad 6b.$$

$$\Delta_{D(A)} = \sum_{i,j} v_{D(A)i} G_{ij}(E) v_{jD(A)}, \quad 6c.$$

and

$$T_{DA} = \sum_{i,j} v_{Di} G_{ij}(E) v_{jA}. \quad 6d.$$

The is and js in the sums run over the bridge orbitals. G is the Green's function for the bridge, i.e. the Green's function (7, 25, 26, 35, 50, 52, 66, 67a) associated with $\mathscr{H}_{el}$ without the donor and acceptor terms, $G = (\mathscr{H}_{bridge} - E)^{-1}$.

Equation 6 is equivalent to Equation 5, i.e. the eigenvalues for the two equations are the same. However, we are only interested in the two states that define the two-level system. The first step in analyzing these states is to determine the tunneling energy. The effective donor energy can be obtained from Equation 6 by solving

$$\bar{\alpha}_D = \alpha_D^{eff}(\bar{\alpha}_D). \quad 7.$$

The root of this equation closest to α_D is the effective donor energy. This result is equivalent to the one used in our laboratory (see 4, 65, for example) when considering the isolated donor-plus-bridge system. A similar calculation can be performed for the acceptor. The tunneling energy, E_T, is obtained for the nuclear configuration **Q** where

$$E_T = \bar{\alpha}_D = \bar{\alpha}_A. \quad 8.$$

After calculating the tunneling energy, we can finally obtain the two-level system by fixing the value of E in Equation 6 equal to E_T. How good is this approximation? Let us refer to the symmetric and antisymmetric state energies that define the two-level system as E_1 and E_2, respectively. $E_2 - E_1$ is twice the tunneling matrix element. E_T is exactly midway between E_2 and E_1. Therefore, if we compare E_2 and E_1 as the eigenvalues of

Equation 6, setting $E = E_T$, we introduce errors of the order $E_2 - E_1$. Because E in Equation 6 only appears in terms like $\alpha_i - E$, the error introduced is approximately $(T_{DA}/|\alpha_{bridge} - E_T|)$. This error is of the order of the overlap between the effective donor and the acceptor states.

In order for the two-level approximation to hold, the separation between levels one and two, $2T_{DA}$, must be small compared to the energy separation between these states and the bridge. Actually, the ratio of these two quantities determines the precision of the approximation. Also, for the Born-Oppenheimer approximation to hold, these energy separations must be large compared to any relevant nuclear excitation energies. Finally, for this approximation to be valid, the investigator must consider one more time (energy) scale. As the electron tunnels from the donor to the acceptor, it spends a certain time in the classically forbidden region (12, 18). If this time is much shorter than the period of the vibrational modes, the atoms stay fixed as the electron tunnels; in other words, the Born-Oppenheimer approximation works. These approximations are reasonably good for electron transfer in proteins, and the reader is referred elsewhere (5, 12, 65) for further details.

To conclude this section, we comment on the nonadiabatic approximation that leads to an electron-transfer rate given by Equation 2. In order for this limit to be valid, the electronic frequency, $T_{DA}/\hbar$, must be low compared to that of the relevant nuclear motion. In the past six years, many papers have addressed this subject (see 67 and references therein for details). In long-distance electron transfer, the tunneling matrix elements are so small that this approximation most likely is adequate.

THE PATHWAY MODEL

The pathway model of electronic coupling in proteins (3, 6, 7, 8, 11) was developed based on earlier studies of electronic coupling in model compounds (4, 7, 63). Tunneling is much more efficient (decays more slowly) through bonded orbitals than through space, because the potential barrier is effectively lower. In proteins, the bonded-path connection length between D and A can be extremely long compared with the direct through-space distance. Our pathway method searches for the combination of bonded and nonbonded interactions that maximizes the total D-A interaction mediated by a combination of through-bond and through-space coupling through the protein. The tunneling pathways obtained contain mostly bonded interactions (with occasional through-space connections).

The intervening protein could provide two distinct mediation mechanisms to couple D and A. One mechanism mediates the interaction by a few very specific combinations of interacting bonds (fragments of amino

acids) between D and A. The bonds would couple D and A through a sequence of directly connected covalent bonds, hydrogen bonds, and noncovalent contacts. Each of these combinations is called a *physical tunneling pathway* and plays a role in the D-A coupling (8). The other distinct way that the protein might couple D and A involves a sufficiently large number of pathways such that modifying a single pathway in this network will have a very small effect on the net coupling and the rate. In this case, no particular detail of the protein will greatly affect the rate.

An elaboration of the discussion of a physical tunneling pathway can help us focus the discussion of the D-A coupling mechanism. For a single physical pathway, one can use exact and perturbation theory methods for calculating the coupling arising from that physical pathway. Numerical strategies (for both exact and perturbation methods) usually write the decay of the wave function as a product of decays per bond [or delocalized group (63, 64)]. Within a lowest-order perturbation theory calculation, the per-bond decay depends only on the tunneling energy and on the nature of the particular bonds in the pathway. This method (applied to lowest order) neglects scattering corrections to the wave function propagation in the protein bridge. The scattering corrections (equivalent to higher-order perturbation-theory corrections) for a given pathway arise from enumerations of bonds in the tunneling pathway longer than the shortest path from D to A. For example, a physical pathway consisting of bonds 1, 2, 3, 4, . . . has the direct pathway 1-2-3-4 . . . and the scattering pathways 1-2-3-2-3-4 . . . , etc. We now discuss how one can exactly account for the scattering pathways in the electronic-coupling calculation for a one-dimensional physical pathway by correcting the self energy of each orbital on the path. Exact methods, particularly Green's function approaches, often write the coupling as a product as well. In this case, the terms in the product explicitly include these scattering corrections. In the same way, the effect of side groups appended to the physical pathway can also be included.

For a single physical pathway, the tunneling matrix element can be written (7, 53)

$$t_{DA} = \text{prefactor} \prod_{i=1}^{N} \varepsilon_i. \qquad 9.$$

Neglecting interactions between pathways within the protein bridge, T_{DA} is a sum over t_{DA}s for all physical pathways. For a pathway, ε_i for each block in the path (66) may be calculated approximately or exactly as discussed above. The prefactor depends on details of the interaction between the D or A with the first or last, respectively, bond of the tunneling

pathway. When experimental systems with similar (or properly scaled) prefactors and *FC* factors are compared, differences in electron-transfer rates are expected to result from differences in the coupling via the physical pathways of the systems. The challenge in proteins, then, is to identify the chains of orbitals that define dominant pathways. The dominant tunneling pathways correspond to the combinations of bonds in the protein that maximize the products in Equation 9.

As an example of how to compute ε_i, we consider a linear chain of identical (Figure 1*a*) orbitals coupling the donor and the acceptor. (The orbital energy is α_B and the coupling between neighbors is v.) If backscattering is neglected, the decay per orbital is

$$\varepsilon = v/(E_T - \alpha_B), \qquad 10.$$

where v is the coupling between neighboring bridge orbitals, and α is the orbital energy. The exact result, including backscattering, can also be written as the product given by Equation 9. The decay ε_i between bonds i and $i+1$ is:

$$\varepsilon_i(E_T) = \frac{G^{i+1}_{1,i+1}}{G^{i}_{1,i}} = \frac{v}{E_T - (\alpha_B + \delta_i^{bs})}, \qquad 11.$$

where δ_i^{bs} is the site self-energy correction due to backscattering. G^j is the Green's function for a linear bridge of j orbitals. In the long chain limit ($i \gg 1$), this result converges to the infinite-chain limit

$$\varepsilon^{\infty}_{exact}(E) + \frac{1}{\varepsilon^{\infty}_{exact}(E)} = \frac{E - \alpha_B}{v}. \qquad 12.$$

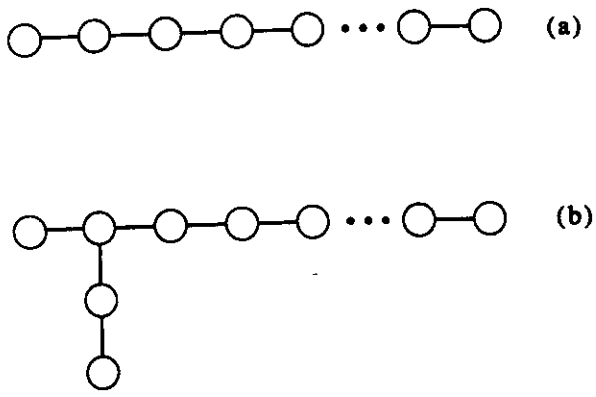


Figure 1 (*a*) Schematic representation of a linear bridge. Only nearest neighbors are coupled. (*b*) Schematic representation of a side group coupled to an orbital in the pathway.

We refer the reader elsewhere (66) for details about how these Green's functions can be calculated and for a description of our stepwise Green's function method.

The effect of pendant (side) groups can also be included without destroying the pathway concept. This concept is associated with the possibility of writing t_{DA} as a product of εs. Figure 1*b* suggests that pathways can include the effects of side groups attached to a single site by modifying the self energy of the orbital that the pendant group is attached to. Assuming that the side chain is coupled to pathway orbital i via orbital $s1$, the side chain can be eliminated by renormalizing the orbital i energy:

$$\alpha_i^{eff} = \alpha_i + v_{i,s1} G^{\infty}_{s1,s1} v_{s1,i}. \qquad 13.$$

$G^{\infty}_{s1,s1}$ is the diagonal matrix element of the Green's function for site $s1$ when only the side chain is included in the Hamiltonian. Using this procedure, many side chains can be immediately eliminated at the early stages of the calculation, greatly simplifying the problem.

The validity of the pathway approximation only becomes suspect when loops involving several paths appear. If interference between pathways is considerable, contributions from independent pathways enter T_{DA} in a rather complex manner. To address this issue, we developed a stepwise Green's function technique, and research is underway to understand the general applicability of the pathway concept. The simple pathway concept without the inclusion of effects like those discussed above can still teach us much about the mediation of electron tunneling in proteins.

Our strategy for mapping tunneling pathways in proteins involves making approximations to the decay factors ε_i and performing computer searches for the combination of interacting bonds with decay factors that maximize the product in Equation 9.

PATHWAY SEARCH STRATEGIES

The Conceptual Basis of the Calculations

The single maximum coupling pathway between two points in a protein indicates, at the very least, the coupling strength between those regions of the molecule. This section describes our simple approximations (based on intuition gained from model compound studies) used to produce a computationally tractable approximation to the coupling given in Equation 9 in which a product of decay factors give the contribution to T_{DA} from a single pathway. Our ansatz partitions electronic mediation through protein into three types of interactions: covalent, hydrogen-bonded, and through-space. This division was based on the fact that bond-mediated interactions are much longer range than through-space interactions (7).

The barrier to tunneling through a bonded medium is considerably lower than tunneling through vacuum (6, 10, 11) so the exponential decay of bond-mediated coupling is slower than through-space tunneling. Hydrogen bonds are weaker than covalent bonds, so it was not immediately apparent that they would be key mediators of tunneling. However, because hydrogen bonds bring lone-pair and bonding orbitals into close proximity, we expect their mediation properties to be substantial (8, 64).

Bonded and nonbonded interaction energies are obviously a function of atom type, hybridization, and orientation. However, the distinction between bonded and nonbonded interactions is so strong that a preliminary understanding of coupling pathways arises from determining the mix of these interactions on the dominant routes. For a single pathway consisting of covalent (C), hydrogen-bonded (H), and through-space (S) interactions, Equation 9 can be rewritten:

$$T_{DA} \propto \prod_i \varepsilon_C(i) \prod_j \varepsilon_S(j) \prod_k \varepsilon_H(k). \qquad 14.$$

Because the rate of electron transfer (Equation 2) is proportional to T_{DA}^2, we can estimate relative rates from Equation 14 for a given nuclear *FC* factor. By writing the T_{DA} expression with a proportionality, we have suppressed prefactors associated with D-bridge and bridge-A coupling. These factors have been discussed elsewhere (22, 64) and for the purposes of this discussion are assumed to be the same for all pathways. Simpler models for electron tunneling in proteins would write T_{DA} (and the transfer rate) as proportional to an exponentially decaying factor arising from a simple one-dimensional square barrier (28):

$$k_{ET}(\text{square}) = A \exp(-\beta R)(FC). \qquad 15.$$

The goal of the algorithm described in the next section is to find the combination of bonds between D and A that maximizes the product in Equation 14 given simple rules for approximating the decay factors ε. Other theoretical strategies for calculating the tunneling matrix element are also being developed (14, 19, 46).

Coupling Decay Factors

We now consider the range of decay parameters that are chemically accessible and describe the computer-search strategy for finding pathways that maximize the product in Equation 14 for a set of specified decay factors. Many covalently coupled D-A model compounds that undergo photoinduced electron transfer have been constructed with both biological and nonbiological redox active chromophores. When one translates the reported decays of rate with bridge size to decay per bond factors of the

tunneling matrix element, through-bond ε_C decay factors are calculated in the range ~0.7–0.4 (55). We have chosen a value of 0.6 because it is a reasonable average value for the decay per bond (see 64 for details). Although ratios of rates depend on the choice, if all ε_Cs are assumed to be the same, the qualitative results of the single pathway calculations are insensitive to the exact value chosen (because ε_C appears as a prefactor in all three terms). The key relationship is between ε_C and the through-space decay constant. Through-space interactions are treated as stretched bonds, with couplings that are weaker than the bonded couplings by an amount commensurate with the length of the interaction beyond the reference covalent-bond length. An additional factor, usually taken as 1/2, is added to account for the generally unfavorable orientation effects associated with through-space interactions. The decay length, 1.7 $Å^{-1}$ for the through-space interaction, arises from the calculation of penetration through a one-dimensional square barrier, which drops with exponential decay constant $(2m_eE_B/\hbar^2)^{1/2}$, where m_e is the electron mass and E_B is the tunneling electron energy, about 10 eV (11). Tunneling energies chosen in the 5–10 eV range have been explored. Again, the results are insensitive to the specific value. The hydrogen-bond decay is treated as two covalent bonds from heteroatom to heteroatom, allowing one to adjust the coupling if the bond length is longer or shorter than the reference length. Thus, we have arrived at the following parameter set (8, 9):

$$\varepsilon_C = 0.6 \qquad 16a.$$

$$\varepsilon_H = \varepsilon_C^2 \exp[-1.7(R-2.8)] \qquad 16b.$$

$$\varepsilon_S = (1/2)\varepsilon_C \exp[-1.7(R-1.4)]. \qquad 16c.$$

In these expressions, the distances, R, are in Å units and the decay factors, ε, are unitless. The reference covalent bond distance is chosen as 1.4 Å (2.8 Å for two bonds). These decay factors include the minimal amount of physical detail needed to understand the structural dependence of electronic coupling in a bridge. As such, they provide a starting point for the development of structure-function relationships that, if promising, will be elaborated to include numerous fascinating complications arising from quantum interference within and between pathways, bond energetic differences, and geometric fluctuations from assumed atomic positions, to name a few.

Finding the Best Path

How are the pathway searches actually performed? These parameters are consistent with typical binding energies for electron-transfer localized

states as well as theoretical and experimental studies of model compounds (9). Each decay factor ε is associated with an effective distance d_{eff} where:

$$d_{\text{eff}}(i) = -\log \varepsilon(i). \qquad 17.$$

We refer both to decay factors and connection lengths throughout the paper. The strength of the coupling arising from a single (noninterfering) pathway is proportional to the product of decay factors for each step on the path: $\Pi_i \varepsilon_i$. The computational challenge is to analyze the highly interconnected network of bonded and nonbonded contacts in a protein and specify the bonds that maximize this product. This is precisely the well-known minimum-distance-in-a-graph problem. The minimum-distance problem addresses finding the shortest pathway between two points in an interconnected network. Because Equation 17 associates the decay factor with an effective distance, we can restate our search for the maximum pathway couplings as a search for the shortest effective distance between donor and acceptor in the corresponding network. Graph-theory strategies for solving the minimum-distance problem are discussed elsewhere (17).

The first step in using graph theory to find electron-transfer pathways in proteins is to construct a labeled graph (17) corresponding to the superset of all potential pathways. Covalent bonds (established as described below) are first mapped onto vertices. Establishing which vertices are to be joined by edges requires progressively more computation for adjacent covalent bonds, hydrogen bonds, and through-space contacts. The lengths of the edges (i.e. the decays) are determined by the distances between the atoms and the nature of the interaction (Equation 16). The covalent bonds are specified implicitly by the Brookhaven Protein Data Bank files. Covalent interactions, those between bonds anchored at a common atom, are easily identified. Commercial software is used to look up these connections for the known amino acids and other residues, which are then appended to the Protein Data Bank data. These amended Protein Data Bank files are used as input to the PATHWAYS software written by J. N. Betts. Directed by data in the parameter files, the program looks up the model-predicted decays (Equation 16) for the various bond types and stores them. Hydrogen bonds are identified as having acceptable: (*a*) hydrogen-donor and hydrogen-acceptor groups (donors: -NHx; acceptors: carbonyl oxygens; both: -OH), (*b*) donor-hydrogen-acceptor angle, and (*c*) donor-acceptor distance (Å). These values are specified in a parameter file. Edges representing the hydrogen bonds are added to the connection list, and lengths that represent these decays are added to the list of segment lengths. Next, potential through-space connections are sought within a limited radius of each atom, typically 6 Å. No through-space connections longer than 6 Å contribute to significant pathways, so they are ignored

to shorten the data-processing time. The through-space connections are established for each atom, X, as follows. First, the investigator composes a list, L, containing all bonds/vertices within range of X and attempts to eliminate as many of the entries as possible. Through-space connections are eliminated between atoms that have a significantly stronger bond-mediated connection. The first through-space connections the program eliminates from L are those that are redundant with preexisting covalent and hydrogen bonds. The vertices remaining in L are sorted on the basis of their distances from X, shortest first. Next, a depth-first shortest-path search (17) is performed with X as the root, finding the shortest distance to atoms with potential through-space connections. The depth of the search is limited to a length that corresponds to the through-space decay from X to atoms within the through-space cutoff radius. If the search returns without having located the potential atom, the through-space contact is the shortest path to it, and the connection is thus added to the master connection (adjacency) list, and its corresponding length is added to the list of lengths. Otherwise, the through-space connection is discarded and the next vertex in L becomes the new target. In this way, shorter through-space contacts can disqualify longer ones, further decreasing the number of connections added to the graph.

Two standard search strategies are used to arrive at the minimum-distance path between two points in an interconnected network, referred to as depth-first and breadth-first searches. Depth-first searches begin at a specified point and step along allowed connections until no additional forward steps remain (a dead-end is reached) or the target site is found. If a dead-end occurs, the search backtracks by one step and then seeks alternative forward steps from that point, and so on until the target atom is found. Breadth-first searches simultaneously consider all paths radiating from the starting point by keeping track of each vertex and its distance. At each step of the search, a new vertex is added. The vertex chosen to be added is always the one that minimizes the effective distance to the donor at that stage. When the acceptor atom is the one that is added, the minimum-distance pathway has been found. We use a depth-first algorithm. The advantage of the depth-first search for our applications is its pathway orientation, i.e. each excursion represents a potentially acceptable pathway and the paths within a given factor of the best pathway are easily tabulated and accumulated.

RUTHENATED CYTOCHROMES

The molecules we have employed in experiments aimed at extracting T_{DA} values are ones in which ruthenium complexes are attached to surface

histidines of structurally characterized proteins (2, 13, 22, 24, 30, 36, 40, 41, 44, 45, 49, 54, 61, 68, 73, 74, 74a, 75, 77–79). Surface modification of a protein is expected to be nonperturbative, so the structure of the modified protein is presumably the same as that of the native protein. Hence, the distance and the intervening medium involved in electron transfer between the native- and synthetic-protein redox sites are known. Altering the site of attachment allows one to vary both the distance and the intervening medium for electron transfer. Changing the ligands in the ruthenium modification reagent also permits one to study free-energy effects on the rate of the reaction.

Cytochrome c

HIS33 DERIVATIVES The first experimental work on the electron-transfer reactions of Ru-modified proteins involved horse-heart cytochrome *c* modified by coordination of pentaammineruthenium to His33 (Figure 2) (75, 77). The rate of intramolecular electron transfer from $Rua_5(His33)^{2+}$ ($a = NH_3$) to the ferriheme ($T = 298$ K), measured using photochemical techniques, is $30(\pm 5)$ s^{-1} (Table 1). The reaction exhibits a rather small activation enthalpy (2 kcal mol^{-1}) and a large negative activation entropy (−43 eu). Measurements of the temperature dependences of the $Rua_5(His)^{3+/2+}$ and $Fe^{3+/2+}$ potentials in $Rua_5(His33)$-Fe-cyt *c* have provided estimates of $\Delta G°$ [$-4.3(\pm 2)$ kcal mol^{-1}, 298 K), $\Delta H°$ [$-11.5(\pm 10)$ kcal mol^{-1}], and $\Delta S°$ [$-25(\pm 3)$ eu] for the Ru(II)Fe(III) intramolecular electron-transfer reaction. Given these thermodynamic quantities, and the temperature dependence (2–40°C) of the electron-transfer rate in $Rua_5(His33)$-Fe-cyt *c*, one can extract values of λ and T_{DA} from Equation 2 using a classical expression for *FC* (51). Nonlinear least-square fits to the data suggest that $\lambda = 1.2$ eV and $T_{DA} = 0.03$ cm^{-1} (74a). This value of the reorganization energy is quite close to that predicted by the Marcus cross relation (59) [$\lambda_{12} = (\lambda_1 + \lambda_2)/2$] using the reorganization energies for the Fe-cyt *c* ($\lambda_{11} = 1.04$ eV) and $Rua_5(py)^{3+/2+}$ ($\lambda_{22} = 1.20$ eV) self-exchange reactions (15, 51).

A clear understanding of the electronic-coupling strengths in metalloprotein electron-transfer reactions depends upon reliable values of λ and T_{DA}. In addition to studies of temperature dependences, analysis of the driving-force dependence of electron transfer rates can also provide electron-transfer parameters. In the low-driving-force regime ($-\Delta G° \ll \lambda$), the variation of rate with free energy does not strongly depend upon λ, and it is difficult to obtain a good value for this parameter. Much better values of λ and T_{DA} can be obtained from high-driving-force measurements (i.e. $-\Delta G° \approx \lambda$). In this region, the driving-force curve flattens out and electron-transfer rates approach their maximum values.

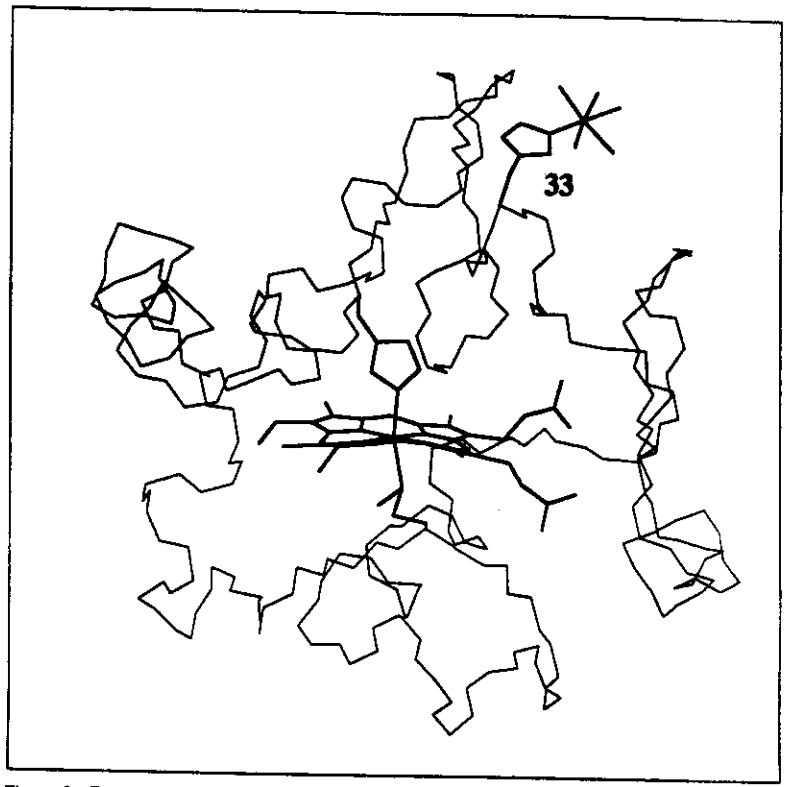


Figure 2 Peptide-backbone structure of $Rua_5(His33)$-Fe-cyt *c*. This complex was prepared by reaction of $Rua_5(OH_2)^{2+}$ with Fe(II)-cyt *c* for 24 h at room temperature. The pure singly modified derivative was isolated using ion-exchange chromatography and was extensively characterized by spectroscopic and chemical methods (77, 78).

It is difficult to prepare a Ru-ammine complex of Fe-cyt *c* in which the driving force for intramolecular electron transfer is much greater than 0.2 eV. Substitution of the native Fe center in cytochrome *c* with Zn, however, has led to high-driving-force intramolecular electron transfer. The lowest triplet-excited state of the Zn-porphyrin in Zn-cyt *c* has a 15-ms lifetime and is a potent reductant [$E° = -0.62$ V vs normal hydrogen electrode (NHE)]. The rates of direct photoinduced electron transfer and thermal recombination have been measured for three $Rua_5L(His33)$-Zn-cyt *c* proteins (L = NH_3, pyridine, isonicotinamide), spanning a 0.39-eV range in $\Delta G°$ (-0.66 to -1.05 eV; Table 1) (30, 54, 74). Fits of these data yield

Table 1 Rate constants and activation parameters for intramolecular electron-transfer reactions of Ru(His)-modified cytochrome *c*

Electron transfer	$-\Delta G^\circ$ (eV)	k_{ET} (s^{-1})	$\Delta H^\ddagger$ (kcal mol^{-1})	$\Delta S^\ddagger$ (eu)
His33 Derivatives (d = 11.1 Å)				
$Ru_5(His)^{2+} \rightarrow Fe(III)^a$	0.18(2)	$3.0(5) \times 10^1$	2.0(5)	−43(5)
$Ru_4(isn)(His)^{2+} \rightarrow ZnP^{+b}$	0.66(5)	$2.0(2) \times 10^5$	<0.5	−35(5)
$ZnP^* \rightarrow Ru_5(His)^{3+c}$	0.70(5)	$7.7(8) \times 10^5$	1.7(4)	−27(5)
$Ru_4(py)(His)^{2+} \rightarrow ZnP^{+b}$	0.74(5)	$3.5(4) \times 10^5$	<0.5	−34(5)
$ZnP^* \rightarrow Ru_4(py)(His)^{3+b}$	0.97(5)	$3.3(3) \times 10^6$	2.2(4)	−22(5)
$Ru_5(His)^{2+} \rightarrow ZnP^{+c}$	1.01(5)	$1.6(4) \times 10^6$	—	—
$ZnP^* \rightarrow Ru_4(isn)(His)^{3+b}$	1.05(5)	$2.9(3) \times 10^6$	<0.5	−30(5)
His39 Derivatives (d = 12.3 Å)^d				
$Ru_4(isn)(His)^{2+} \rightarrow ZnP^+$	0.66(5)	$6.5(7) \times 10^5$	−1.7(4)	−39(5)
$ZnP^* \rightarrow Ru_5(His)^{3+}$	0.70(5)	$1.5(2) \times 10^6$	1.3(3)	−27(5)
$Ru_4(py)(His)^{2+} \rightarrow ZnP^+$	0.74(5)	$1.5(2) \times 10^6$	−1.8(4)	−37(5)
$ZnP^* \rightarrow RuA_4(py)(His)^{3+}$	0.97(5)	$8.9(9) \times 10^6$	0.2(2)	−27(5)
$Ru_5(His)^{2+} \rightarrow ZnP^+$	1.01(5)	$5.7(6) \times 10^6$	−0.2(2)	−29(5)
$ZnP^* \rightarrow Ru_4(isn)(His)^{3+}$	1.05(5)	$1.0(1) \times 10^7$	0.2(2)	−27(5)
His62 Derivatives (d = 14.8 Å)^e				
$ZnP^* \rightarrow Ru_5(His)^{3+}$	0.70(5)	$6.5(7) \times 10^3$	1.4(3)	−37(5)
$Ru_4(py)(His)^{2+} \rightarrow ZnP^+$	0.74(5)	$8.1(8) \times 10^3$	—	—
$ZnP^* \rightarrow Ru_4(py)(His)^{3+}$	0.97(5)	$3.6(4) \times 10^4$	—	—
$Ru_5(His)^{2+} \rightarrow ZnP^+$	1.01(5)	$2.0(2) \times 10^4$	0.7(7)	−37(5)

[a] References 61, 75.
[b] Reference 54.
[c] Reference 30.
[d] Reference 74.
[e] Reference 73.

$\lambda = 1.10$ eV and $T_{DA} = 0.12$ cm^{-1} for the photoinduced reactions, and $\lambda = 1.19$ eV and $T_{DA} = 0.09$ cm^{-1} for the recombinations. The electron-transfer parameters are not extremely sensitive to the nature of the reaction (photoinduced or recombination), and these reactions can be adequately described by a single pair of parameters: $\lambda = 1.15(10)$ eV and $T_{DA} = 0.1(2)$ cm^{-1} (74a).

The similarity in reorganization energies for the Ru-Fe-cyt *c* and Ru-Zn-cyt *c* intramolecular electron-transfer reactions is to be expected. The total reorganization energy is a sum of inner-sphere (λ_i) and outer-sphere (λ_o) elements. Inner-sphere contributions arise from nuclear rearrangements in the Ru-ammine and metalloporphyrin complexes accompanying electron transfer. These rearrangements are rather small and have been estimated to contribute no more than 0.2 eV to λ for both Ru-Fe-cyt *c* and Ru-Zn-cyt *c* (54). The two sources of outer-sphere rearrangements are the solvent and the peptide matrix. Calculations based on a single-sphere dielectric continuum model (16) indicate a 0.6-eV contribution to λ_o from the solvent (54). From the structures of ferri- and ferrocytochrome *c*s, the peptide contribution to λ_o has been calculated to be about 0.2 eV (20). The sum of these individual components (1.0 eV) is in good agreement with the experimentally derived reorganization energy for the Ru-M-cyt *c* (M = Fe, Zn) systems.

HIS39 DERIVATIVES Ru-ammine complexes have been bound to His39 of Zn-substituted cytochrome *c* from *Candida krusei* (68, 74). Intramolecular electron-transfer rates (Table 1) are approximately three times faster than those of corresponding reactions in His33 derivatives of horse-heart cytochrome *c*. The variation of rates with driving force in these derivatives suggests a 1.2(1) eV reorganization energy, indistinguishable from that found in the His33 complexes. The faster electron-transfer rates have been attributed to stronger donor-acceptor electronic coupling in the His39-modified protein (74).

The direct D-A distances in Ru(His33)-Zn-cyt *c* and Ru(His39)-Zn-cyt *c* are 11.1 and 12.3 Å, respectively; however, the T_{DA} is twofold larger for the His39 system. The pathway model is somewhat more consistent with the data: both the His33 and His39 pathways consist of 11 covalent bonds and 1 hydrogen bond (Figure 3). The n_{eff} values for His33 and His39 are 13.9 and 14.0 bonds, respectively (74a).

HIS62 DERIVATIVES Site-directed mutagenesis creates many new opportunities for studying electron transfer in Ru-modified proteins. A yeast (*Saccharomyces cerevisiae*) cytochrome *c* variant has been characterized as having a surface histidine at position 62 (13). The Ru_5(His62) derivative of this mutant protein was prepared, and the rate of electron transfer from Ru(II) to Fe(III) was found to be 1.7 s^{-1} (Table 1) (13). Ru_5(His62) and Ru_4(His62) derivatives of Zn-substituted *S. cerevisiae* cytochrome *c* have also been examined. The rates of the photoinduced and thermal recombination reactions are more than two orders of magnitude slower than the rates of analogous reactions in His33 derivatives of horse-heart cytochrome *c* (73). The driving-force data are more limited than for the other His derivatives of cytochrome *c*, but again suggest that $\lambda \approx 1.2$ eV. The slower rates for the His62 derivatives are attributed to weaker electronic coupling. The direct D-A separation is 14.8 Å, while the effective number of bonds in the pathway is 20.6 (Figure 3) (74a). Both measures suggest that the His62 electron transfer reactions should be substantially slower than those found in His33 or His39 derivatives.

NATURE OF THE PATHWAYS Qualitative differences can arise in the collection of best pathways found, depending on the protein structure. We have

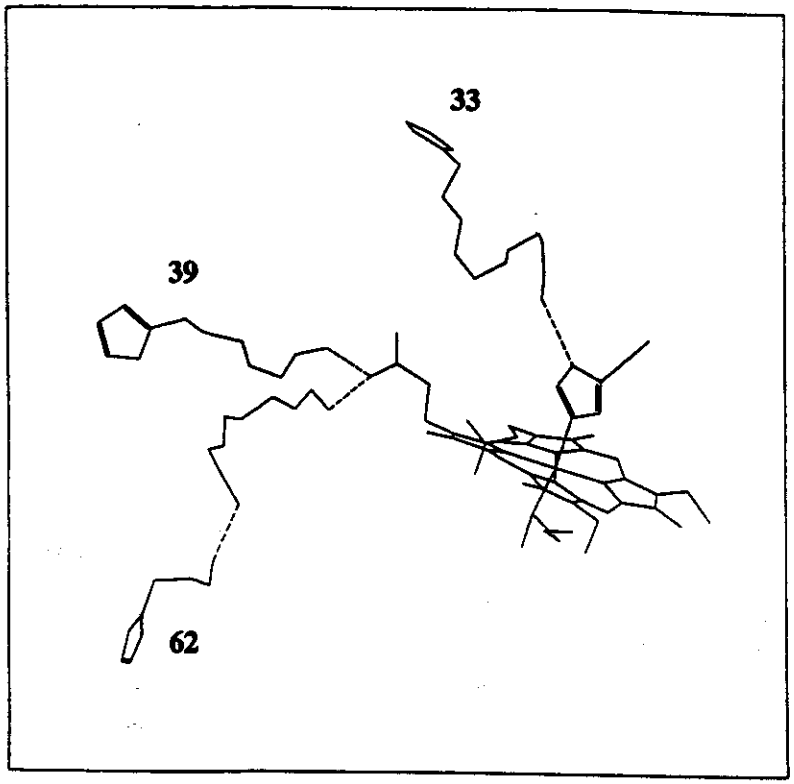


Figure 3 Predicted electronic-coupling pathways in Ru(His33)-, Ru(His39)-, and Ru(His62)-modified cytochrome *c*. Covalent bonds are depicted as solid lines and hydrogen bonds as dashed lines.

examined the paths within a factor of 10 of the best one in ruthenated His39 and His62 cytochrome *c*. In the His39 derivative, three routes feed into a single propionic acid side chain of the heme (Figure 4). The three pathways are more or less parallel and not highly interconnected. His62 has only two classes of pathways, but paths between and within each class have intertwined pathways near the His62 group, which are independent at intermediate distance and connect to independent parts of the heme. Pathway coupling calculations can be displayed in map form: Figure 5 is a coupling map showing the maximum pathway coupling to each α-carbon in cytochrome *c*.

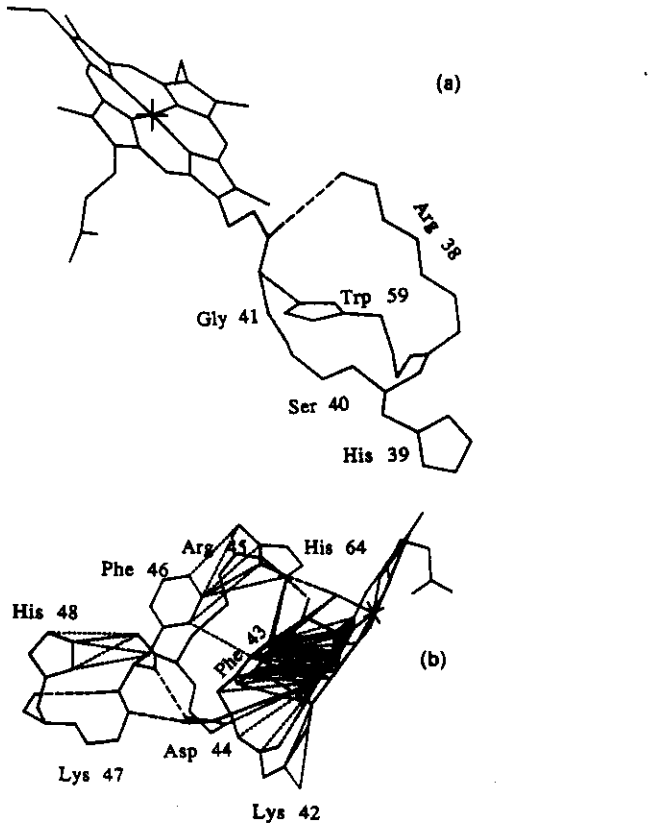


Figure 4 The best paths for (*a*) His39 cyt *c* and (*b*) His48 Mb are shown. Dotted lines are through-space contacts. Note that the best paths in cyt *c* are structurally related to one another, while several classes of pathways exist in Mb.

Cytochrome b_5

HIS26 DERIVATIVES Three surface His residues of trypsin-solubilized bovine cytochrome b_5 (Tb_5) have been modified by coordination to Ru-penta-ammine complexes (His15, His80, His26) (40). Rates of intramolecular electron transfer from Fe(II) to Ru(III) have been measured in three His26 derivatives: $Ru a_5$(His26)-Tb_5; mutant (Asn57 to Asp, Gln13 to Glu, Glu11 to Gln, His15 to Asn, His80 to Asn) lipase-solubilized cytochrome b_5 [$Ru a_5$(His26)LMb_5]; and deuteroporphyrin-substituted (DP) Tb_5 [$Ru a_5$(His26)DPb_5] (40, 41). Electron-transfer rates vary by more than an

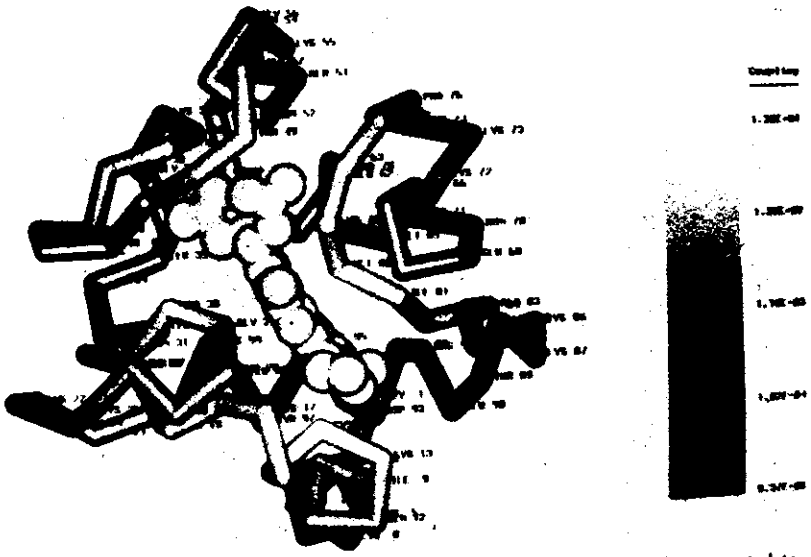


Figure 5 Electronic coupling map for cytochrome *c*. Amino acids directly connected to those coordinating the Fe or hydrogen-bonding to the heme are anomalously strongly coupled in reference to their through-space distance from the heme.

order of magnitude for the three proteins (Table 2). The small differences in driving force or estimated D-A separation cannot readily account for the variations in rate. Driving-force data are not available for this system, but changes in λ probably could not be responsible for the differences in electron-transfer rates. The pathway model has been invoked to account for the differences in rates. A critical through-space jump (from Leu25 to the heme) in the pathway from His26 to the heme is not constant in the three different proteins (Figure 6). The dramatic reduction in rate in Rua_5(His26)DPb_5 has been attributed to the absence of the heme 2-vinyl

Table 2 Electron-transfer rates in Ru(His26)-modified cytochrome b_5[a]

Electron transfer	$-\Delta G^\circ$ (eV)	k_{ET} (s^{-1})	d (Å)
Fe(II)-Tb_5 → Rua_5(His26)$^{3+}$	0.08(2)	1.4(1)	12.1
Fe(II)-LMb_5 → Rua_5(His26)$^{3+}$	0.10(2)	5.9(5)	12.0
Fe(II)-DPb_5 → Rua_5(His26)$^{3+}$	0.13(2)	0.2(1)	12.9

[a] Reference 41.

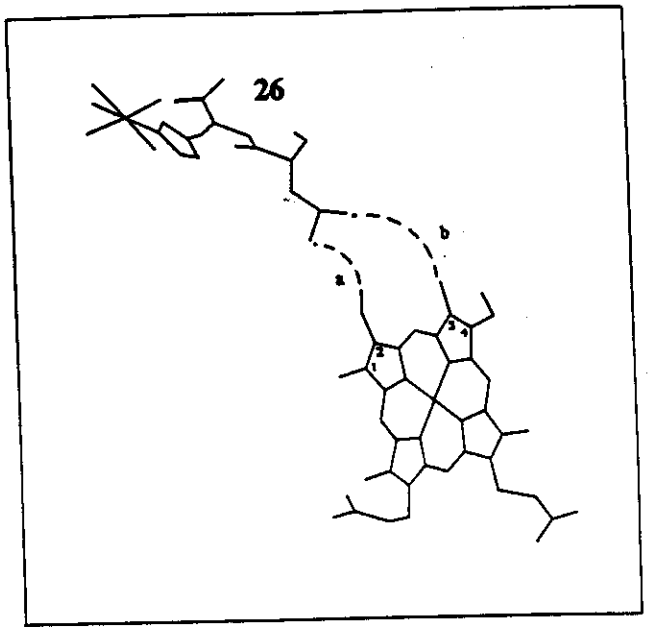


Figure 6 Best-pathway through-space jump from Leu25 to the heme in Ru(His26)-modified cytochrome Tb_5(*a*) and DPb_5 (*b*).

group, which is the terminus of the Leu25-to-heme through-space jump in the other two proteins (40, 41). A longer jump to the heme 3-methyl is predicted for Rua_5(His26)DPb_5, leading to a slower electron-transfer rate.

NATURE OF THE PATHWAYS Pathways from His26 to the heme in cytochrome b_5 are somewhat less sparse than in cytochrome *c*. When the His63-Fe coupling (2.04-Å bond length) is treated as a through-space interaction, pathways through the vinyl group dominate as described above. Two other classes of pathways can be identified. Pathways arising from through-space interactions between the heme and residues His63 and Phe58 form a second tier of paths with weaker coupling than those described above.

Reorganization Energies and Electronic Couplings

Based on the few systems in which a reliable number has been extracted, $\lambda = 1.2$ eV appears to be a reasonable value for Ru-ammine-modified cytochromes (74a). Perhaps because of a lack of data and limited precision in the derived parameters, λ has not been found to be particularly sensitive to D-A separation or to the site of modification. In fact, the simple Marcus

cross relation provides a reasonably good estimate of the reorganization energies in these reactions.

Unlike the reorganization energy, the electronic-coupling strengths in the Ru-modified cytochromes show a great deal of variability. Equation 15 expresses a simple distance dependence for T_{DA} that adequately describes electron transfer in model complexes with values of β between 0.8 and 1.2 $Å^{-1}$. This distance dependence, assuming a maximum electron-transfer rate of 10^{13} s^{-1} at close contact ($d = 3$ Å), is represented by the solid ($\beta = 1.0$ $Å^{-1}$) and dashed ($\beta = 0.8$, 1.2 $Å^{-1}$) lines in Figure 7. Estimates of maximum electron-transfer rates (i.e. the rate at $-\Delta G^\circ = \lambda$) for Ru-modified cytochromes (Table 3) are plotted as a function of D-A separation (λ was assumed to be 1.2 eV for the cytochrome b_5 derivatives). Clearly all of the maximum rates lie below the values predicted by Equation 15; there is no simple correlation. The obvious conclusion is that, for a given D-A separation, the electronic coupling in the Ru-modified proteins is

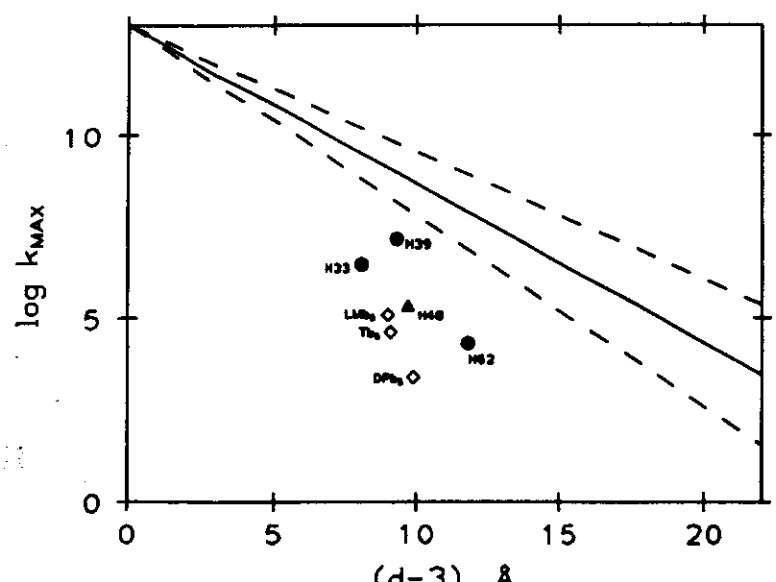


Figure 7 Plot of $\log k_{MAX}$ versus D-A distance (d) minus 3 Å (van der Waals contact) for Ru-modified proteins. Solid and dashed lines represent Equation 15 with $\beta = 0.8$, 1.0, and 1.2 $Å^{-1}$. Filled symbols indicate systems in which λ was estimated from a driving-force study. Open symbols indicate that an assumed value for λ (1.2 eV) was used to estimate k_{MAX}. ●, Cytochrome c; ▲, His48 derivative of Mb; ◇, His26 derivatives of cytochrome b_5.

Table 3 Maximum rates, D-A distances, coupling strengths, and effective bonds in pathways for Ru-modified proteins

	k_{max} (s^{-1})	d(Å)	T_{DA} (cm^{-1})	n_{eff} (bonds)
Ru(His39)cyt c[a]	1.4×10^7	12.3	0.24	14.0
Ru(His33)cyt c[b]	2.9×10^6	11.1	0.11	13.9
Ru(His62)cyt c[c]	2.0×10^4	14.8	0.01	20.6
Ru(His26)Tb_5[d]	4.1×10^4	12.1	0.01	19.0
Ru(His26)LMb_5[d]	1.2×10^3	12.0	0.02	18.7
Ru(His26)DPb_5[d]	2.4×10^3	12.9	0.003	20.3
Ru(His48)Mb[e]	2.1×10^5	12.7	0.03	22.6

[a] Reference 74.
[b] Reference 54.
[c] Reference 73.
[d] Reference 41.
[e] Reference 74a.

substantially weaker than that predicted by a simple exponential decay with distance.

Our pathway model predicts the failure of exponential-decay correlations based on edge-edge distances. It also predicts that maximum electron-transfer rates correlate with the effective number of bonds in the pathway. [Multiplying n_{eff} by a canonical value of 1.4 Å/bond gives a tunneling length (σl) that replaces d in rate-distance correlations.] Maximum electron-transfer rates in the Ru-modified cytochromes are plotted against σl in Figure 8. A linear least-square fit yields the solid line with a slope of 0.7 $Å^{-1}$. Though the data are limited, the intercept at one bond (i.e. 1.4 Å) corresponds to a maximum electron-transfer rate of 3.4×10^{12} s^{-1}, which is in reasonable agreement with data from covalently coupled D-A complexes (33, 37, 76).

RUTHENATED MYOGLOBIN

His48 Derivatives

Myoglobin (Mb) is an oxygen-storage protein with 153 amino acids and a heme prosthetic group (1). Unlike cytochrome c, the heme is not covalently bound to the protein in Mb. This feature greatly facilitates metal substitution and has enabled the preparation of Ru(His48) proteins with six different metalloporphyrin active sites.

For cytochrome c, the evidence indicated that the reorganization energy for the electron-transfer reactions of the Zn-substituted protein would be nearly the same as that of the native-Fe protein. This, however, is not

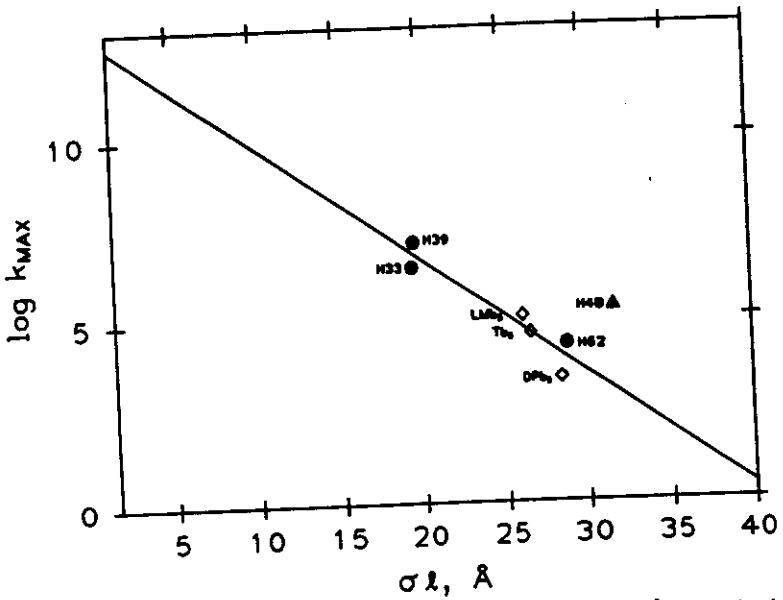


Figure 8 Plot of $\log k_{MAX}$ versus the tunneling length, σl ($=n_{eff} \times 1.4$ Å/bond), in the physical pathway between donor and acceptor for three Ru-modified derivatives of cytochrome c (●), three His26 derivatives of cytochrome b_5 (◇), and one Mb derivative (▲). The solid line is a linear least-square fit to the three cytochrome c points.

likely to be true in myoglobin: Fe(II)-Mb is a five-coordinate low-spin complex that, upon one-electron oxidation, binds a water molecule to form a six-coordinate species (1). This change in coordination number should be reflected in a greater reorganization energy in native Fe Mb, or even by electron-transfer rates limited by ligand binding or dissociation. The electron-transfer reactions of Ru-modified, metal-substituted Mb, however, are not accompanied by changes in metal coordination, and a single set of electron-transfer parameters should adequately describe these reactions. The rates of 18 different electron-transfer reactions in Ru(His48) Mb have been reported (Table 4) (2, 22, 24, 44, 45, 74a), spanning nearly 0.8 eV in driving force. Fitting the photoinduced and thermal recombination rates to a classical expression for *FC* yields $\lambda = 1.26$ eV and $T_{DA} = 0.03$ cm^{-1} (74a). The D-A separation in Ru(His48)Mb (12.7 Å) is quite similar to the distances in His33 and His39 derivatives of cytochrome c and does not readily explain the smaller value of T_{DA}. The best pathway in Ru(His48)Mb, however, is comprised of 22.6 effective bonds (eight more than found in the His33 and His39 pathways).

Table 4 Electron-transfer rates in Ru(His48)-modified myoglobin

Electron transfer	$-\Delta G^\circ$ (eV)	k_{ET} (s^{-1})
$H_2P^* \rightarrow Ru^{II}_5(His)^{3+}$ [a]	0.39(±2)	7.6(8) × 10^2
$Ru^{II}_4(isn)(His)^{2+} \rightarrow ZnP^{+}$ [b]	0.61(±2)	1.4(5) × 10^4
$PdP^* \rightarrow Ru^{II}_5(His)^{3+}$ [c]	0.64(±2)	9.1(9) × 10^3
$Ru^{II}_4(py)(His)^{2+} \rightarrow ZnP^{+}$ [b]	0.69(±2)	2.0(5) × 10^4
$CdP^* \rightarrow Ru^{II}_5(His)^{3+}$ [a,d]	0.79(±2)	4.5(5) × 10^4
$MgP^* \rightarrow Ru^{II}_5(His)^{3+}$ [a,d]	0.81(±2)	9.5(9) × 10^4
$ZnP^* \rightarrow Ru^{II}_5(His)^{3+}$ [b]	0.82(±2)	7.0(7) × 10^4
$PdP^* \rightarrow Ru^{II}_4(py)(His)^{3+}$ [c]	0.91(±2)	9.0(9) × 10^4
$Ru^{II}_5(His)^{2+} \rightarrow MgP^{+}$ [d]	0.91(±2)	4.9(5) × 10^4
$Ru^{II}_5(His)^{2+} \rightarrow CdP^{+}$ [d]	0.92(±2)	2.1(2) × 10^5
$Ru^{II}_5(His)^{2+} \rightarrow ZnP^{+}$ [b]	0.96(±2)	1.5(5) × 10^5
$ZnP^* \rightarrow Ru^{II}_4(py)(His)^{3+}$ [b]	1.09(±2)	2.0(2) × 10^5
$ZnP^* \rightarrow Ru^{II}_4(isn)(His)^{3+}$ [b]	1.17(±2)	2.9(3) × 10^5
$Ru^{II}_5(His)^{2+} \rightarrow Fe(II)(CN\text{-}Br)$ [e]	0.09(±2)	2.0(5)
$Fe(II)(CN\text{-}Br) \rightarrow Ru^{II}_4(isn)^{3+}(His)$ [e]	0.26(±2)	5.5(5)
$Fe(II) \rightarrow Ru^{II}_5(His)^{3+}$ [f]	0.02(±2)	4.0(5) × 10^{-2}
$Fe(II) \rightarrow Ru^{II}_4(py)(His)^{3+}$ [g]	0.28(±2)	2.5(5)
$Fe(II) \rightarrow Ru^{II}_4(isn)(His)^{3+}$ [e]	0.35(±2)	3.0(4)

[a] Reference 22.
[b] References 44, 74a.
[c] Reference 45.
[d] Reference 74a.
[e] Reference 79.
[f] Reference 24.
[g] Reference 49.

Cyanogen-bromide modification of His64 in the distal heme pocket of myoglobin inhibits coordination of a water ligand to the ferric heme (43, 56, 69). Therefore the reorganization energy for electron transfer in cyanogen-bromide-treated Ru(His48)-Fe-Mb will probably be nearly the same as that of the metal-substituted myoglobins. Fitting the two electron-transfer rates measured for this system (Table 4) to Equation 2, with λ equal to 1.26 eV, yields an electronic-coupling matrix element of 0.01 cm^{-1} (79). As in the case of Ru-modified cytochrome c, the apparent coupling strength in Ru-ammine–iron-heme reactions is somewhat smaller than that found for reactions involving Ru-ammines and metal-substituted porphyrins.

Three electron-transfer rates have been measured with Ru-ammines bound to His48 of native Fe myoglobin (Table 4) (24, 49). The reorganization energy for these reactions can be estimated by assuming that the coupling strength is the same as that found in the cyanogen bromide–treated systems (0.01 cm^{-1}) and by optimizing λ. The data suggest $\lambda = 1.48$

eV (74a), a 0.2-eV increase over the value found in systems that have no change in coordination number.

Nature of the Pathways

The pathways in cytochrome *c* are rather sparse and fairly independent compared to pathways in His48 Mb (Figure 4). This Mb derivative has two or three families of pathways, one connected to the heme by a hydrogen bond to Arg45, one connected through space to Phe43, and one connected through space to Lys42. In contrast to cytochrome *c*, loops interconnect these paths (by hydrogen-bond and through-space interactions) throughout the protein. The limitations of the single pathway approximation depend on the number of loops in the intervening medium, and it will be interesting to see (theoretically as well as experimentally) how well single-pathway models work for proteins of varied structural motifs.

Figure 8 also plots the maximum electron transfer rate for His48-modified Mb (n_{eff} = 22.6 bonds) (74a). This Mb point lies substantially above the line based on the cytochrome *c* and b_5 data, and clearly indicates a problem with the pathway model. In the simple form of this model, a single route is assumed to dominate the D-A coupling. The pathway-searching algorithm tends to support this assumption in the cytochromes, where single coupling paths stand out. In Mb, however, the pathway-searching algorithm identifies many nearly equivalent pathways: the one used for the point in Figure 8 represents the best route, but there are several close competitors. The problem is again the tunneling distance: with many nearly equivalent paths contributing to D-A coupling, n_{eff} will be substantially below 22.6 bonds for His48 Mb. Efforts are being made to refine the pathway model to accommodate multiple paths. If enough paths contribute to the overall electronic coupling in a given protein, the composition of any one path becomes relatively unimportant and tunneling lengths should closely parallel edge-edge distances.

CONCLUDING REMARKS

In addition to the utility of the pathway-mapping method in examining couplings between pairs of points in proteins, its simplicity allows the generation of global coupling maps from all atoms in a protein to the redox center. Such maps reveal important characteristic effects of primary, secondary, tertiary, and quaternary folded structure on the nature of the coupling to a specific patch of the protein. Regions anomalously strongly or weakly coupled to the redox center, given their distances, are easily identified (3), and evidence of these anomalous regions should appear in the experimental data. The simple pathway model appears to work relatively well for the cytochromes, which have few pathways. The His48 Mb data are not entirely consistent with the single-pathway analysis discussed here. Differences can arise from inadequacy of the single-path approximation, uncertainties in fitting of the experimental data, and inadequacies of using a simple classical expression for *FC*.

Our pathway technique is now being used to study other biological systems (19, 21, 29, 31, 32, 39, 57). New theoretical strategies are being implemented to further our understanding of tunneling pathways so that greater molecular detail can be included in the treatment of very large systems (see 66 for a description of the methods used). The long-term goal of this work is to obtain compact symbolic representations for proteins that include all the relevant pathways in the protein.

ACKNOWLEDGMENTS

Work in San Diego was funded by the National Science Foundation (Grant No. DMB-9018768) and a research contract from the Jet Propulsion Laboratory, supported by the Department of Energy's Catalysis/Biocatalysis Program. This work was performed in part at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Department of Energy's Catalysis/Biocatalysis Program (Advanced Industrial Concepts Division), through an agreement with the National Aeronautics and Space Administration. Experimental work at the Beckman Institute was supported by the National Science Foundation and the National Institutes of Health.

Literature Cited

1. Antonini, E., Brunori, M. 1971. *Hemoglobin and Myoglobin in Their Reactions with Ligands*. Amsterdam: North Holland
2. Axup, A. W., Albin, M., Mayo, S. L., Crutchley, R. J., Gray, H. B. 1988. *J. Am. Chem. Soc.* 110: 435
3. Beratan, D. N., Betts, J. N., Onuchic, J. N. 1991. *Science* 252: 1285
4. Beratan, D. N., Hopfield, J. J. 1984. *J. Am. Chem. Soc.* 106: 1584
5. Beratan, D. N., Hopfield, J. J. 1984. *J. Chem. Phys.* 81: 5753
6. Beratan, D. N., Onuchic, J. N. 1989. *Photosynth. Res.* 22: 173
7. Beratan, D. N., Onuchic, J. N. 1991. *Adv. Chem. Ser.* 228: 71
8. Beratan, D. N., Onuchic, J. N., Betts, J. N., Bowler, B. E., Gray, H. B. 1990. *J. Am. Chem. Soc.* 112: 7915
9. Beratan, D. N., Onuchic, J. N., Gray, H. B. 1991. See Ref. 72, p. 97
10. Beratan, D. N., Onuchic, J. N., Hopfield, J. J. 1985. *J. Chem. Phys.* 83: 5325
11. Beratan, D. N., Onuchic, J. N., Hopfield, J. J. 1987. *J. Chem. Phys.* 86: 4488
12. Bialek, W., Bruno, W. J., Joseph, J., Onuchic, J. N. 1989. *Photosynth. Res.* 22: 15
13. Bowler, B. E., Meade, T. J., Mayo, S. L., Richards, J. H., Gray, H. B. 1989. *J. Am. Chem. Soc.* 111: 8757
14. Broo, S., Larsson, S. 1989. *J. Quant. Chem. Quant. Biol. Symp.* 16: 185
15. Brown, G. M., Sutin, N. 1979. *J. Am. Chem. Soc.* 101: 883
16. Brunschwig, B. S., Ehrenson, S., Sutin, N. 1986. *J. Phys. Chem.* 90: 3657
17. Buckley, F., Harary, F. 1990. *Distance in Graphs*. New York: Addison-Wesley

18. Caldeira, A. O., Leggett, A. J. 1983. *Ann. Phys.* 149: 374
19. Christensen, H. E. M., Conrad, L. S., Ulstrup, J., Mikkelsen, K. V. 1991. See Ref. 72, p. 57
20. Churg, A. K., Weiss, R. M., Warshel, A., Takano, T. 1983. *J. Phys. Chem.* 87: 1683
21. Conrad, D. W., Scott, R. A. 1989. *J. Am. Chem. Soc.* 111: 3461
22. Cowan, J. A., Upmacis, R. K., Beratan, D. N., Onuchic, J. N., Gray, H. B. 1988. *Ann. N.Y. Acad. Sci.* 550: 68
23. Cramer, W. A., Knaff, D. B. 1990. *Energy Transduction in Biological Membranes*. New York: Springer-Verlag
24. Crutchley, R. J., Ellis, W. R., Gray, H. B. 1985. *J. Am. Chem. Soc.* 107: 5002
25. da Gama, A. A. S. 1985. *Theor. Chim. Acta* 68: 159
26. da Gama, A. A. S. 1990. *J. Theor. Biol.* 142: 251
27. Davydov, A. S. 1987. *Phys. Status Solidi B* 90: 457
28. DeVault, D. 1984. *Quantum Mechanical Tunneling in Biological Systems*. New York: Cambridge Univ. Press. 2nd ed.
29. Durham, B., Pan, L. P., Long, J. E., Millett, F. 1989. *Biochemistry* 28: 8659
30. Elias, H., Chou, M. H., Winkler, J. R. 1988. *J. Am. Chem. Soc.* 110: 429
31. Farver, O., Pecht, I. 1989. *FEBS Lett.* 244: 379
32. Feher, G., Allen, J. P., Okamura, M. Y., Rees, D. C. 1989. *Nature* 339: 111
33. Fox, L. S., Kozik, M., Winkler, J. R., Gray, H. B. 1990. *Science* 247: 1069
34. Garg, A., Onuchic, J. N., Ambegaokar, V. 1985. *J. Chem. Phys.* 83: 4491
35. Goldman, C. 1991. *Phys. Rev. A* 43: 4500
36. Gray, H. B., Malmström, B. G. 1989. *Biochemistry* 28: 7499
37. Holten, D., Hoganson, C., Windsor, M. W., Schenck, C. C., Parson, W. W., et al. 1980. *Biochim. Biophys. Acta* 592: 461
38. Hopfield, J. J. 1974. *Proc. Natl. Acad. Sci. USA* 71: 3640
39. Jackman, M. P., McGinnis, J., Powls, R., Salmon, G. A., Sykes, A. G. 1988. *J. Am. Chem. Soc.* 110: 5880
40. Jacobs, B. A. 1991. *Preparation, characterization, and intramolecular electron transfer in pentaammineruthenium-modified derivatives of cytochrome* b_5 *and azurin*. PhD Thesis. Calif. Inst. Technol., Pasadena, Calif.
41. Jacobs, B. A., Mauk, M. R., Funk, W. D., MacGillivray, R. T. A., Mauk, A. G., Gray, H. B. 1991. *J. Am. Chem. Soc.* 113: 4390
42. Jortner, J. 1980. *Biochim. Biophys. Acta* 594: 139
43. Kamiya, N., Shiro, Y., Iwata, T., Iizuka, T., Iwasaki, H. 1991. *J. Am. Chem. Soc.* 113: 1826
44. Karas, J. L. 1989. *Long-range electron transfer in ruthenium-labelled myoglobin*. PhD Thesis. Calif. Inst. Technol., Pasadena, Calif.
45. Karas, J. L., Lieber, C. M., Gray, H. B. 1988. *J. Am. Chem. Soc.* 110: 599
46. Kuki, A., Wolynes, P. G. 1987. *Science* 236: 1647
47. Larsson, S. 1981. *J. Am. Chem. Soc.* 103: 4034
48. Larsson, S. 1983. *J. Chem. Soc. Faraday Trans.* 279: 1375
49. Lieber, C. M., Karas, J. L., Gray, H. B. 1987. *J. Am. Chem. Soc.* 109: 3779
50. Lin, S. H. 1989. *J. Chem. Phys.* 90: 7103
51. Marcus, R. A., Sutin, N. 1985. *Biochim. Biophys. Acta* 811: 265
52. Magarshak, Y., Malinsky, J., Joran, A. D. 1991. *J. Chem. Phys.* 95: 418
53. McConnell, H. M. 1961. *J. Chem. Phys.* 35: 508
54. Meade, T. J., Gray, H. B., Winkler, J. R. 1989. *J. Am. Chem. Soc.* 111: 4353
55. Mikkelsen, K. V., Ratner, M. A. 1988. *Chem. Rev.* 87: 113
56. Morishima, I., Shiro, J., Wakino, T. 1985. *J. Am. Chem. Soc.* 107: 1063
57. Natan, M. J., Baxter, W. W., Kuila, D., Gingrich, D. J., Martin, G. S., Hoffman, B. M. 1991. *Adv. Chem. Ser.* 228: 201
58. Newton, M. D. 1988. *J. Phys. Chem.* 92: 3049
59. Newton, M. D. 1991. *Chem. Rev.* 91: 767
60. Newton, M. D., Sutin, N. 1984. *Annu. Rev. Phys. Chem.* 35: 437
61. Nocera, D. G., Winkler, J. R., Yocom, K. M., Bordignon, E., Gray, H. B. 1984. *J. Am. Chem. Soc.* 106: 5145
62. Onuchic, J. N. 1987. *J. Chem. Phys.* 86: 3925
63. Onuchic, J. N., Beratan, D. N. 1987. *J. Am. Chem. Soc.* 109: 6771
64. Onuchic, J. N., Beratan, D. N. 1990. *J. Chem. Phys.* 92: 722
65. Onuchic, J. N., Beratan, D. N., Hopfield, J. J. 1986. *J. Phys. Chem.* 90: 3707
66. Onuchic, J. N., de Andrade, P. C. P., Beratan, D. N. 1991. *J. Chem. Phys.* 92: 1131
67. Onuchic, J. N., Wolynes, P. G. 1988. *J. Phys. Chem.* 92: 6495
67a. Ratner, M. A. 1990. *J. Phys. Chem.* 94: 4877
68. Selman, M. A. 1989. *Preparation and characterization and intramolecular electron transfer in a pentaammineruthenium derivative of* Candida krusei *cytochrome c*. PhD Thesis. Calif. Inst. Technol., Pasadena, Calif.
69. Shiro, Y., Morishima, I. 1984. *Biochemistry* 23: 4879
70. Siddarth, P., Marcus, R. A. 1990. *J. Phys. Chem.* 94: 2985
71. Siddarth, P., Marcus, R. A. 1990. *J. Phys. Chem.* 94: 8430
72. Sigel, H., Sigel, A., eds. 1991. *Metal Ions in Biological Systems*, Vol. 27. New York: Marcel Dekker
73. Therien, M. J., Bowler, B. E., Selman, M. A., Gray, H. B., Chang, I.-J., Winkler, J. R. 1991. *Adv. Chem. Ser.* 228: 191
74. Therien, M. J., Selman, M. A., Gray, H. B., Chang, I.-J., Winkler, J. R. 1990. *J. Am. Chem. Soc.* 112: 2420
74a. Winkler, J. R., Gray, H. B. 1992. *Chem. Rev.* In press
75. Winkler, J. R., Nocera, D. G., Yocom, K. M., Bordignon, E., Gray, H. B. 1982. *J. Am. Chem. Soc.* 104: 5798
76. Wasielewski, M. R., Niemczyk, M. P., Svec, W. A., Pewitt, E. B. 1985. *J. Am. Chem. Soc.* 107: 5562
77. Yocom, K. M. 1981. *The synthesis and characterization of inorganic redox reagent-modified cytochrome c*. PhD Thesis. Calif. Inst. Technol., Pasadena, Calif.
78. Yocom, K. M., Shelton, J. B., Shelton, J. R., Schroeder, W. E., Worosila, G., et al. 1982. *Proc. Natl. Acad. Sci. USA* 79: 7052
79. Zewert, T. E. 1990. *Electron transfer in chemically and genetically modified myoglobins*. PhD Thesis. Calif. Inst. Technol., Pasadena, Calif.

Mapping Electron Tunneling Pathways: An Algorithm that Finds the "Minimum Length"/Maximum Coupling Pathway between Electron Donors and Acceptors in Proteins

Jonathan N. Betts,[†] David N. Beratan,*[,‡] and José Nelson Onuchic[§]

Contribution from the Beckman Institute, California Institute of Technology, Pasadena, California 91125, Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, and Department of Physics, University of California, San Diego, La Jolla, California 92093. *Received May 13, 1991*

Abstract: The covalent, hydrogen bonded, and van der Waals connectivity of proteins can be represented with geometrical objects called graphs. In these graphs, vertices represent bonds and the connections between them, edges, represent bond–bond interactions. We describe a model in which edge lengths are associated with the wave function decay between interacting pairs of bonds, and a minimum distance graph-search algorithm is used to find the pathways that dominate electron donor–acceptor interactions in these molecules. Predictions of relative electron transfer rates can be made from these pathway lengths. The results are consistent with many experimentally measured electron-transfer rates, although some anomalies exist. Presentation of the pathway coupling between the donor (or acceptor) and every other atom in a given protein as a color-coded map provides a design tool for tailored electron-transfer proteins.

Introduction

Graph theory is often used in chemistry to describe the relationship between molecular structure and chemical properties.[1] Although alternative approaches have been used,[1a,b] traditional chemical graphs consist of a direct mapping of atoms to vertices as well as bonds to edges linking vertices. The graph representations of proteins that are described here employ a slightly different mapping where vertices correspond to bonds and edges to covalent, hydrogen bond, and van der Waals interaction between bonds. The edge lengths represent the wave function decay through these bonded or nonbonded contacts rather than physical lengths. Longer effective lengths represent larger decays. Minimum length pathways often make the dominant contribution to the protein-mediated coupling between electron donor (D) and acceptor (A).

Tunneling Pathways. Many biological reactions shift an electron a considerable distance (>5 Å) via electron tunneling. Such long distance transfers are in the nonadiabatic limit, so the rate is proportional to the square of the protein mediated donor–acceptor coupling, T_{DA}.[2] We recently developed a tunneling pathway model for electron transfer in proteins that identifies the bonded and nonbonded interactions that give rise to the coupling.[3,4] This model is based on an effective one-electron tight-binding hamiltonian. These one-electron interaction parameters are renormalized couplings arising from the more complete hamiltonian. The validity of this reduction and the simple parameter set (discussed below) have been discussed in prior papers.[3d,4d] Using this simple hamiltonian, relative values of T_{DA} have been estimated for a variety of proteins using a single pathway approximation. The single pathway parameters include corrections (in an average sense) due to scattering of electron amplitude in side chains connected to the main pathway.[5a] Alternative electronic structure methods for computing T_{DA} in large systems are being actively pursued.[5] The goal of this method, described in detail here, is to apply the *minimal description* required to incorporate the basic features of the mechanism for electron tunneling in proteins. In spite of these simplifications, this model successfully predicts the relative rates of electron transfer in a large number of experimental systems[4d] and provides the starting point from which the complicating effects of multiple pathways, loop structures in pathways, and many-electron effects can be investigated in a systematic manner.

The algorithm for determining the set of bonds that dominates this D–A interaction is the subject of this paper. Although based on a simple expression for the protein-mediated coupling, the model successfully predicts the relative rates of electron transfer in ruthenated cytochrome *c*,[4,6] myoglobin,[7] and cytochrome b_5.[8] The pathway model explains order of magnitude differences in couplings for specific metal-labeled proteins despite nearly identical D–A separation.[3d,4d,6-8]

The rate of nonadiabatic electron transfer from D to A is

[†] California Institute of Technology. Present Address: Massachusetts Institute of Technology, Mail Stop E34-201, Cambridge, MA 02139.
[‡] University of Pittsburgh.
[§] University of California.

(1) (a) Arteca, G. A.; Mezey, P. G. *Int. J. Quantum Chem.* **1988**, *34*, 517. (b) Mitchell, E. M.; Artymiuk, P. J.; Rice, D. W.; Willett, P. *J. Mol. Biol.* **1989**, *212*, 151. (c) Wilson, R. J.; Watkins, J. J. *Graph Theory, an Introductory Approach*; Wiley: New York, 1990; Chapter 8. (d) Buckley, F.; Harary, F. *Distance in Graphs*; Addison-Wesley: New York, 1990. (e) Maurer, S. B.; Ralston, A. *Discrete Algorithmic Mathematics*; Addison Wesley: New York; 1991; Chapter 3. (f) Balaban, A. T., Ed. *Chemical Applications of Graph Theory*; Academic Press: New York, 1976. (g) Trinajstic, N. *Chemical Graph Theory*; CRC Press: New York, 1983. (h) Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer-Verlag: Berlin, 1986.

(2) (a) Hopfield, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3640. (b) Jortner, J. *J. Chem. Phys.* **1976**, *64*, 4860. (c) DeVault, D. *Quantum Mechanical Tunneling in Biological Systems*, 2nd ed.; Cambridge University Press: New York, 1984.

(3) (a) Onuchic, J. N.; Beratan, D. N. *J. Chem. Phys.* **1990**, *92*, 722. (b) Beratan, D. N.; Onuchic, J. N. *Photosynth. Res.* **1989**, *22*, 173. (c) Beratan, D. N.; Onuchic, J. N.; Hopfield, J. J. *J. Chem. Phys.* **1987**, *86*, 4488. (d) Beratan, D. N.; Onuchic, J. N. In *Electron Transfer in Inorganic, Organic, and Biological Systems*; ACS Adv. Chem. Ser. No. 228, Bolton, J. R., Mataga, N., McLendon, G., Eds.; American Chemical Society: Washington, DC, 1991.

(4) (a) Beratan, D. N.; Betts, J. N.; Onuchic, J. N. *Science* **1991**, *252*, 1285. (b) Beratan, D. N.; Onuchic, J. N.; Betts, J. N.; Bowler, B. E.; Gray, H. B. *J. Am. Chem. Soc.* **1990**, *112*, 7915. (c) Beratan, D. N.; Onuchic, J. N.; Gray, H. B. In *Metal Ions in Biological Systems*; Sigel, H., Sigel, A., Eds.; Marcel Dekker Press: New York, 1991; Vol. 27, 97–127. (d) Onuchic, J. N.; Beratan, D. N.; Winkler, J. R.; Gray, H. B. *Annu. Rev. Biophys. Biomol. Struct.* In press.

(5) (a) Onuchic, J. N.; de Andrade, P. C. P.; Beratan, D. N. *J. Chem. Phys.* **1991**, *95*, 1131. (b) Kuki, A., preprint, 1991. (c) Siddarth, P.; Marcus, R. A. *J. Phys. Chem.* **1990**, *94*, 8430. (d) Broo, A.; Larsson, S. *J. Phys. Chem.* **1991**, *95*, 4925. (e) Christensen, H. E. M.; Conrad, L. S.; Mikkelsen, K. V.; Nielsen, M. K.; Ulstrup, J. *Inorg. Chem.* **1990**, *29*, 2808.

(6) (a) Bowler, B. E.; Meade, T. J.; Mayo, S. L.; Richards, J. H.; Gray, H. B. *J. Am. Chem. Soc.* **1989**, *111*, 8757. (b) Therien, M. J.; Selman, M. A.; Gray, H. B.; Chang, I.-J.; Winkler, J. R. *J. Am. Chem. Soc.* **1990**, *112*, 2420. (c) Bowler, B. E.; Raphael, A. L.; Gray, H. B. *Prog. Inorg. Chem.: Bioinorg. Chem.* **1990**, *38*, 259–322. (d) Wuttke, D. S.; Bjerrum, M. J.; Winkler, J. R.; Gray, H. B. *Science*, in press.

(7) Cowan, J. A.; Upmacis, R. K.; Beratan, D. N.; Onuchic, J. N.; Gray, H. B. *Ann. N.Y. Acad. Sci.* **1988**, *550*, 68.

(8) Jacobs, B. A.; Mauk, M. R.; Funk, W. D.; MacGillivray, R. T. A.; Mauk, A. G.; Gray, H. B. *J. Am. Chem. Soc.* **1991**, *113*, 4390.

0002-7863/92/1514-4043$03.00/0

$$k_{ET} = (2\pi/\hbar)|T_{DA}|^2(\text{F.C.}) \quad (1)$$

where (F.C.) is the Franck–Condon factor associated with the nuclear motion along the reaction coordinate. A single electron tunneling pathway is defined as a combination of interacting bonds that link D with A via covalent (C), hydrogen bonded (H), or through-space (S) connections. For a single path, the coupling is approximated as[3]

$$T_{DA} \propto \prod_i \epsilon_i^C \prod_j \epsilon_j^S \prod_k \epsilon k^H \quad (2)$$

The goal of the algorithm described here is to choose the combination of bonds between D and A that maximizes the product in eq 2. To provide a simple implementation of the pathway concept, to test its validity, and to show its predictive power, we chose the following parameters:[3,4]

$$\epsilon^C = 0.6 \quad (3a)$$

$$\epsilon^H = 0.36 \exp[-1.7(R - 2.8)] \quad (3b)$$

$$\epsilon^S = 0.6 \exp[-1.7(R - 1.4)] \quad (3c)$$

The distances, R, are in angstroms and the decay factors, ϵ, are unitless. These parameters are consistent with typical binding energies for electron-transfer-localized states as well as theoretical and experimental studies of model compounds.[3] Each decay factor ϵ is associated with an effective distance d_{eff} where

$$d_{eff} = -\log \epsilon \quad (4)$$

We will refer to both decay factors and connection lengths throughout the paper.

Maximum Coupling Pathways. The strength of the coupling arising from a single pathway is proportional to the product of decay factors for each step on the path: $\prod_i \epsilon_i$. The computational challenge before us is to analyze the highly interconnected network of bonded and nonbonded contacts in a protein and specify the bonds that maximize this product. This is precisely the well-known "minimum distance in a graph" problem. The minimum distance problem addresses finding the shortest pathway between two points in an interconnected network. Since eq 4 associates the decay factor with an effective distance, we can restate our search for the maximum pathway coupling as a search for the shortest effective distance between donor and acceptor in the corresponding network. General graph theory strategies for solving the minimum distance problem are discussed in refs 1d and 1e.

Methods

The first step in using graph theory to find electron-transfer pathways in proteins is to construct a labeled graph[1] corresponding to the superset of all interesting potential pathways. Covalent bonds (established as described below) are first mapped onto vertices.[9] Establishing which vertices are to be joined by edges requires progressively more computation for adjacent covalent bonds, hydrogen bonds, and potential through-space (TS) contacts. The lengths of the edges (i.e., the decays) are determined by the distances between the atoms and the nature of the interaction, eq 3. The covalent bonds are specified implicitly by the Brookhaven Protein Data Bank (PDB) files.[10] Covalent interactions, those between bonds anchored at a common atom, are easily identified. Existing software[11] is used to look up these connections for the known amino acids and other residues, which are then appended to the PDB data. Figure 1 outlines the chain of events between PDB file reading and pathway prediction. The degree of connectivity in the resulting graph is shown for a typical protein in Figure 2, and averages about 2.3 connections per atom. These amended PDB files are used as input to the PATHWAYS software[12] written by the authors. On the basis of data in the parameter files, the program looks up the model-predicted decays, eq 3, for the covalent bonds and stores them.

Hydrogen bonds are identified by the following criteria:[13] (1) hydrogen-donor and hydrogen-acceptor groups (donors, $-NH_n$; acceptors, carbonyl oxygens; both, –OH); (2) donor–hydrogen–acceptor angle ($\leq 90°$), and (3) donor–acceptor distance (≤ 3.5 Å). These values are specified in a parameter file. Edges representing the hydrogen bonds are

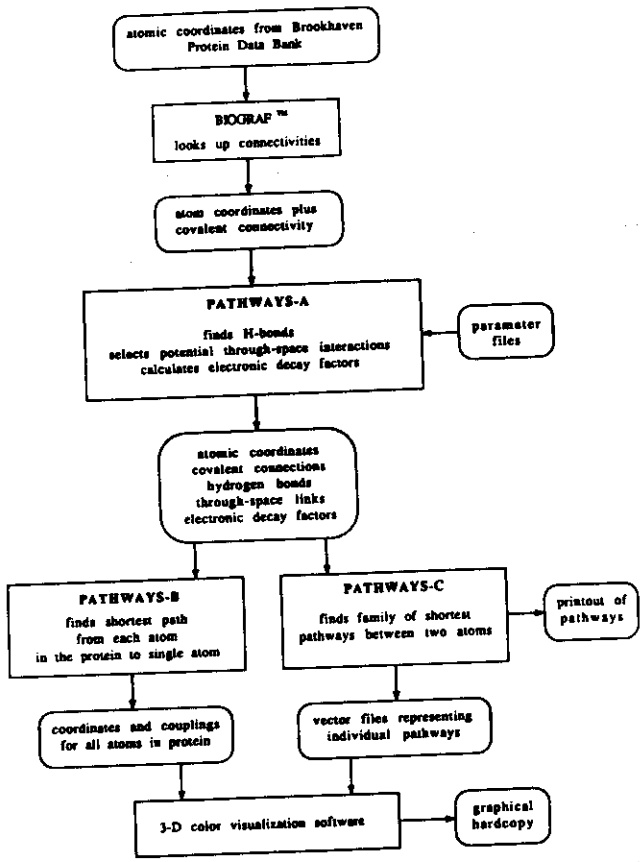


Figure 1. Information flow for calculating electron-transfer pathways in proteins. Rounded cells refer to data and square cells to processes performed by computer programs. PATHWAYS-A, -B, and -C are each part of the PATHWAYS program available from the authors.[12]

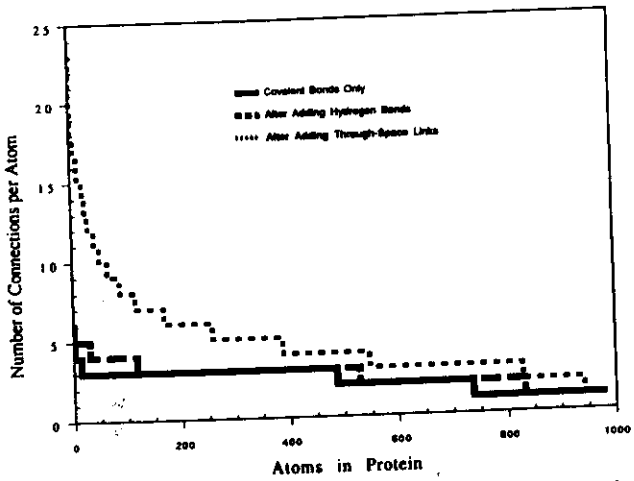


Figure 2. Distribution of connectivity for the heavy atoms in a typical protein (azurin) at various stages in the graph-building process. The connectivities from each stage are sorted on the basis of the number of connections to the atoms.

Hydrogen bonds are identified by the following criteria:[13] (1) hydrogen-donor and hydrogen-acceptor groups (donors, $-NH_n$; acceptors, carbonyl oxygens; both, –OH); (2) donor–hydrogen–acceptor angle ($\leq 90°$), and (3) donor–acceptor distance (≤ 3.5 Å). These values are specified in a parameter file. Edges representing the hydrogen bonds are

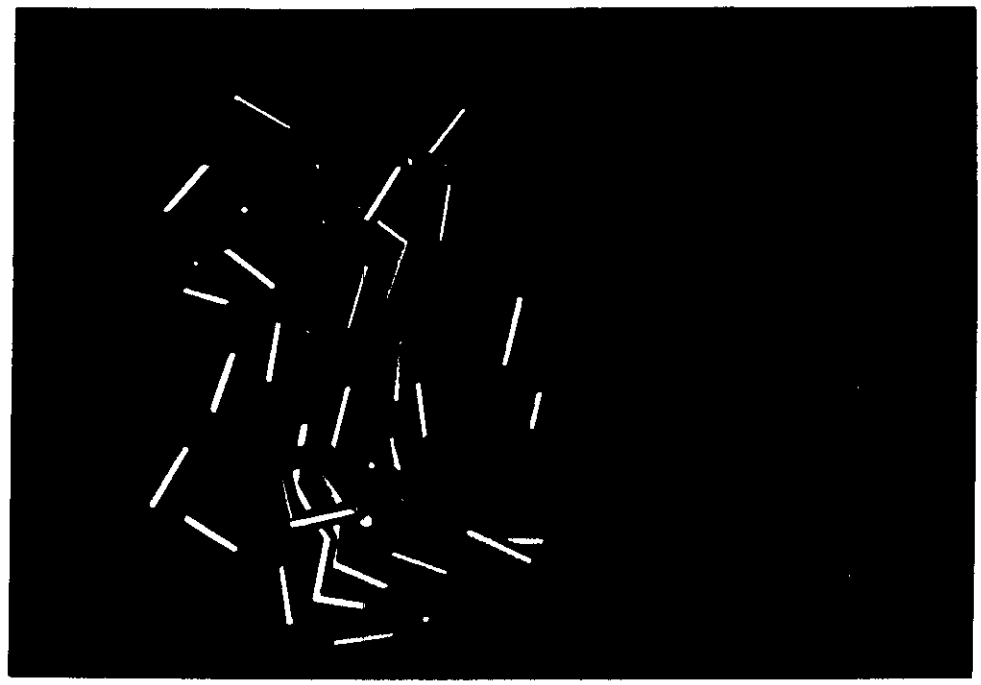
(9) These assignments are an obvious oversimplification. Inaccuracies in treatment of the through-space coupling are introduced by neglecting or adding some lone pair electrons, neglecting hydrogens bound to atoms other than heteroatoms, and suppressing through-space orientation effects. However, if all of these effects were included, corrections to the overall decay would likely be of order unity because there are very few through-space connections in the dominant paths. Errors in the through-space decays do not affect any of the qualitative predictions of the pathway analysis.

(10) Bernstein, F. C.; Koetzle, T. E.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, M.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.

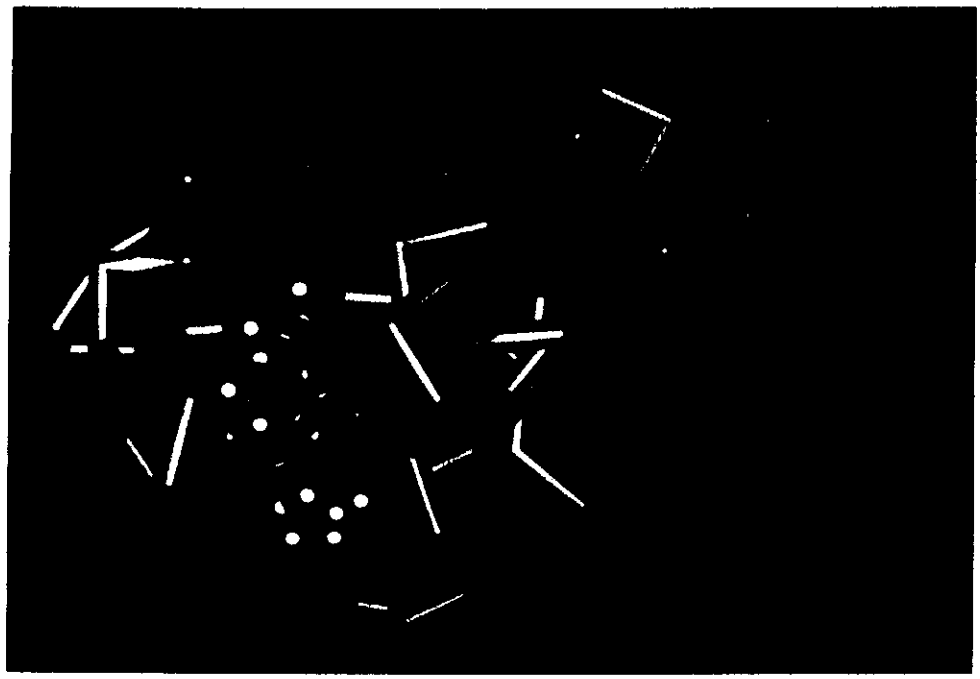
(11) For example, BIOGRAF, a product of Biodesign, Inc., Pasadena, CA 91101, was used here.

(12) The software (PATHWAYS v. 2.2) and user's manual are available from D.N.B.

(13) See, for example: Stryer, L. *Biochemistry*, 2nd ed.; W. H. Freeman and Co.: New York, 1981.



(a) Plastocyanin



(b) Cytochrome b5

Figure 3. Pathway coupling ratio maps are shown for (a) plastocyanin and (b) cytochrome b_5. Note that the antiparallel β-sheet (barrel) structure in plastocyanin provides "hot" spots in the strands ligating the Cu center but not in the other strands. In cytochrome b_5, however, the β-sheet structure (shown here behind the heme in a plane roughly perpendicular to it) does not radiate from the porphyrin, so it does not assist coupling along the full length of the protein as it does in plastocyanin. Displayed is $\prod \epsilon_i/[A \exp(-R\beta/2)]$ where the numerator is the pathway mediated coupling to an α-carbon and the denominator is the best fit exponential expression for T_{DA} for all α-carbons in the entire protein, evaluated for each α-carbon at distance R from the heme or Cu site.

added to the connection list, and the lengths that represent these decays are added to the list of segment lengths. This increase the degree of connectivity for the protein by about 0.25 per atom (see Figure 2).

Potential through-space (TS) connections are sought within a limited radius of each atom, typically 6 Å. It was found that no TS connections longer than this contributed to significant pathways, so the irrelevant long distance ones beyond this cutoff distance are eliminated to shorten the data-processing time. The TS connections are established for each atom, A, as follows. First, a list, L, containing all bonds/vertices within range of A is made and an attempt is made to eliminate as many of the entries as possible. Eliminating the TS connections between two atoms having a significantly better alternative through-bond connection was found to decrease the average added connectivity from about 21.3 to 1.9 per atom. The first connections the program eliminates from L are those that are redundant with preexisting covalent and hydrogen bonds. The vertices remaining in L are sorted on the basis of their distances from A, shortest first. Next, a depth-first shortest-path search[1] is performed (see next section) with A as the root, finding the shortest distance to F (the first

vertex in L) through the already existing connections. The depth of the search is limited to a length which corresponds to the TS decay from A to F. If the search returns without having located F, then the TS contact is the shortest path, and is thus added to the master connection (adjacency) list, and its corresponding length is added to the list of lengths, otherwise, F is discarded. Then the next vertex in L becomes the new F. In this way, shorter TS contacts may contribute to favorable paths and can disqualify longer ones, further decreasing the amount of connectivity added to the graph. The resulting change in connectivity is shown in Figure 2.

The Search Algorithm. There are two standard search strategies for arriving at the minimum-distance path between two points in an interconnected network, referred to as depth-first and breadth-first searches.[1] A depth-first search begins at a specified point and steps along allowed connections until no additional forward steps exist (a dead-end is reached) or the target site is found. If a dead-end occurs, the search backtracks by one step and then seeks alternative forward steps from that point, and so on until the target atom is found. A breadth-first search simultaneously considers all paths radiating from the starting point by keeping track of each vertex and its distance. At each step of the search a new vertex is added. The vertex chosen to be added is always the one that minimizes the effective distance to the donor at that stage. When the acceptor atom is the one that is added, the minimum distance pathway has been found. We use a depth-first algorithm in this work. The advantage of the depth-first search for our application is its "pathway orientation", i.e., each excursion represents a potentially acceptable pathway and the paths within a given factor of the best one are easily tabulated and accumulated.

Once the adjacency list is complete, and the lengths of the edges are calculated, the graph is ready for shortest-path searches to be executed. The depth-first search algorithm used in PATHWAYS can be described recursively in approximate terms as follows:

```
begin SEARCH(base, length)
   ispath(base) = true
   branch = 1
   probe = adj(base, branch)
   do while probe ≠ 0
      if length + len(base, probe) < sofar(probe) and ispath(probe) =
            false then
         call SEARCH(probe, length + len(base,probe))
      else
         branch = branch + 1
         probe = adj(base, branch)
      end if
   end do
   ispath(base) = false
return
```

where len(i, j) is a 2-D array containing the length of the jth edge connected to vertex i, sofar(i) is an array with the length of the shortest approach made to vertex i so far, ispath(i) is an array that notes whether or not vertex i is part of the currently searched path, and adj(i,j) is the adjacency list, a 2-D array holding the number of the jth vertex connected to vertex i.

After SEARCH(root, 0) is called, the sofar(i) array contains the shortest pathway from root to all other vertices (see Figure 1). In our actual implementation, recursion is not used, and a stack is explicitly maintained. This allows the pathways to be recorded mid-search, and the search to be terminated more easily.

SEARCH is used several times in the program. During the process of locating and eliminating TS connections, the search routine sets a flag and returns immediately if a path to a given target atom is found. The sofar array is not erased between searches from a given atom, so the searches accelerate progressively.

Searches are executed between two atoms or from one atom to all others in the protein. During searches of the entire protein, the routine is allowed to run to completion. The sofar(i) array is used to generate statistics such as a regression of the pathway-based couplings versus through-space distance. The sofar(i) array is also incorporated in output files which are used by our custom graphics-display software to view the electronic couplings as color-coded maps (Figure 3).

For searches between specific atoms, the routine is allowed to run to completion, *but* is interrupted whenever the target atom is encountered in order to record the current pathway. The criteria in this search are relaxed using a sloppiness parameter so that all paths within a variable factor of the best one are retained. Branches are only skipped if the length accumulated to reach them is longer than that atom's entry in the sofar(i) array *minus* the length specified by the sloppiness parameter. In this way, nearly equivalent pathways will not prevent one another from being found. Thus, families of pathways are recorded. After the call to SEARCH, pathways and their lengths are output as tabular reports and as graphics-compatible files.

Discussion

We have described a search algorithm to find electron tunneling pathways with maximal coupling given a simple prescription for through-bond and through-space electronic decay. The method has been used with success to predict relative rates of transfer in several transition metal labeled proteins. The capability of performing global searches for best pathways in a protein from a single site (for example donor or acceptor) to all heavy atoms allows (1) the construction of global protein coupling maps, (2) the identification of "hot" and "cold" spots[4a] for electron transfer at a given distance, and (3) determination of secondary and tertiary motif effects on the coupling. [Hot and cold spots are defined by fitting the pathways couplings for every site in a specific protein to a single exponential in distance to determine the average decay. Sites that are coupled more strongly (weakly) compared to the average value for that distance are termed hot (cold).] Equipped with improved bond and orientation dependent ϵ values, the algorithm could provide lists of the lowest-order perturbation theory pathways for a given level of electronic structure theory (e.g., extended-Hückel).

A key test of the theory involves the attachment of transition metal probes to residues at similar distances that are predicted by the pathway model to have vastly different coupling.[6] The blue copper proteins are systems in which dramatic effects are predicted. Figure 3 shows hot and cold spots in plastocyanin and cytochrome b_5. In plastocyanin, hot spots radiate from the β-strands ligating the Cu. Shorter-range hot spots in cytochrome b_5 are associated with amino acid/heme hydrogen bonding interactions. By changing the protein interactions with the redox centers or by modifying the electronic structure of the ground/excited states (heme or Cu orbitals), it should be possible to change the rates as well (via the prefactors not explicitly in eq 2).

The pathway method has pointed to anomalous electron transfer rates in some systems,[4a,6c] which are now being investigated in further detail experimentally. Further application of this search algorithm should provide a deeper understanding of electron transfer reactions in proteins and nucleic acids, and the manipulation of pathways may also allow the design of stabilized high-energy charge-separated species for more efficient energy conversion schemes.[14] The method is now being refined to include multiple interfering pathways and bond type differences.[5a]

Programming Environment

The software [12] was developed using Silicon Graphics FORTRAN under the IRIX (UNIX) operating system on a Silicon Graphics IRIS 4D/210 VGX. The software will run on any Silicon Graphics IRIS, and should be portable to most UNIX systems supporting FORTRAN. BIOGRAF [11] is used to create the covalent list, but this file could also be generated by other means. Timing for albacore cytochrome c on the 4D/210: 50 sec to construct connection list; 9 sec to calculate all best paths to the heme.

Acknowledgment. This work was performed in part at the Jet Propulsion Laboratory, California Institute of Technology and was sponsored by the Department of Energy's Catalysis/Biocatalysis Program (Advanced Industrial Concepts Division), through an agreement with the National Aeronautics and Space Administration. J.N.O. thanks the National Science Foundation (Grant No. DMB-9018768) and the Department of Energy's Catalysis/Biocatalysis Program (through a research contract from the Jet Propulsion Laboratory) for support of this work. The pathway search software, written in the FORTRAN for Silicon Graphics IRIS computers, is available from D.N.B. J.N.O is in residence at the Instituto de Física e Química de São Carlos, Universidade de São Paulo, 13560, São Carlos, SP, Brazil, during the summers.

Registry No. Cytochrome b_5, 9035-39-6.

(14) Beratan, D. N.; Betts, J. N.; Onuchic, J. N. *J. Phys. Chem.* In press.

Reprint Series
31 May 1991, Volume 252, pp. 1285–1288

SCIENCE

Protein Electron Transfer Rates Set by the Bridging Secondary and Tertiary Structure

D. N. BERATAN, J. N. BETTS,* AND J. N. ONUCHIC

Protein Electron Transfer Rates Set by the Bridging Secondary and Tertiary Structure

D. N. BERATAN, J. N. BETTS,* J. N. ONUCHIC

The rate of long-distance electron transfer in proteins rapidly decreases with distance, which is indicative of an electron tunneling process. Calculations predict that the distance dependence of electron transfer in native proteins is controlled by the protein's structural motif. The helix and sheet content of a protein and the tertiary arrangement of these secondary structural units define the distance dependence of electronic coupling in that protein. The calculations use a tunneling pathway model applied previously with success to ruthenated proteins. The analysis ranks the average distance decay constant for electronic coupling in electron transfer proteins and identifies the amino acids that are coupled to the charge localization site more strongly or weakly than average for their distance.

MANY BIOLOGICAL ELECTRON transfer reactions involve charge transport over considerable distance (>5 Å). Therefore, electron tunneling occurs through the protein, and the rate of transport is proportional to the square of the electronic coupling between donor and acceptor species (*1*). We recently developed a tunneling pathway model for electron transfer in proteins that identifies the dominant bonding and nonbonding interactions responsible for the donor-acceptor coupling (*2*). The model successfully predicts the relative rates of electron transfer in ruthenated cytochrome c, myoglobin, and cytochrome b_5. In the cytochrome c and b_5 derivatives, order of magnitude rate differences between isomers with nearly identical transfer distances were successfully explained with the pathway model (*3*).

In weakly coupled donor-acceptor systems, the rate of electron transfer is

$$k_{ET} = \frac{2\pi}{\hbar}|T_{DA}|^2(\text{F.C.}) \qquad (1)$$

where T_{DA} is the electronic tunneling matrix element between donor and acceptor localized states and (F.C.) is the Franck-Condon factor determined by the activated (or tunneling) nuclear processes coupled to the transport. A physical tunneling pathway is defined as a combination of interacting bonds that link donor with acceptor. Segments of the pathway are characterized as covalent (*C*), hydrogen-bonded (*H*), or through-space (*S*), depending on whether the bonds share a common atom (*C*), are linked by a hydrogen bond (*H*), or are in van der Waals contact (*S*). The pathway model assumes that noninteracting tunneling pathways dominate T_{DA}, and the numerical implementation of the model seeks those dominant paths (*3*). The contribution to the donor-acceptor electronic coupling mediated by an individual physical tunneling pathway is as follows (*2*):

$$\frac{\beta_A\beta_D\beta_1}{(E-\alpha_L^1)(E-\alpha_R^1)-\beta_1^2}\prod_{i=2}^{Nc}\epsilon_i^C\prod_{j=1}^{Ns}\epsilon_j^S\prod_{k=1}^{NH}\epsilon_k^H \qquad (2)$$

where α_L^1 and α_R^1 are the orbital energies of the two hybrid atomic orbitals in the first bond of the pathway, and the subscripted values of β are coupling matrix elements between orbitals at the chain ends (*2, 3*). The energy of the tunneling electron is *E*.

We recently implemented a graph-search algorithm to determine the dominant tunneling pathways in proteins and to calculate their relative electronic couplings. The following values of the decay parameters in Eq. 2 were used (*3*):

$$\epsilon^C = 0.6 \qquad (3a)$$

$$\epsilon^H = 0.36 \times \exp[-1.7(R-2.8)] \qquad (3b)$$

$$\epsilon^S = 0.6 \times \exp[-1.7(R-1.4)] \qquad (3c)$$

The distances, *R*, are in angstroms, and the decay factors, ϵ, are unitless (*4*). We estimated the through-bond decay factor by using data from existing model compounds and protein electron transfer rates. We calculated the through-space decay factor, 1.7 $Å^{-1}$, using the known typical binding energies for the localized states, and the H bond decay factor arises from approximating the H bond as two stretched covalent bonds (*2, 3*). The parameterization neglects bond type differences, but for mapping the residues in proteins that mediate tunneling and estimating relative couplings, this method is adequate. Differences in the ϵ values for various bond types are small enough that this approximation is adequate for these initial calculations. In fact, the nature of the predicted pathways is insensitive to the details of the parameters if they are chosen in a physically realistic range. The model explains a substantial amount of experimental data that are inconsistent with simpler single exponential decay expressions for T_{DA} (*3*). Typically only a few viable pathways or families of pathways exist because of the relatively weak interaction between nonbonded groups.

Earlier estimates of coupling matrix elements in proteins relied on the calculation of tunneling barrier heights based on optical properties or electronic structure calculations on simplified systems (*5*). More recently, strategies that treat details of the polypeptide electronic structure have been under study (*3, 6*). The pathway search method outlined above is apparently the first to allow global searches from a single center (for example, Cu atom, Fe-S cluster, or metalloporphyrin) to every other (nonhydrogen) atom in the protein. These searches are possible because of the relative simplicity of the model. The calculation is broken into two phases. First, bonded and nonbonded connections within a radius (usually 6 Å) of each atom are identified and their coupling is calculated. Local through-space connections are eliminated if stronger bonded links exist between the atoms. More than half of the computing time is spent establishing and screening the local connections. The connection-coupling list is used by the program to seek pathways between the donor and acceptor that maximize the product in Eq. 2.

An expanding set of experimental data is now emerging for electron transfer in proteins between residues at fixed distances. Surface modification with transition metal complexes, chemical modification of chromophores in multisubunit proteins, site-directed mutagenesis, and semisynthesis have provided a wide variety of electron transfer proteins with

D. N. Beratan, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, and The Beckman Institute, California Institute of Technology, Pasadena, CA 91125.
J. N. Betts, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109.
J. N. Onuchic, Department of Physics, University of California, San Diego, La Jolla, CA 92093.

*Present address: Woods Hole Oceanographic Institute, E34-201, Woods Hole, MA 02139.

Table 1. Best fits of pathway couplings to $\exp[-(\beta/2)R_{DA}]$. Distances in heme proteins are measured to the closest atom in the porphyrin ring. The 95% confidence limits on $\beta/2$ are ± 0.03 Å^{-1}. The standard deviation of the decay length fit for each protein is σ (*9*). The structures were taken from the Brookhaven Protein Data Bank coordinate files of *Pseudomonas* cytochrome c_{551} (351c and 451c), sperm whale myoglobin (1mbo and 5mbn), Bovine cytochrome b_5 (2b5c), albacore cytochrome c (3cyt and 5cyt), *Chromatium* high-potential Fe protein (1hip), *Alcaligenes* azurin (2acu, some residue data not available), and poplar plastocyanin (1pcy and 5pcy). Values of $\beta/2$ for the oxidized and reduced structures of a single protein (for cytochrome c and plastocyanin) differ by no more than 0.01 Å^{-1}. The same is true of $\beta/2$ in oxymyoglobin and deoxymyoglobin.

Protein	$\beta/2$ (Å^{-1})	σ ($\beta/2$)	Helix (%)	Sheet (%)
Cytochrome c_{551}	0.76	0.22	51	
Cytochrome b_5	0.73	0.22	57	29
Myoglobin	0.71	0.22	79	
Cytochrome c	0.61	0.21	51	
HiPIP	0.60	0.16	11	16
Azurin	0.55	0.16	14	47
Plastocyanin	0.49	0.19	5	62

known bridging structures (*7*) to which this method can be applied.

For tunneling through a structureless one-dimensional barrier between localized vibronic states, the rate is

$$k'_{ET} = A^2 \exp(-\beta R_{DA})(\text{F.C.}) \qquad (4)$$

where R_{DA} is the donor-acceptor distance and the tunneling matrix element is $T_{DA} = A \exp(-R_{DA}\beta/2)$. Experimental data from a variety of proteins with different donors and acceptors are often fitted to this expression to estimate β. The barrier height is determined by the redox potentials of the donor and acceptor and the electronic structure of the protein. To connect our method with this common formulation of the rate, and to calculate the average decay length resulting from the pathway analysis of a specific protein, we calculated the best fit exponential decay constant for T_{DA} using the couplings calculated with the pathway model. The pathway coupling for a specific metalloprotein was calculated between the transition metal site and every other nonhydrogen atom in the protein (846 sites in albacore cytochrome c). Myoglobin, cytochrome c, cytochrome c_{551}, cytochrome b_5, high-potential iron protein (HiPIP), azurin, and plastocyanin were examined (*8*). We fitted the single exponential expression for each protein, using the largest pathway coupling found between the native charge localization site and each other nonhydrogen atom (correlation coefficients ~0.85 to 0.90). The atomic positions were fixed at the crystallographic coordinates. The matrix element decay factor β was found to be protein-dependent; the results appear in Table 1. The variation in β arises simply from connectivity differences in the proteins. This effect is distinct from others that might arise from differences in the electron tunneling energy or the atom types and hybridization between proteins (*5*). The pathways are calculated from the point in the heme ring edge nearest to the target atom, from the Cu site to the target atom, or from the Fe-S cluster to the target atom (the HiPIP values are averages of eight independent sets of searches from each atom in the Fe-S cluster).

The values chosen for ϵ^C (0.6) and the through-space decay factor (1.7 Å^{-1}) in Eq. 3 determine the maximum and minimum possible values for β; coupling can decay no faster than it would decay purely through space and no more slowly than it would if there were a fully extended bonded chain between donor and acceptor. For a covalent chain, $\epsilon^C = 0.6$ generates $\beta/2 = 0.42$ Å^{-1} (we assume tetrahedral atoms). The calculated pathway-mediated values of $\beta/2$ vary from 0.76 to 0.49 Å^{-1}, or 73 to 95% of the bond-mediated value.

The marked difference between the heme and blue Cu protein distance dependence arises from secondary and tertiary structure differences. One can understand this difference by examining the atoms that are especially strongly or weakly coupled (for their distance) in the proteins. The ratio (ζ_N) of the largest pathway coupling to site N to the coupling calculated with the average value of $\beta/2$ fitted for that specific protein, identifies sites that are more strongly or more weakly coupled than average for their distance.

$$\zeta_N = \frac{\prod_i \epsilon_i}{A \exp[-R_{DN}\beta/2]} \qquad (5)$$

Atoms with $\zeta_N > 1$ are called "hot" spots, and those with $\zeta_N < 1$ are called "cold" spots. Maps of ζ_N and the pathway couplings appear in Fig. 1. In the heme proteins, the smallest values of ζ_N are found for amino acid atoms in mid-helix, and the largest values of ζ_N occur for amino acids directly coupled to the porphyrin or near turns between helical segments. In the blue Cu β-barrel proteins, β strands coordinated to the Cu have many atoms with $\zeta_N > 1$ in the entire strand, and the ratio is generally somewhat less than one

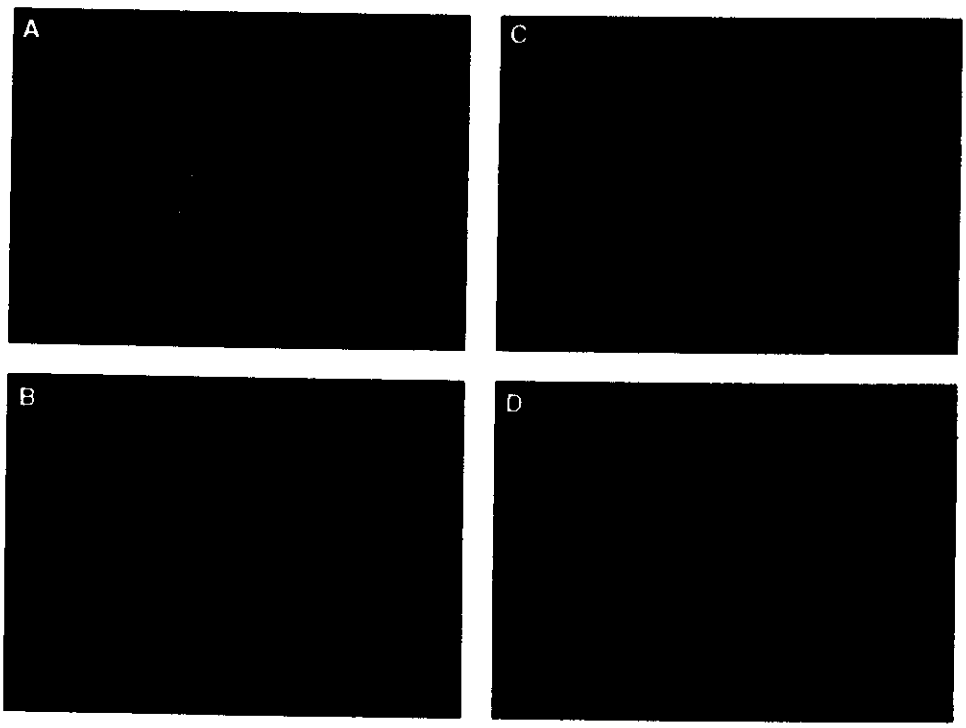


Fig. 1. Labeled traces of the α-carbons in azurin (**A**) and cytochrome c (**B**) showing the strength of the pathway coupling (Eq. 5 numerator only) to the blue Cu or heme sites (red signifies strong coupling, blue weak, and green intermediate), which decays approximately exponentially with distance. Also shown are the sites with "hot" (red), "cold" (blue), or "average" (green) electronic coupling to the Cu or heme center given their separation (Eq. 5) in azurin (**C**) and cytochrome c (**D**). Residues in the β barrel of azurin are "hot" or "cold" for their distance (depending on whether or not the strand contains a strong connection to the Cu), while those in the α-helical region of the protein are weakly coupled for their distance. In cytochrome c, amino acids near the porphyrin axial ligands, the porphyrin covalent bonds to the protein, and the porphyrin hydrogen bonds to the protein are more strongly coupled than expected for their distance. Amino acid numbers are shown.

in the other β strands.

The greater anisotropy of ζ_N in the secondary structural units of helical proteins compared to β-barrel proteins is also reflected in the larger standard deviation of the fitted value of β/2 (Table 1) for the helical proteins as compared to those of β-barrel structures (*9*). In the β-barrel proteins, a considerably smaller standard deviation is associated with each single (nearly linear) strand that spans the structure from the top of the barrel to the bottom.

Another way to understand the anisotropy of electronic coupling in the helical versus β-barrel proteins is to consider searches from all atoms in the protein to a single probe atom (with the probe not at the Cu or heme site). We have carried out such searches in azurin and myoglobin. In azurin, the fitted value of β/2 is insensitive to the probe atom position when it is in the β-barrel structure (slightly smaller values for β/2 occur for probe atoms near the top or bottom of the barrel and slightly larger values occur for probe atoms near the centers of the strands). This differs from the result in myoglobin, in which β/2 depends on the location of the probe atom in the helix. Even in azurin, if the probe atom is in the helical segment of the protein (Ala_{60} for example), β/2 increases to values typical of the heme proteins. These calculations illustrate that helical and β-barrel motifs differ in both their efficiency and anisotropy as tunneling mediators. The origin of these structural effects lies in the accessibility or inaccessibility of the chosen residue to all other groups through efficient pathways (Fig. 2).

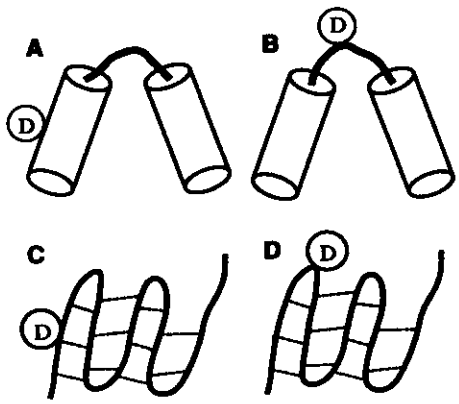


Fig. 2. Large differences in average coupling between one probe site (D) and the rest of the protein occur in helical proteins when the probe is at a helix center (**A**) or at a turn between helical regions (**B**) because of the differential accessibility of these sites through strongly coupled pathways. Cylinders represent helical segments of protein. Amino acids at turns are, on average, more strongly coupled to the rest of the protein than those in other positions. Small differences in average coupling are seen in β-barrel proteins for a probe in the middle of a strand (**C**) versus the turn between strands (**D**).

The three-dimensional structure of a protein sets limits on the average decay of the tunneling matrix element in the protein. Simple models give the upper limit of the decay as proportional to the square root of the binding energy of the transferring electron, that is, the decay for tunneling through vacuum. The lower bound on the decay is related to the electronic coupling between neighboring covalent bonds and the tunneling energy relative to the bond energies (*2, 3*). The decay of the coupling matrix element in specific proteins falls between these two values and depends on the secondary and tertiary protein structure. If the unique charge localization site (the Cu, porphyrin, or Fe-S center) is embedded in a highly interconnected β-sheet structure or is at the turn between α-helical chains, it couples more strongly with the protein than if it were embedded in the middle of an α-helical segment. A substantial qualitative difference is seen, therefore, between the average coupling of heme and blue Cu proteins. Moreover, amino acids in β strands directly ligated to the Cu in the β-barrel proteins are predicted to be more strongly coupled than those at a nearly identical distance but on β strands (or helical segments) not coordinating the Cu.

Amino acids exist in electron transfer proteins that are anomalously strongly or weakly coupled (ζ_N much larger or smaller than 1) for their distance from the charge localization site. Experimental evidence of this behavior was recently reported in ruthenated cytochrome c (*3, 7*). In those experiments, the pathway model accounts for observed rate differences not predicted with simple exponential decay expressions for T_{DA}. Differences in the average distance decay of the tunneling matrix element for various proteins should substantially affect observed electron transfer rates because β/2 appears as an exponential contribution. The set of protein pathway couplings that produced the average decay constants in Table 1 were used to produce coupling maps and maps showing anomalously strongly or weakly coupled residues for their distance. Figure 1 shows these maps for cytochrome c and azurin. The coupling maps display the roughly exponential decay of coupling with distance. In the ratio maps (Eq. 5), amino acids predicted to be "hot" ($\zeta_N > 1$) or "cold" ($\zeta_N < 1$) with respect to electron transfer (given their distance) are easily identified. These figures show α-carbons color-coded according to the pathway coupling or ζ_N value (tubes connecting the α-carbons show the connectivity of the protein and are color-coded according to the value for the α-carbon nearest the COOH-terminus of the protein). The extent to which the protein secondary and tertiary structure provides relatively direct (almost linear) pathways that radiate from the charge localization site to a particular amino acid determines whether it is "hot" or "cold." For example, cytochrome b_5 consists of 29% β sheet, but the sheet structure runs in a plane perpendicular to the porphyrin and is not well connected to it, so the average distance decay (Table 1) is close to that of the other highly helical proteins, and no particularly "hot" spots exist in the β-sheet region. The heme in cytochrome c_{551} is the most weakly coupled of all the proteins studied, apparently because of the relatively short and randomly oriented helical segments in the protein.

Improved estimates of ϵ^C for specific bonds are emerging from quantum chemical calculations (*10*), as are new methods for summing the contributions to T_{DA} from intersecting pathways (*11*). This work should produce more reliable predictions of T_{DA} for proteins. The reported trends in β/2 are expected to be generic because the values of ϵ and their decay with distance were chosen in a physically reasonable range, and many protein pathways are not highly interconnected.

Bimolecular electron transfer in proteins may be mediated by interactions in a single docking configuration or a family of configurations; in either limit, the pathway model can yield testable predictions of the couplings. One might expect the specificity of electron transfer reactions to be controlled, to some extent, by the structural motifs of the proteins that surround the redox centers. Proteins with different motifs are predicted to display average distance dependencies that are qualitatively different, and the coupling at a given distance is expected to be somewhat anisotropic. The availability of theoretical electronic coupling maps for proteins with known structures should assist in future molecular design projects.

REFERENCES AND NOTES

1. R. A. Marcus and N. Sutin, *Biochim. Biophys. Acta* **811**, 265 (1985); D. DeVault, *Quantum Mechanical Tunneling in Biological Systems* (Cambridge Univ. Press, New York, ed. 2, 1984).
2. J. N. Onuchic and D. N. Beratan, *J. Chem. Phys.* **92**, 722 (1990); D. N. Beratan and J. N. Onuchic, *Photosynth. Res.* **22**, 173 (1989); ———, J. J. Hopfield, *J. Chem. Phys.* **86**, 4488 (1987).
3. D. N. Beratan, J. N. Onuchic, J. N. Betts, B. E. Bowler, H. B. Gray, *J. Am. Chem. Soc.* **112**, 7915 (1990); D. N. Beratan, J. N. Onuchic, H. B. Gray, in *Metal Ions in Biological Systems*, H. Sigel and A. Sigel, Eds. (Dekker, New York, 1991), vol. 27, pp. 97–127.
4. One can calculate the ϵ's for a physical pathway using perturbation theory, band theory neglecting the influence of side chains, or exact methods for a single physical pathway. Multiple pathways and

interference effects have been considered and appear to be of limited importance when one uses these qualitative arguments.

5. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3640 (1974); J. Jortner, *Biochim. Biophys. Acta* **594**, 139 (1980); D. N. Beratan, J. N. Onuchic, J. J. Hopfield, *J. Chem. Phys.* **83**, 5325 (1985); D. N. Beratan and J. J. Hopfield, *J. Am. Chem. Soc.* **106**, 1584 (1984).
6. A. Kuki and P. G. Wolynes, *Science* **236**, 1647 (1987); A. Broo and S. Larsson, *Int. J. Quantum Chem. Quant. Biol. Symp.* **16**, 185 (1989).
7. J. A. Cowan, R. K. Upmacis, D. N. Beratan, J. N. Onuchic, H. B. Gray, *Ann. N.Y. Acad. Sci.* **550**, 68 (1988); B. E. Bowler, T. J. Meade, S. L. Mayo, J. H. Richards, H. B. Gray, *J. Am. Chem. Soc.* **111**, 8757 (1989); M. J. Therien, M. A. Selman, H. B. Gray, I.-J. Chang, J. R. Winkler, *ibid.* **112**, 2420 (1990); S. E. Peterson-Kennedy, J. L. McGourty, J. A. Kalweit, B. M. Hoffman, *ibid.* **108**, 1739 (1986); D. W. Conrad and R. A. Scott, *ibid.* **111**, 3461 (1989); B. Durham, L. P. Pan, J. E. Long, F. Millett, *Biochemistry* **28**, 8659 (1989); O. Farver and I. Pecht, *FEBS Lett.* **244**, 379 (1989); M. P. Jackman, J. McGinnis, R. Powls, G. A. Salmon, A. G. Sykes, *J. Am. Chem. Soc.* **110**, 5880 (1988); B. A. Jacobs *et al.*, *ibid.*, in press.
8. Heteroatom hydrogens were added to the crystallographic coordinates of the other atoms using the software BIOGRAF: BioDesign, Inc., Pasadena, CA 91101.
9. We define the standard deviation of $\beta/2$

$$\sigma(\beta/2) = \left\{(1/N)\sum_i \left[\frac{\ln A - \ln \pi_i}{R_i} - \frac{\beta}{2}\right]^2\right\}^{1/2} \quad (6)$$

where π_i is the pathway coupling to site i in the protein, which contains N nonhydrogen atoms.
10. V. Balaji, L. Ng, K. D. Jordan, M. N. Paddon-Row, H. K. Patney, *J. Am Chem. Soc.* **109**, 6957 (1987).
11. J. N. Onuchic, P. C. P. Andrade, D. N. Beratan, *J. Chem. Phys.*, in press; C. Goldman, *Phys. Rev. A* **43**, 4500 (1991).
12. This work was performed in part at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Department of Energy's Catalysis/Biocatalysis Program (Advanced Industrial Concepts Division) through an agreement with the National Aeronautics and Space Administration. We thank the National Science Foundation and the Conselho Nacional de Des envolvimento Científico e Tecnológico (Brazil) for a binational research grant that allowed international visits during which this work was initiated. Work in San Diego was funded by a research contract from the Jet Propulsion Laboratory, supported by the Department of Energy's Catalysis/Biocatalysis Program and the National Science Foundation (grant DMB-9018768). The pathway search software, written in FORTRAN for Silicon Graphics IRIS computers, is available from D.N.B. at the Beckman Institute address. J.N.O. is in residence at the Instituto de Física e Química de São Carlos, Universidade de São Paulo, 13560, São Carlos, São Paulo, Brazil during the summers.

19 November 1990; accepted 14 March 1991

Electron tunneling pathways in proteins: A method to compute tunneling matrix elements in very large systems

José Nelson Onuchic and Paulo C. P. de Andrade
Department of Physics, University of California, San Diego, La Jolla, California 92093 and Instituto de Física e Química de São Carlos, Universidade de São Paulo, 13560 São Carlos, S. P., Brazil

David N. Beratan
Jet Propulsion Laboratory and Beckman Institute, California Institute of Technology, Pasadena, California 91109

(Received 14 January 1991; accepted 5 April 1991)

A tight-binding Hamiltonian and Dyson's equation method are described that allow the computation of the tunneling matrix elements between electron donor and acceptor sites in a protein. The method is exact and computationally tractable. The Green's function matrix elements of the bridge are computed using a strategy that builds up the bridge one orbital at a time, allowing inclusion of all orbitals on proposed tunneling pathways and elsewhere. The tunneling matrix element is determined directly from the bridge Green's function. A simple representation of a helical protein segment is used to illustrate the method and its ability to include contributions from high-order backscattering and multiple pathway interference in the donor–acceptor coupling.

I. INTRODUCTION

Electron transfer rates in chemical and biological compounds are modulated by electronic and nuclear factors.[1] Recent work has emphasized the separability of these factors (Born–Oppenheimer and Condon approximations), and reduction to a two-level electronic system.[1(d),2,3(b)] If these approximations are valid and the donor and acceptor states are energetically separated from the other electronic states, a two-level electronic model is valid. The effective renormalized Hamiltonian is

$$H_{ET} = T_{DA}\sigma_x + \frac{1}{2}\delta\sigma_z + H_Q. \tag{1}$$

σ_x and σ_z are the Pauli matrices. $\sigma_z = \pm 1$ is associated with the donor and acceptor localized states, H_Q provides the nuclear dynamics, δ is the instantaneous energy difference between the reactant and product states, and T_{DA} is the tunneling matrix element between the donor and acceptor electronic states.[1-4]

In this paper we analyze the electronic part of this problem and its reduction to a two-level system composed of localized effective donor and acceptor sites coupled by a tunneling matrix element. (In the above approximation, the tunneling matrix element is computed with frozen nuclear coordinates such that $\delta = 0$). For most proteins, the electronic coupling is small enough that the rate is nonadiabatic,[3,4] so the transfer rate is proportional to T_{DA}^2.

Recently, we developed a model to compute tunneling matrix elements and map key mediating bonds in proteins by generalizing approaches used on small molecules.[5] We use a one-electron tight-binding Hamiltonian [see Ref. 5(d)], with mixing between the donor(D) and acceptor(A) sites provided by a bridge. The effective donor and acceptor orbitals in the two-state approximation are a linear combination of all of these orbitals, but are dominated by the D and A sites.[1(e),2] The electronic Hamiltonian for D, A, and bridge is

$$H_{el} = \alpha_D a_D^\dagger a_D + \alpha_A a_A^\dagger a_A + \sum_{i_D} v_{Di_D}(a_D^\dagger a_{i_D} + a_{i_D}^\dagger a_D) + \sum_{i_A} v_{Ai_A}(a_A^\dagger a_{i_A} + a_{i_A}^\dagger a_A) + \sum_i \alpha_i a_i^\dagger a_i + \sum_{j>i} v_{ij}(a_i^\dagger a_j + a_j^\dagger a_i), \tag{2}$$

where the $a_\mu^\dagger$ (a_μ) creates (destroys) an electron in the μth orbital. The first two terms in the Hamiltonian represent the donor and acceptor sites. The third (fourth) term contains the coupling between the donor (acceptor) and the bridge. Bridge orbitals coupled to the donor (acceptor) are labeled i_D (i_A). The last two terms are the bridge Hamiltonian. We have not yet converted to the effective donor and acceptor states from which T_{DA} is calculated. The Hamiltonian is general in the sense that each orbital can be coupled to any bridge orbital. Each bridge "site" may have one or more orbitals localized on it, and a basis of atomic or bonding/antibonding orbitals can be used.

This electronic Hamiltonian can be reduced to a two-level system if the splitting between the donor and acceptor localized states is small relative to their energetic distance from the bridge (delocalized) states. All electronic transitions (except the donor to acceptor transition) must occur at a large energy compared to nuclear energies coupled to the transfer. For example, the reorganization energy has to be small compared to this energetic distance. If this is not the case, transport of the electron might result in electronic excitation of the bridge itself.[2,3] In this limit, the effective electronic Hamiltonian is[2]

$$H_{el}^{eff} = \alpha_D^{eff} a_D^\dagger a_D + \alpha_A^{eff} a_A^\dagger a_A + T_{DA}(a_D^\dagger a_A + a_A^\dagger a_D), \tag{3}$$

where

$$\alpha_{D(A)}^{eff} = \alpha_{D(A)} + \Delta_{D(A)}, \tag{4a}$$

 0021-9606/91/141131-06$03.00

$$\Delta_{D(A)} = \sum_{ij} v_{D(A)i} G_{ij}(E) v_{jD(A)}, \tag{4b}$$

and

$$T_{DA} = \sum_{ij} v_{Di} G_{ij}(E) v_{jA}. \tag{5}$$

The necessity of reducing to a two-level system to calculate T_{DA} has recently been reemphasized.[2,3,6,7] G is the Green's function for the bridge, i.e., the Green's function[8] associated with H_{el} without the donor and acceptor terms, and E is the tunneling electron energy. It is important to note that E is typically neither the effective donor or acceptor state energy, but is some intermediate value set by vibronic coupling,[3(d)] a detail often suppressed when reporting tunneling matrix elements obtained from molecular orbital calculations. For example, if all of the vibronic coupling is on the donor, the tunneling energy is α_A^{eff}. Computation of G is the subject of Secs. II and III. The electronic Hamiltonian is written in second quantized rather than spin notation for convenience. For example, the operator $a_D^\dagger a_A + a_A^\dagger a_D$ in Eq. (3) is equivalent to the spin operator σ_x.

Other Green's function strategies for calculating T_{DA} have been communicated recently.[9] One uses a Green's function based perturbation theory to calculate coupling.[9(b)] Another method, similar to that used here, calculates the exact coupling for a simple linear chain.[9(c)] The method described below is both exact and able to include a large number of orbitals without restricting their connectivity.

We defined a physical tunneling pathway as a collection of interacting bonds between and around the donor and acceptor that makes some contribution to the donor–acceptor coupling.[10,11] The single pathway contribution to the tunneling matrix element is

$$t_{da} = \text{prefactor} \prod_{i=1}^{N_B} \epsilon_i(\text{B}) \prod_{j=1}^{N_H} \epsilon_j(\text{H}) \prod_{k=1}^{N_S} \epsilon_k(\text{S}), \tag{6}$$

where ϵ is the tunneling matrix element decay across a covalent bond (B), hydrogen bond (H), or nonbonded contact (S). Approximations to these decay parameters allowed the mapping of specific pathways in synthetically modified proteins and the successful correlation with experimental rates (not provided by simple exponential expressions for T_{DA}).[10-12]

The contribution to the tunneling matrix element from a single pathway [Eq. (6)] is equivalent to reducing the bridge to a selected chain of sites in the protein between the donor and acceptor. Strategies to derive improved expressions for the ϵ's used to calculate t_{da} *for a single tunneling pathway* have been sketched.[11] For example, the interaction parameters common to extended-Hückel theory and a perturbation approach can be used.[9(b)] A substantial improvement is described in this paper, namely a method that allows (1) inclusion of multiple intersecting physical pathways in the T_{DA} calculation and (2) exact, not perturbation theory, calculation of the coupling matrix element.

This paper describes a method to calculate T_{DA}, Δ_D, and Δ_A for any extended-Hückel or tight-binding Hamiltonian with a very large number of orbitals. The method is particularly useful for large systems because it generates the coupling matrix element without requiring diagonalization of a large Hamiltonian matrix followed by the calculation of a very small energy matrix element by adding large numbers of different sign [see Eq. (16)]. The method builds up the decaying tunneling matrix element as a product of decay factors for each "block" of the bridge. This is advantageous because it allows computer experiments to determine the effects of specific theoretical protein mutations and provides comparison between single pathway calculations and more detailed ones.

Schrödinger equation approaches to the T_{DA} calculation rely on the ability to calculate eigenstates of the bridge and to sum their highly oscillatory coefficients with sufficient accuracy to generate the coupling matrix element that decays approximately exponentially with the distance separation between donor and acceptor sites [Eq. (16)]. The generation of a decaying function is a generic aspect of the Green's function method which is not lost if errors are introduced to the calculation of specific ϵ's. Finally, the numerical method that utilizes Dyson's equation progressively adds sites to the bridge Hamiltonian and recalculates the Green's function after each addition. Hence a relatively small amount of information must be stored during the calculation, especially for systems with limited interconnectivity (see Sec. II). The computational demands of conventional eigenvalue and eigenvector evaluation scale as N^3, where N is the number of eigenstates. The Green's function strategy scales as a a^3N, where a is average number of local connections.

II. THE STEPWISE GREEN'S FUNCTION METHOD

Consider the general one-electron tight-binding bridge Hamiltonian:

$$H = \sum_i \alpha_i a_i^\dagger a_i + \sum_{ij>i} v_{ij}(a_i^\dagger a_j + a_j^\dagger a_i). \tag{7}$$

Each orbital can be coupled to any other one, and atomic, bonding/antibonding, or molecular orbitals can be used as the basis.[11(b)] As described in Sec. I, several authors have applied the Green's function approach to the electron transfer problem.[9] Their calculations, however, were limited to specific kinds of bridge. The method described here is general for any one-electron tight binding Hamiltonian and can be applied to electron transfer in proteins and other complex systems.

We now calculate the bridge Green's function matrix elements needed for the two-level T_{DA} calculation and the effective donor and acceptor energies. The bridge consists of N orbitals. We begin our calculation with orbital one and continue adding one new orbital in each step of the calculation. G^i is the Green's function after the ith orbital is included. This is related to the $(i-1)$th orbital Green's function by Dyson's equation

$$G^i = G^{i-1} + G^{i-1} V^i G^i, \tag{8}$$

where

$$H^i = \sum_{j=1}^{i} \alpha_j a_j^\dagger a_j + \sum_{j=1,k>j}^{i} v_{jk}(a_j^\dagger a_k + a_k^\dagger a_j), \tag{9}$$

$$V^i = H^i - H^{i-1} - \alpha_i a_i^+ a_i \text{ and } G_{ii}^{i-1} = (E-\alpha_i)^{-1}. \tag{10}$$

The calculation begins with $i=1$ (only one site), and the final Green's function is obtained when $i=N$. T_{DA} is calculated from the Green's function matrix elements. In the first step, there is just one orbital, so $G_{11}^1 = (E-\alpha_1)^{-1}$. In the following steps, the matrix elements for a given value of i are

$$G_{ii}^i = \left[E - \alpha_i - \sum_{m,k=1}^{i-1} v_{im} G_{mk}^{i-1} v_{ki}\right]^{-1}, \tag{11a}$$

$$G_{ij}^i = G_{ii}^i \left(\sum_{m=1}^{i-1} v_{im} G_{mj}^{i-1}\right), \quad j \neq i, \tag{11b}$$

$$G_{nj}^i = G_{nj}^{i-1} + G_{ii}^i \left(\sum_{m=1}^{i-1} G_{nm}^{i-1} v_{mi} \delta_{ij} + \sum_{m,k=1}^{i-1} G_{nm}^{i-1} v_{mi} v_{ik} G_{kj}^{i-1}\right), \quad n \neq i. \tag{11c}$$

In these calculations, the orbital numbering is unimportant and $G_{ij} = G_{ji}$.

T_{DA} is calculated from the matrix elements G_{lm} for the bridge orbitals l and m that are coupled to the donor and/or acceptor. The other matrix elements are unnecessary for the T_{DA} calculation. For this reason, we do not need to calculate G_{lm}^i for all orbitals l and m at each step. Only the matrix elements G_{lm}^i for orbitals l and m coupled to the donor and/or acceptor, or to any orbital that has not yet been included, are needed. This makes the computation much less numerically demanding than direct diagonalization of the full Green's function or Schrödinger equation. A schematic representation of the elimination procedure for a square lattice bridge is shown in Fig. 1.

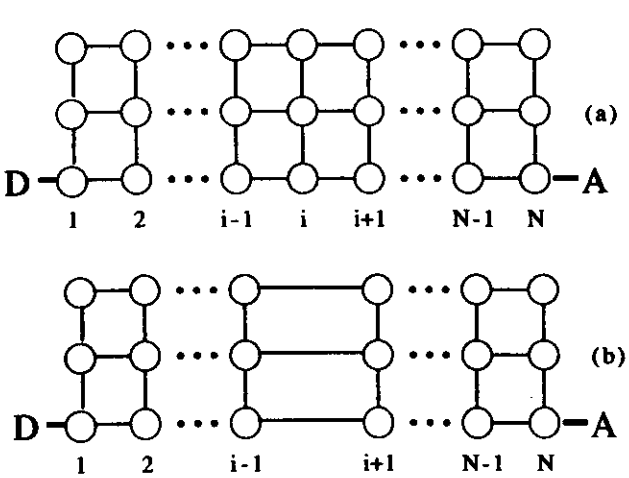


FIG. 1. (a) Schematic representation of a square lattice bridge; (b) after all orbitals of column $i-1$, i, and $i+1$ are included, the orbitals of column i can be eliminated because they are not coupled to any other orbital not yet included in the calculation or to the donor or acceptor.

III. INTERPRETATION OF THE METHOD AND ITS NUMERICAL ADVANTAGES

The two-level electronic state description of electron transfer is appropriate because the donor and acceptor are weakly coupled in the long distance electron transfer problem. Electronic couplings between pairs of orbitals separated by vacuum decay roughly exponentially with distance and have a decay related to the binding energies of the states.[5(d)] Interactions mediated by intervening orbitals have a considerably softer distance decay. It has been shown in a variety of models that the direct electronic coupling between distant sites (greater than about 5–10 Å) is orders of magnitude smaller than the intervening orbital mediated interaction.[1,5]

Typical exponential coupling decay constants for orbital mediated interactions (the decay of T_{DA}) in proteins are 0.5–0.8 $Å^{-1}$ in the tunneling matrix element (1.0–1.6 $Å^{-1}$ in the rate),[5,10] much smaller than the direct through space coupling constants. Hence the electronic structure is well described by including interactions between groups within a rather small cutoff radius. This simple observation, mirrored in many electronic structure calculations, is central to the strategy presented in Sec. II: stepwise introduction of coupled orbitals. As an example of the range of the interactions, covalent and hydrogen bond connections within a segment of α-helix are relatively strong, but interconnections between different helical segments are relatively weak.[10(b)] Indeed, the weak dominantly through space coupling between segments of helix, compared to other secondary structural motifs, recently led to qualitative prediction of the range of electronic couplings in proteins with distinct secondary and tertiary structure.[10(b)]

Methods used to calculate T_{DA}[5,7,9,13] utilize either perturbation theory (and, implicitly, a pathway concept) or "exact" one-electron calculations on a limited portion of the protein judged to mediate the coupling.[13] These methods can be understood to include coupling along a main physical pathway corrected to some order by the effect of wave function backscattering in the system. (This backscattering amplitude can be visualized as shown in Fig. 3, Sec. IV.) Reliable predictions can be made about the propagation of amplitude in a periodic bridge because the bridge Bloch states ($\epsilon - E$ relations of Ref. 5) include all orders of backscattering. Such calculations, applied directly to (finite) model compounds, provide reasonable approximations because of their proper treatment of backscattering within the section of bridge between donor and acceptor. Each order of backscattering is associated with a perturbation theory term in the coupling.[5(c),14]

When applying the infinite chain results to finite chains the error introduced is very small compared to completely neglecting backscattering. The infinite chain calculations yield $\epsilon^\infty(E)$, the energy dependent decay per bridge repeat unit.[5] While finite calculations can be carried out, $\epsilon^\infty(E)$ provides an excellent approximation to the bridge mediated decay and shows bridge symmetry effects. To demonstrate the effects of finite bridge length on ϵ consider a linear bridge consisting of identical orbital repeat units with interaction v between nearest neighbors. The decay factor–energy relation is[5]

$$\epsilon^{\infty}_{\text{exact}}(E)+\frac{1}{\epsilon^{\infty}_{\text{exact}}(E)}=\frac{E}{v}. \tag{12}$$

The finite chain length equivalent of the decay factor ϵ^{∞} between bonds i and $i+1$ is

$$\epsilon^{\text{finite}}_{\text{exact}}(E)=\frac{G^{i+1}_{1,i+1}}{G^{i}_{1,i}}, \tag{13}$$

where the Green's functions are defined in Eq. (11) [Ref. 9(d) solves the finite linear chain with one orbital per unit and nearest neighbor coupling]. This is the ratio of T_{DA} for donor and acceptor attached to bridges of length i and $i+1$, and it converges rapidly to the infinite chain limit Eq. (12) with differences of order $(v/E)^{2i}$. For typical organic molecules, very few bonds are needed in the bridge for $\epsilon^{\text{finite}}_{\text{exact}}$ and $\epsilon^{\infty}_{\text{exact}}$ to be nearly identical.

The lowest order (l.o.) perturbation theory expression for the decay across a site is

$$\epsilon^{\text{finite}}_{\text{l.o.}}=\frac{v}{E}. \tag{14}$$

The exact value of ϵ_i falls between the infinite chain limit (upper limit) and the lowest perturbation theory limit (lower limit). For most organic bridges, the infinite chain limit is superior to the l.o. approach for estimating ϵ_i. ϵ^{∞} produces nearly exact ratios of matrix elements for the finite chain. The deficiency of the lowest order perturbation result is that it does not include any scattering between orbitals (see Fig. 3). It is not necessary to include very high order scattering terms; corrections that arise from scattering in the "nearby" bonds are adequate for excellent estimates of ϵ_i. For this reason, the infinite chain approximation works very well for estimates of ϵ_i in the corresponding finite periodic chains.

Even for finite aperiodic chains, the decay of the wave function can be written exactly as a product of ϵ_i's.[5,9(d)] Use of Eq. (6) to map tunneling pathways, which writes the decay through a protein as a product, was motivated by this fact. For a linear aperiodic chain, which approximation is best to compute ϵ_i, the lowest order perturbation or the infinite form? A very good approximation is to use the analog of the infinite chain result for each orbital in the chain,

$$\epsilon_i+\frac{1}{\epsilon_i}=\frac{E-\alpha_{i+1}}{v_{i+1,i}}. \tag{15}$$

This approximation works for two reasons. Backscattering is important but only nearby bonds are relevant. Therefore, the distant bonds in a long periodic system do not make important contributions to ϵ_i. Second, although α_i varies for different bond types, these fluctuation are generally small relative to an average $E-\alpha$.

We have shown the importance of backscattering when computing tunneling matrix elements through pathways. Two other points are yet to be discussed: how side groups appended to the main pathway modify it and, more importantly, how physical pathways interfere with one another. A side group is defined as a collection of atoms bound to the physical pathway at a single orbital (see Fig. 4 in Sec. IV). A collection of atoms bound to the main pathway by two or more atoms creates multiple physical pathways. The intersections between physical pathways create loops (see Fig. 5 in Sec. IV). The side groups can be easily eliminated; they renormalize the site energy of the orbital in the main pathway to which they are bound (a self-energy correction).[9(d)] The physical pathway remains intact, but some of the orbital energies are slightly modified by the side groups. Loops, however, create alternative interfering pathways. Questions concerning multiple pathways are addressed in Sec. IV, and this new method will allow their further study.

To include all of the physical pathways, side chains, and loops in complex bridges naturally requires numerical calculations. The conventional numerical approach to calculating the tunneling matrix element of the corresponding two-level system by matrix diagonalization is to compute the eigenstates of the isolated bridge and, from them

$$T_{DA}=\sum_{ij}v_{Di}\left[\sum_{\nu}\frac{c_{i,\nu}c_{j,\nu}}{E-\epsilon_{\nu}}\right]v_{jA}, \tag{16}$$

where ϵ_ν is the energy of bridge molecular orbital ν. This is identical to Eq. (5) where the Green's function is written in a basis diagonal in the eigenstates of the isolated bridge. Equation (16) is the expression evaluated in one-electron calculations of T_{DA}.[7,13(a)] As mentioned in Sec. I, the tunneling energy E is determined by the vibronic coupling on the donor and acceptor.

While Eq. (16) is valid in principle, it is difficult to calculate the summation reliably for a large number of delocalized bridge states. The source of this problem is that a large number of oscillating coefficients are added to give a very small number. The numerical problems with the summation get worse as the tunneling energy moves further from the bridge states because the energy denominators of all terms in the sum are the same order but the sign of the numerator varies. The new Green's function method, however, captures the roughly exponentially decaying nature of T_{DA}. The exponential decay arises because the coupling between sites (v_{ij} in the tight-binding Hamiltonian) is only appreciable for orbitals close in space. The method described in Sec. II exploits the local connectivity that gives rise to T_{DA}. This makes the numerical method described in Sec. II much more accurate than methods that do not reflect this fact. Also, because the orbitals are only locally connected, the strategy discussed at the end of Sec. II minimizes the number of intermediate Green's function matrix elements that need to be retained during the calculation. Using the elimination procedure discussed in Sec. II, the number of orbitals at any stage of the calculation is much smaller than the total number of orbitals in the protein. This should make the calculation of tunneling matrix elements in very large proteins possible.

The numerical advantage of the stepwise Green's function method vs eigenvector/eigenfunction evaluation and sum over states is shown in a model problem. Consider a chain of N orbitals with the same site energy ($\alpha=0$) coupled to nearest neighbors and calculate G_{1N} with the two approaches as a function of chain length, N. The calculation is performed for $E/v=-2.5$ ($\epsilon^{\infty}=-0.5$). In the one orbital calculation, the bridge states extend from $2/v$ to $-2/v$ and the energy zero is the site energy of the bridge basis

orbital. Figure 2 shows the results for the two calculations using single and double precision arithmetic. The wave function coefficients for the finite bridge $c_{i,\nu}$ are known analytically so there are no errors associated with matrix diagonalization in this case. It is intriguing that the sum on states expression for the Green's function matrix element [Eq. (16)] is intractable after fewer than 100 states for typical values of ϵ^{∞}, even in double precision. Using the Green's function method, we find very small numerical errors in just the single precision calculations for more than 300 states, and errors appear in the fourth or fifth significant figure.

If all of the coupling energies between orbitals in the Green's function calculation are small compared to the energy difference between the tunneling energy and the orbital energies, numerical difficulties arising from division by very small numbers will be avoided. For example, these conditions are met with bonding/antibonding orbitals but not necessarily with simple atomic orbitals.[11] In the latter case, the tunneling energies may be extremely close to the bridge (atomic) orbital energies.

IV. THE PATHWAY CONCEPT

Our recent effort to compute T_{DA} for proteins uses the concept of a physical tunneling pathways (see Sec. I). The pathway concept is useful if a "small" number of paths dominates the tunneling matrix element. In the limit of many pathways, changes in a specific one will have a small effect on the rate. Therefore, if a large number of pathways is important, no specific details of the protein will influence the rate. In this section, we describe how to define effective pathways and the conditions that must be met so that a small number of them determines the tunneling matrix element. We show how the method presented here can be used to predict these effective pathways, and their contribution to the tunneling matrix element.

As discussed in the previous section, a physical pathway has many scattering pathways associated with it. For example, if we consider the donor coupled to the acceptor via a single physical pathway, and we compute the tunneling matrix element to lowest order in perturbation theory, this is equivalent to neglecting scattering corrections. The scattering corrections for a given pathway arise from enumerations of orbitals in the tunneling pathway longer than the shortest path from donor to acceptor. The importance of backscattering can be observed by considering a chain of N identical orbitals coupled by nearest neighbor interactions. $G_{1,N}$ can be computed exactly using Eq. (11), but to clarify our discussion we present it as a series in powers of $(v/E)^2$ (v is the coupling between nearest neighbors and α is set to zero, the origin of the energy scale)

$$G_{1,N} = \frac{v^{N-1}}{E^N}\left\{1 + (N-1)\left(\frac{v}{E}\right)^2 + [(N-1) + (N-2) + \frac{(N-1)(N-2)}{2}]\left(\frac{v}{E}\right)^4 + \cdots\right\}. \quad (17)$$

Figure 3 shows some of the scattering pathways to order $(v/E)^4$. The important conclusion is that although the first correction to the Green's function goes as $(N-1)(v/E)^2$, each individual orbital is only responsible for $(v/E)^2$. The reason for this is that there are $N-1$ scattering pathways giving corrections of this order, but each of them involves a different scattering orbital. Equation (14) gives the exact expression for ϵ_{∞}, and we see that for $|E/v| > 2.27$ ($|\epsilon_{\infty}| < 0.6$) it converges to the infinite chain limit (better than 5% accuracy) just by including scattering with first and second nearest neighbors. For this reason, finite chains are well approximated with infinite chain values for ϵ. Scattering is important, but only nearby orbitals have an effect.

We now describe the effect of pendant (side) chains on T_{DA} and on the pathway concept. Figure 4 shows that pathways including the side chains must be included when computing the tunneling matrix element, but that their effect can be incorporated with a self-energy correction to the pendant group attachment site. Assuming that the side chain is cou-

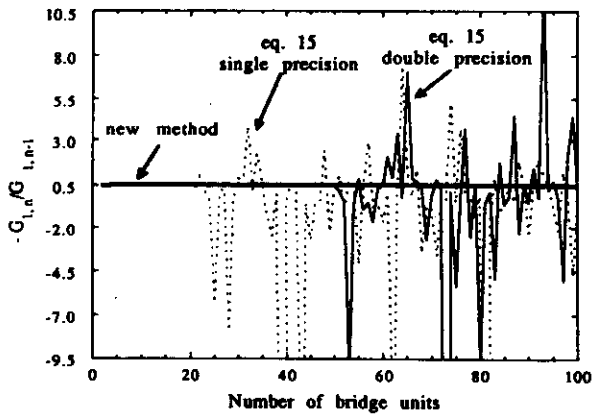


FIG. 2. Calculation of $G_{1,N}/G_{1,N-1}$ for $E/v = -2.5$, ($\epsilon_{\infty} = -0.50$) for the new method described here and the standard matrix diagonalization method of Eq. (15). Plots using Eq. (15) are displayed for both single and double precision calculations. The results from the new method for long chains converge to the infinite chain limit without any noticable numerical differences. In the one orbital per repeat unit system there is an analytical result for $G_{1,N}/G_{1,N-1}$ as discussed in the text.

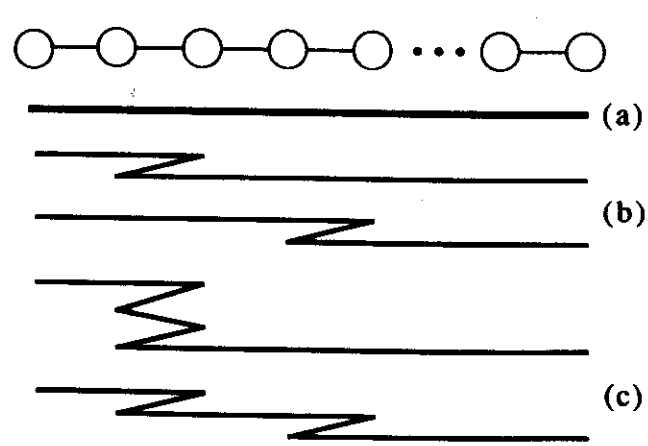


FIG. 3. Scattering pathways for a linear bridge (only one physical pathway). (a) Main pathway; (b) examples of pathways that generate corrections of order $(v/E)^2$ to the coupling; (c) examples of pathways that generate corrections of order $(v/E)^4$ to the coupling.

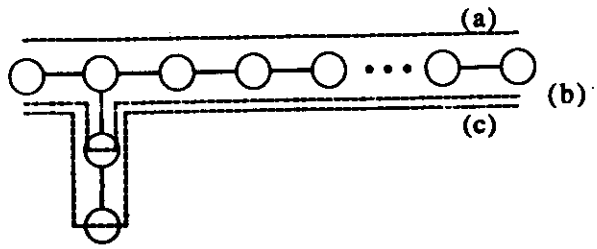


FIG. 4. Pathways in a bridge with a side group. (a) Main pathway; (b) example of one pathway that generates corrections of order $(v/E)^2$ to the coupling; (c) example of one pathway that generates corrections of order $(v/E)^4$ to the coupling.

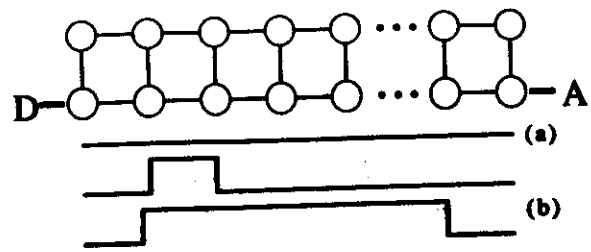


FIG. 5. Loop pathways for a square lattice composed of two lines. (a) Main pathway; (b) examples of pathways that generate corrections of order $(v/E)^2$ to the coupling.

pled to pathway orbital i via orbital $s1$, the side chain can be eliminated by renormalizing the orbital i energy.

$$\alpha_i^{\text{eff}} = \alpha_i + v_{i,s1} G^{\infty}_{s1,s1} v_{s1,i}. \tag{18}$$

$G^{\infty}_{s1,s1}$ is the diagonal matrix element of the Green's function for site $s1$ when *only* the side chain is included in the Hamiltonian. Using this procedure, many side chains can be immediately eliminated at the start of the calculation, greatly simplifying the problem.

The effects discussed to this point, scattering in a single physical pathway and in the side chains, can be incorporated in the pathway concept. The effective pathway is described using effective site energies. The tunneling matrix element takes into account scattering pathways and side chains by modifying the self-energies of sites on the physical pathway. This situation becomes more difficult when loops of orbitals connect two atoms in a physical pathway.

When loops exist, the identification of specific physical pathways may be less useful from the standpoint of identifying dominant contributions to T_{DA}. To address this point concretely, we consider a square lattice bridge (see Fig. 1). This bridge has many loops because it is highly interconnected. As discussed in Sec. II, the exact solution for G is much simplified by column elimination. For the initial discussion, a bridge composed of just two parallel interconnected rows with equal coupling between neighboring sites is sufficient (Fig. 5). In analogy with the treatment for a bridge composed of a single row, we can expand $G_{1,N}$ in powers of $(v/E)^2$

$$G_{1,N} = \frac{v^{N-1}}{E^N}\left\{1 + \left[(N-1) + N + \frac{N(N-1)}{2}\right]\left(\frac{v}{E}\right)^2 + \left[\frac{9N^2+5N-14}{2} + \frac{N(N-1)(N-2)}{6} + \frac{N(N-1)(N-2)(N-3)}{24}\right]\left(\frac{v}{E}\right)^4 + \cdots\right\}. \tag{19}$$

Loops that make contributions to $G_{1,N}$ of order $(v/E)^2$ are shown schematically in Fig. 5. The lowest order contributions from loops enter as N^2, rather than linearly in N. Figure 6 shows a plot of $G_{1,N}(E)$ in several limits. All of these plots are made for $v = -0.4$ eV and $E = 2.0$ eV ($v/E = -0.2$). For these values, some corrections become of order unity or larger for $N > 7$. We see in Fig. 6 that near this value the approximate calculations become a poor estimate. This small value of v/E was chosen so that the effect of backscattering would be small. The figure shows the importance of including loops for highly interconnected bridges. In such cases, there are no dominant pathways. A particularly interesting result, showing how loops can destroy the relevance of specific pathways, is shown in Figs. 7(a) and 7(b). Here the bridge is composed of two rows of ten orbitals each. The central bond in the lower row (main pathway) of Fig. 7(a) is eliminated and in Fig. 7(b) the central bond of the upper row is eliminated. Performing the Green's function calculation we obtain

$$\frac{G^{a}_{1,N}}{G_{1,N}} \sim 0.4 \tag{20a}$$

and

$$\frac{G^{b}_{1,N}}{G_{1,N}} \sim 0.6, \tag{20b}$$

where $G_{1,N}$ is the exact result if no bonds are eliminated. If single pathways were dominant, these results would be ap-

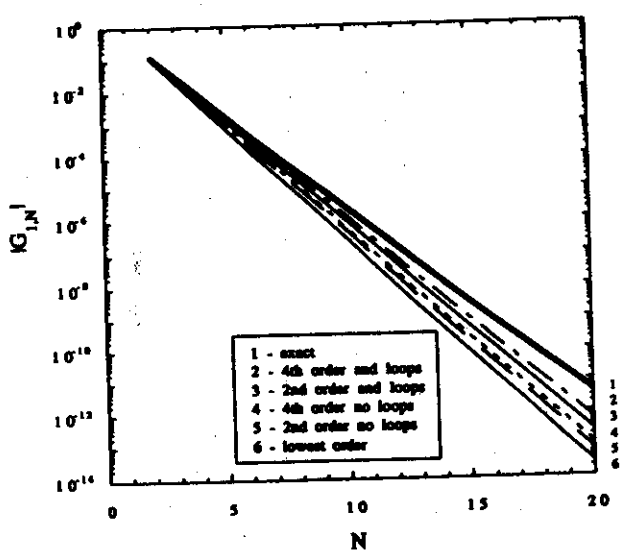


FIG. 6. $|G_{1,N}(E)|$ (eV^{-1}) for the square lattice bridge of Fig. 5 in several limits. The absolute value is plotted because $G_{1,N}$ alternates in sign. All of these plots are made for $v = -0.4$ eV and $E = 2.0$ eV, so $v/E = -0.2$. For the values used here, some corrections becomes of order unity or larger for $N > 7$. Notice that around this length the approximate calculations become a poor estimate. This small value of v/E was chosen so that the effect of backscattering would be small, and most of the corrections would arise from loops.

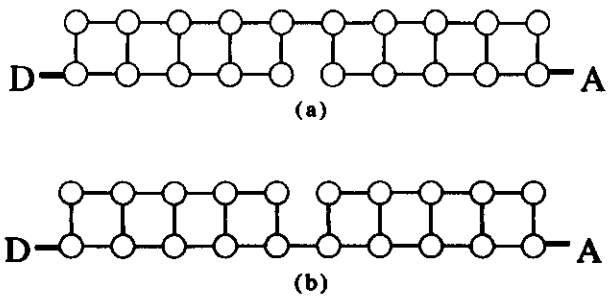


FIG. 7. Square lattice bridge for $N = 10$ with one bond eliminated. (a) Central bond in lower row (main pathway) eliminated; (b) Central bond in upper row eliminated.

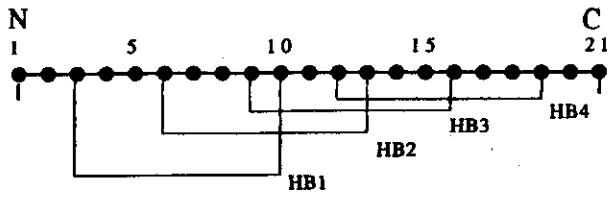


FIG. 8. Schematic representation of an alpha-helical bridge for $N = 7$ amino acids. Each covalent bond is represented by just one orbital in the calculation ($E/v = -2.27$, $\epsilon_\alpha = -0.6$). Hydrogen bonds are treated as two normal covalent bonds.

proximately 0.04 and 1.0, respectively. By eliminating one bond in the lower row, the main pathway is increased by two orbitals, and by eliminating one bond in the upper row, the main pathway remains the same. The discrepancy between the exact result and the "pathway" result occurs because the bridge is highly interconnected, i.e., it contains a large number of loops. In this case, it is not productive to discuss contributions to T_{DA} in terms of single pathways.

When more rows are added to the bridge, no substantial change occurs. Rows further from the one that the donor and acceptor are bound to are responsible for still higher order corrections. [The lowest order corrections for row i is proportional to $(v/E)^{2i}$]. For this reason, a fixed chain length N causes $G_{1,N}$ to approach a limiting value as the number of parallel rows is increased. This is expected as long as the tunneling electron energy remains in the gap between the highest occupied and lowest unoccupied molecular orbitals of the bridge. Recall that for an infinite one chain bridge with equal coupling between neighboring sites the bridge states occur between $2v$ and $-2v$ (corresponding the group of bonding or antibonding states in a molecular system). An infinite 2D bridge of the connectivity shown in Fig. 1 has twice the bandwidth. Therefore, energies that were associated with localized states in the 1D case may be in the band for the 2D bridge. As rows are added to the 2D bridge, there is a transition between these two limiting widths, but as long as the tunneling energy is associated with a localized state, $G_{1,N}$ saturates after a few rows are included.

From earlier work (see the discussion in Sec. V), we noted that tunneling matrix elements in proteins can be dominated by few or many pathways depending on the secondary and tertiary structure of the protein.[10(b)] Theoretical and experimental work aimed at defining when the pathway concept is valid continues.[10-12] This new method permits the quantitative determination of when the tunneling matrix element is indeed dominated by a few pathways and how protein secondary and tertiary structure might influence this. As an example of the potential of the method in this regard, we performed a calculation on a bridge composed by N simple model amino acids in an alpha-helix for different values of N. The bridge is shown schematically in Fig. 8. Each bond is represented by one orbital and coupling between neighbor backbone orbitals is equal.[2(b),5] Hydrogen bonds are represented by two additional bonds between sites i and $i+3$. Calculations were performed for $E/v = -2.27$ ($\epsilon_\alpha = -0.6$). The results are presented in Table I, which shows the importance of the hydrogen bonds. $G_{1,3N-1}$ for $N = 3m+1$ with m integer is dominated by a short pathway involving the hydrogen bonds. This conclusion can be demonstrated by disconnecting some bonds in the chain for $N = 10$, see Table II. Elimination of hydrogen bonds in the main pathway reduces the coupling about 1 order of magnitude.

V. CONCLUSIONS

An open question in the field of protein electron transfer is whether or not a small number of physical pathways between donor and acceptor dominates the tunneling matrix element. The answer to this question will depend on the level of connectivity (loops) in the specific protein. The concept of a pathway is not invalidated by the presence of multiple nonintersecting paths, since their contributions to the tunneling matrix element enter as a simple sum.

TABLE I. Value of $G_{1,3N-1}$ (eV^{-1}) for a bridge of N amino acids in an alpha-helix. A single orbital per bond model is used. $E/v = -2.27$ ($\epsilon_\alpha = -0.6$) and $E = 2.0$ eV.

Number of amino acids N	Linear path (No H bonds)	Best path	Exact result
1	-0.273	-0.273	-0.529
2	$+0.563\times10^{-1}$	$+0.563\times10^{-1}$	$+0.732\times10^{-1}$
3	-0.121×10^{-1}	-0.121×10^{-1}	-0.103×10^{-1}
4	$+0.026\times10^{-1}$	$+0.202\times10^{-1}$	$+0.202\times10^{-1}$
5	-0.056×10^{-2}	-0.435×10^{-2}	-0.519×10^{-2}
6	$+0.121\times10^{-3}$	$+0.938\times10^{-3}$	$+2.269\times10^{-3}$
7	-0.026×10^{-3}	-1.565×10^{-3}	-1.182×10^{-3}
8	$+0.056\times10^{-4}$	$+3.371\times10^{-4}$	$+4.396\times10^{-4}$
9	-0.121×10^{-5}	-7.264×10^{-5}	-19.512×10^{-5}
10	$+0.026\times10^{-5}$	$+12.117\times10^{-5}$	$+8.581\times10^{-5}$

TABLE II. Value of $G_{1,3N-1}$ (eV^{-1}) for a bridge of $N = 10$ amino acids in an alpha-helix. Parameters are discussed in the text and Table I.

Exact result	Removing H-bond 1	Removing H-bonds 1,4	Removing H-bonds 1,4,7
$+8.58\times10^{-5}$	$+2.99\times10^{-5}$	$+1.95\times10^{-5}$	-0.70×10^{-5}

Previous work has applied a graph search method to map physical pathways and estimate their contribution to the tunneling matrix element. The new method described here provides a way to calculate the contribution to T_{DA} for each physical pathway, and to examine the role of loops, which has not been thoroughly investigated.

Pathway searches in heme, blue copper, and iron-sulfur proteins have suggested that the density of pathways and the prevalence of loop structures varies from protein to protein.[10a,10b] For example, in cytochrome *c* the number of physical pathways with significant contributions to the tunneling matrix element is relatively small and the number of loop structures is also rather limited.[10] This contrasts with the case of electron transport through a glassy medium where the large number of intersecting physical pathways should eliminate the possibility of single dominant ones. In proteins such as cytochrome *c*, the pathway model is expected to be quite useful for interpreting experimental tunneling matrix elements and intervening protein structure may substantially affect the rate. Indeed, such specific effects have been observed experimentally and are consistent with the existing theory. The new Green's function method described here can be used to calculate the coupling for physical pathways that have been predicted. It can investigate the importance of side groups attached to the main pathways (suppressed in the simple model), loops, and multiple pathways.

The stepwise Green's function approach provides a numerically robust strategy for calculating tunneling matrix elements in very large systems. In contrast with existing methods,[5,7,13] one does not have to select "important" residues to include in an energy splitting calculation. Also, the influence of specific orbitals or contacts on the tunneling matrix element can be probed by performing the calculation with and without these orbitals or interactions in the Hamiltonian. For example, the Green's function method can test the proposal that hydrogen bonds form essential connections in the dominant physical tunneling pathways of ruthenated cytochrome *c*.[10a]

ACKNOWLEDGMENTS

We thank Arnóbio da Gama for useful discussions and careful reading of this manuscript. J. N. O. and D. N. B. thank the National Science Foundation and CNPq (Brazil) for a Binational Research Grant that allowed international visits during which this work was initiated and the Brazilian agencies FINEP and CNPq for additional support. P.C.P.A. is supported by a CNPq fellowship. This work was performed in part at the Jet Propulsion Laboratory, California Institute of Technology and was sponsored by the Department of Energy's Catalysis/Biocatalysis Program, Advanced Industrial Concepts Division, through an agreement with the National Aeronautics and Space Administration. Work in San Diego was funded by the National Science Foundation (Grant No. DMB-9018768) and a research contract from the Jet Propulsion Laboratory, supported by the Department of Energy's Catalysis/Biocatalysis Program. J. N. O. is in residence at the Instituto de Física e Química de São Carlos, Universidade de São Paulo, 13560, São Carlos, SP, Brazil during the summers.

[1] (a) R. A. Marcus and N. Sutin, Biochim. Biophys. Acta **811**, 265 (1985); (b) M. D. Newton and N. Sutin, Ann. Rev. Phys. Chem. **35**, 437 (1984); (c) *Photoinduced Electron Transfer*, edited by M. A. Fox and M. Chanon (Elsevier, Amsterdam, 1988), Vols. A–D; (d) *Metal Ions in Biological Systems*, Vol. 27, edited by H. Sigel and A. Sigel (Marcel Dekker, New York, 1991); (e) M. D. Newton, Chem. Revs. (in press).

[2] D. N. Beratan and J. N. Onuchic, in *ACS Adv. Chem. Ser., Electron Transfer in Inorganic, Organic, and Biological Systems*, Vol. 228, edited by J. Bolton, G. L. McLendon, N. Mataga (American Chemical Society, Washington, D.C., in press).

[3] (a) J. N. Onuchic, and P. G. Wolynes, J. Phys. Chem. **92**, 6495 (1988); (b) W. Bialek, W. J. Bruno, J. Joseph, and J. N. Onuchic, Photosynth. Res. **22**, 15 (1989); (c) J. N. Onuchic, J. Chem. Phys **86**, 3925 (1987); (d) J. N. Onuchic, D. N. Beratan, and J. J. Hopfield, J. Phys. Chem. **90**, 3707 (1986).

[4] (a) J. Halpern and L. E. Orgel Discuss. Faraday Soc. **29**, 32 (1960); (b) H. M. McConnell, J. Chem. Phys. **35**, 508 (1961); (c) J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **71**, 3640 (1974).

[5] (a) D. N. Beratan and J. J. Hopfield, J. Am. Chem. Soc. **106**, 1584 (1984); (b) D. N. Beratan, *Ibid.* **108**, 4321 (1986); (c) J. N. Onuchic and D. N. Beratan, *Ibid.* **109**, 6771 (1987); (d) D. N. Beratan, J. N. Onuchic, and J. J. Hopfield, J. Chem. Phys. **83**, 5325 (1985).

[6] R. A. Marcus, Chem. Phys. Lett. **133**, 471 (1987).

[7] (a) S. Larsson, J. Am. Chem. Soc. **103**, 4034 (1981); (b) S. Larsson, J. Chem. Soc. Faraday Trans. 2 **79**, 1375 (1983); (c) A. Broo and S. Larsson, Int. J. Quant. Chem. Quant. Biol. Symp. **16**, 185 (1989).

[8] (a) E. N. Economou, *Green's Functions in Quantum Physics*, 2nd ed. (Springer, New York, 1983); (b) S. Doniach and E. H. Sondheimer, *Green's Functions for Solid State Physicists*, revised ed. (Benjamin/Cummings, Reading, 1978).

[9] (a) A. A. S. da Gama, Theor. Chim. Acta **68**, 159 (1985); (b) M. A. Ratner, J. Phys. Chem. **94**, 4877, 1990; (c) S. H. Lin, J. Chem. Phys. **90**, 7103 (1989); (d) A. A. S. da Gama, J. Theor. Biol. **142**, 251 (1990); (e) C. Goldman, Phys. Rev. A **43**, 4500 (1991); (f) Y. Magarshak, J. Malinsky, and A. D. Joran (preprint 1990).

[10] (a) D. N. Beratan, J. N. Onuchic, J. N. Betts, B. E. Bowler, and H. B. Gray J. Am. Chem. Soc. **112**, 7915 (1990); (b) D. N. Beratan, J. N. Betts, J. N. Onuchic, Science (in press); (c) D. N. Beratan, J. N. Onuchic, H. B. Gray, in *Metal Ions in Biological Systems*, Vol. 27, edited by H. Sigel and A. Sigel (Marcel Dekker, New York, 1991), p. 97; (d) J. A. Cowan, R. K. Upmacis, D. N. Beratan, J. N. Onuchic, and H. B. Gray, Ann. New York Acad. Sci. **550**, 68 (1988).

[11] (a) D. N. Beratan, J. N. Onuchic, and Hopfield, J. J. J. Chem. Phys. **86**, 4488 (1987); (b) J. N. Onuchic and D. N. Beratan, *Ibid.* **92**, 722 (1990); (c) D. N. Beratan and J. N. Onuchic, Photosynth. Res. **22**, 173 (1989).

[12] (a) B. E. Bowler, T. J. Meade, S. L. Mayo, J. H. Richards, and H. B. Gray, J. Am. Chem. Soc. **111**, 8757 (1989); (b) M. J. Therien, M. A. Selman, I.-J. Chang, J. R. Winkler, H. B. Gray, Ibid. **112**, 2420 (1990); (c) M. J. Therien, B. E. Bowler, M. A. Selman, and H. B. Gray, in *ACS Adv. Chem. Ser., Electron Transfer in Inorganic, Organic, and Biological Systems*, Vol. 228, edited by J. Bolton, G. L. McLendon, and N. Mataga (American Chemical Society: Washington, D.C., in press).

[13] (a) P. Siddarth and R. A. Marcus, J. Phys. Chem. **94**, 2985 (1990); (b) H. E. M. Christensen, L. S. Conrad, J. Ulstrup, and K. V. Mikkelsen, in *Metal Ions in Biological Systems*, Vol. 27, edited by H. Sigel and A. Sigel (Marcel Dekker, New York, 1991), p. 57; (c) P. Siddarth and R. A. Marcus, J. Phys. Chem. **94**, 8430 (1990).

[14] L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1968).

47

Electron Transfer

From Model Compounds to Proteins

David N. Beratan[1] and José Nelson Onuchic[1,2]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109
[2]Department of Physics, University of California at San Diego, La Jolla, CA 92093 and Instituto de Física e Química de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil

We summarize the formulation of the protein-mediated electronic coupling calculation as a two-level system with weakly interacting bridge units. Using model compounds as a starting point from which to derive coupling parameters, we present a strategy for defining the pathways for electron tunneling in biological and biomimetic systems. The specific bonding and nonbonding interactions in cytochrome c *and myoglobin that mediate the tunneling between the porphyrin and an attached transition metal probe are described. The method appears to succeed where traditional structureless tunneling barrier or periodic bridge models are not adequate. An algorithm to search for these tunneling pathways in proteins is described, and the nature of the paths is discussed.*

THE PROCESS OF ELECTRON TRANSPORT IS CENTRAL in chemistry, biology, and physics. This field is frequently subjected to detailed reanalysis and review (*1–4*). We begin the discussion here by presenting the Hamiltonian that has been used extensively to model the generic electron-transfer problem

$$H_{ET} = H_{rp}\sigma_z + \frac{1}{2}\delta\sigma_z + H_Q \qquad (1)$$

[3]Current address: Department of Physics, University of California at San Diego, La Jolla, CA 92093

0065–2393/91/0228–0071$06.00/0

H_{rp} is the tunneling matrix element between donor and acceptor (reactants and products); σ_x and σ_z are the Pauli matrices, where the eigenvalue $\sigma_z = 1$ is associated with the reactant state and the eigenvalue $\sigma_z = -1$ is associated with the product state; H_Q supplies the dynamics for the nuclear motion (reaction coordinates and bath), and δ is the instantaneous energy difference between the reactants and products (*5–8*). This Hamiltonian leads to the ubiquitous rate equation for transfer between weakly coupled donors and acceptors (*1–10*).

$$k_{ET} = \frac{2\pi}{\hbar} |H_{rp}|^2 \text{ (FC)} \qquad (2)$$

Assuming that this separation can be performed and that the process is not relaxation-controlled, the rate is proportional to the electronic coupling factor $|H_{rp}|^2$ times a nuclear Franck–Condon (FC) weighted density of states (activated) factor; $\hbar$ is Planck's constant divided by 2π.

Equation 2 gives the rate in the weak coupling limit, often called the nonadiabatic limit. Two important conditions must hold to write this equation. First, an energy separation is required to reduce the problem to the Hamiltonian given by eq 1 (i.e., a two-level system coupled to nuclear modes; renormalization procedure). A separation of electronic energies is also required so that the electronic problem can be reduced to a two-level (donor and acceptor) system. Second, even when the Hamiltonian of eq 1 is valid, the transfer must be nonadiabatic to write the rate in eq 2 (*1–10*). The nonadiabatic limit is valid for the model systems and proteins discussed in this chapter. Next we discuss why the simple Hamiltonian (eq 1) is appropriate for such complex problems.

Bridge-Mediated Electron Tunneling and Two-Level Systems

First, consider the electronic part of this problem. Because of the complexity of proteins, we hope to reduce it to smaller appropriate parts (if possible) that can be analyzed and understood. This is achieved by gradually eliminating higher energies. The first step in this procedure is to assume that the energies involved in chemical bonding are very small compared to core-electron excitations. This assumption allows the elimination of the core electrons. The core electrons provide a pseudopotential in which the valence electrons move. Next we make the further assumption that coupling energies associated with hopping between neighboring bonding orbitals are small compared to atomic excitation energies. This assumption leads to a tight-binding molecular orbital picture (*11*).

These assumptions are generally valid for the electron-transfer problem and are adopted throughout this chapter (*4, 5*). This modification justifies an effective one-electron Hamiltonian and permits computation of the tunneling matrix element H_{rp}. Finally, to reduce the one-electron Hamiltonian of the entire system to a two-level Hamiltonian, the electronic energy separation between donor (or acceptor) and bridge sites must be much larger than the coupling energy between donor and acceptor. If this is not the case, there are electronic excitations with energies of the same order as the donor–acceptor coupling, invalidating a two-level approach. [These energy comparisons are best made with bond orbitals rather than atomic orbitals (*12*).]

We now include the vibrational modes. If the energy scales associated with excitations of these modes are much smaller than the electronic excitation energies, we can use the Born–Oppenheimer approximation. This approximation allows us to solve the electronic problem for fixed nuclear coordinates, so the nuclear coordinates enter as parameters. A two-level system then results, with energies that are functions of nuclear coordinates. The tunneling matrix element is calculated by fixing the nuclear coordinates so that the reactant and product states have the same energy (Condon approximation) (*2, 4, 5*). If all of the energy separations discussed here are appropriate, the problem is reduced to eq 1. References 5 and 13 describe details of the electronic–nuclear energy separation. A tutorial showing the reduction of a three-level system to a two-level system is given in ref. 5. Electronic excitation energies are about 1–3 eV, so this separation is valid for most of the nuclear modes.

Next, we include the high-frequency nuclear modes (C=O stretches, for example, with $\hbar\Omega$ in the range 0.1–0.25 eV; Ω is the frequency of the mode). In this case, $\hbar\Omega$ is much larger than other vibrational excitation energies and k_BT (k_B is Boltzmann's constant and T is the temperature). These modes, which typically arise from local vibrations, have a nearly discrete spectrum [very low damping (*14*)], which should be treated in the quantum limit; they simply renormalize the tunneling matrix element and the driving force (*7*). For example, in a two-level system with thermodynamic driving force δ_0 coupled to one high-frequency mode, $|n_D\rangle$ ($|n_A\rangle$) represents the vibrational state of the high-frequency mode when the electron is on the donor (D) or acceptor (A). (The equilibrium position of this high-frequency mode shifts, depending on whether the electron is on the donor or acceptor.) Because $k_BT << \hbar\Omega$, the donor vibrational state is always $|0_D\rangle$. One of the acceptor states $|n_A\rangle$ will dominate the process, depending on δ_0. The renormalized parameters are

$$H_{rp}^{eff} = H_{rp} \langle 0_D | n_A \rangle \qquad (3a)$$

$$\delta_0^{eff} = \delta_0 - n_A\hbar\Omega \qquad (3b)$$

The effective donor state is $|D^{el}\rangle|0_D\rangle$, and the effective acceptor state is $|A^{el}\rangle|n_A\rangle$ (el signifies an electronic state). Finally, if the electronic excita-

tions are of the same order as $\hbar\Omega$ and the reorganization energy is a few times $\hbar\Omega$, the renormalization procedure is a bit different and we must construct an energy cutoff that includes the electronic states and the high-frequency mode. We cannot separate this process into two stages as we did before (electronic part first, then the high-frequency mode). The final result is very similar to the present one (i.e., a single donor state and a set of discrete acceptor states), but these states will be mixtures of electronic and high-frequency nuclear states rather than simple products.

Energy separation is not the only requirement for the validity of the Born–Oppenheimer approximation. Although energy separation guarantees that we can neglect the donor (or acceptor) excited electronic states, care is required when computing the tunneling matrix element that depends on details of the electronic wave function tail. Formally, as the electron moves from donor to acceptor it spends an imaginary time (a traversal time) in the forbidden region (*15*). If the nuclear modes are slow compared to this time, the Born–Oppenheimer approximation works (i.e., the nuclei stay essentially fixed as the electron tunnels). The traversal time increases with the tunneling distance and decreases with the tunneling barrier height (*5*). For very-long-range transfer the Born–Oppenheimer approximation must break down, but this approximation is reasonable for the systems discussed here (*4*, *16*).

To this point, we have described why the Hamiltonian in eq 1 is, in many cases, an appropriate starting point for the electron-transfer problem. We will now describe how to obtain the two-level representation of the electronic portion of the problem for bridged systems. The questions to be addressed are (1) What are these two states in a complex bridged system? (2) How is the coupling H_{rp} between the two states related to energy splittings that are obtained from electronic structure calculations?

The simplest example of bridge-mediated electron transfer in a tight-binding or molecular orbital model results in the Hamiltonian of eq 4. The donor and acceptor are only coupled by their mutual interactions with one bridge (B) orbital via the exchange interactions β_{DB} and β_{BA}. The Hamiltonian matrix in this case ($\alpha_B > \alpha_D, \alpha_A$) is

$$H = \begin{pmatrix} \alpha_{D(\vec{y})} & \beta_{DB} & 0 \\ \beta_{DB} & \alpha_B & \beta_{BA} \\ 0 & \beta_{BA} & \alpha_{A(\vec{y})} \end{pmatrix} \tag{4}$$

The nuclear coordinates are represented by $\vec{y}$, and eq 4 is written in the Born–Oppenheimer approximation. The $\vec{y}$ dependence of the site energies reflects the separation between electronic and nuclear motion and the assumption that only the donor and acceptor orbitals are coupled to nuclear distortions. Because the donor and acceptor (unmixed) orbitals are degenerate at the crossing of the nuclear surfaces, $\alpha_D(\vec{y}) = \alpha_A(\vec{y}) = \alpha$ (Condon

approximation). The symmetric–antisymmetric splitting (ΔE) between the two lowest states localized dominantly on donor and acceptor (eq 4) is

$$\Delta E = \sqrt{\frac{(\alpha_B - \alpha)^2}{4} + \beta_{DB}^2 + \beta_{BA}^2} - \frac{(\alpha_B - \alpha)}{2} \simeq \frac{(\beta_{DB}^2 + \beta_{BA}^2)}{(\alpha_B - \alpha)} \tag{5}$$

This splitting is nonzero even when there is no donor–bridge or bridge–acceptor coupling! Contrary to the common claim, this splitting is not proportional to H_{rp}. The resolution of this issue arises from the fact that we have calculated the splitting between the wrong states. The splitting in eq 3 is nonzero because it includes contributions of pure donor–bridge and bridge–acceptor mixing. The net bridge-mediated donor–acceptor interaction, H_{rp}, is not the splitting between states in the overall Hamiltonian with $\alpha_D = \alpha_A$. Also, $2|H_{rp}|$ is the energy associated with mixing the donor *plus* bridge state with the acceptor *plus* bridge state (i.e., the splitting of the states in the corresponding two-level system). The splittings between eigenvalues of the full Hamiltonian of eq 4 are not directly related to H_{rp}. From the standpoint of perturbation theory, the donor–acceptor degeneracy at the crossing point of the nuclear surfaces is broken only in second order by the bridge, so that the coupling between the states in this order is $-\beta_{DB}\beta_{AB}/(\alpha_B - \alpha)$ (*17*). Only in a true two-site model is there direct equivalence between ΔE and $2|H_{rp}|$. Also, strictly speaking, the orbital energies that were made equivalent in the Condon approximation should be the energies of the two-level system, not the individual site energies of the donor and acceptor.

A general technique to reduce a bridged donor–acceptor system to the corresponding two-level system is Löwdin diagonalization (*18*, *19*). Working in a basis diagonal in the bridge orbitals, the total Hamiltonian is

$$H = \left(\begin{array}{cc|ccccc} \alpha_D & \beta_{DA} & \beta'_{D1} & \cdots & \cdots & \beta'_{DN} \\ \beta_{AD} & \alpha_A & \beta'_{A1} & \cdots & \cdots & \beta'_{AB} \\ \hline \beta'_{1D} & \beta'_{1A} & \alpha_{B1} & 0 & \cdots & 0 \\ \beta'_{2D} & \beta'_{2A} & 0 & \alpha_{B2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \beta'_{ND} & \beta'_{NA} & 0 & 0 & \cdots & \alpha_{BN} \end{array}\right) \tag{6}$$

Primes denote interactions between a single atomic orbital and a molecular orbital and N is the number of bridge orbitals. Unprimed interactions are between single atomic orbitals. The exact corresponding two-level Hamiltonian is

$$H = \begin{pmatrix} \alpha_D - \sum_i \left[\frac{\beta'^2_{Di}}{\alpha_{Bi} - E}\right] & \beta_{DA} - \sum_i \left[\frac{\beta'_{Di}\beta'_{Ai}}{\alpha_{Bi} - E}\right] \\ \beta_{DA} - \sum_i \left[\frac{\beta'_{Di}\beta'_{Ai}}{\alpha_{Bi} - E}\right] & \alpha_A - \sum_i \left[\frac{\beta'^2_{Ai}}{\alpha_{Bi} - E}\right] \end{pmatrix} \quad (7)$$

The off-diagonal elements in eq 7 are the electron tunneling matrix elements of the corresponding two-level system. The tunneling energy E is determined by the diagonal energies (these are donor plus bridge and acceptor plus bridge energies) and the vibronic coupling in the molecule (a simple average is appropriate, for example, if the vibronic coupling on the two sites is identical) (*4*).

There are other methods of calculating tunneling matrix elements in bridged systems. An elegant method that is experiencing growing interest is the Green's function technique. The matrix elements of the bridge Green's function contain the effective coupling between sites in the bridge (*20–22*). Numerical techniques applicable to Green's functions are somewhat different from those usually applied in a Schrödinger equation approach, and some powerful theorems allow both exact and perturbation evaluation of the couplings for tight-binding Hamiltonians. The Green's function for a system, G, is defined by Dyson's equation:

$$(E - H)G = 1 \quad (8)$$

If the Green's function of the isolated bridge is given by $\tilde{G}$, the donor is coupled to bridge orbitals i with strength β_{Di}, and the acceptor is coupled to sites n with strength β_{nA}.

$$H_{tp} = \beta_{DA} + \sum_i \sum_n \beta_{Di}\tilde{G}_{in}\beta_{nA} \quad (9)$$

$\tilde{G}_{in}$ describes the propagation of amplitude within the bridge from site i to site n; β_{DA} is the direct "through-space" donor–acceptor coupling and can generally be neglected relative to the bridge-mediated terms for distant electron transfer.

Information Learned from Model Compounds

Donor plus bridge and acceptor plus bridge states are needed for a two-level calculation of H_{tp}. As such, techniques that calculate this mixing reliably

were the first targets of study. Ab initio techniques are now being successfully applied to relatively small bridged electron-transfer model compounds (*23–25*) and idealized systems (*26*). Our approach has relied on one-electron and effective potential methods because these methods are adequate for addressing issues of tunneling energy dependence and bridge topology effects and because it is possible to perform these calculations in very weakly coupled systems without serious concern about basis set artifacts. Qualitative issues related to through-bond and through-space coupling are addressed conveniently with carefully parameterized exactly soluble square barrier models (*27*).

The generic results of the bridge studies are summarized in Figure 1. Most bridges can be "reduced" to chains of interacting pairs of orbitals with two characteristic interactions. The details of the reduced orbitals are determined by the topology of the chain and energetics of the bonds in the bridge (*28*). Tunneling through a bridge of such repeating units where the mixing into the bridge is weak and decay is rapid enough (decay per bridge unit squared is small compared to 1, not a very stringent condition) allows H_{tp} to be written as in eq 10. Writing the decay of H_{tp} per bond as ϵ (*12, 28–32*)

$$H_{tp} = \frac{\beta_A\beta_D\beta_1}{(E - \alpha_L^1)(E - \alpha_R^1) - \beta_1^2} \prod_{i=1}^{N} \epsilon_i \quad (10a)$$

Neglecting backscattering between bonds,

$$\epsilon_i \simeq \left[\frac{\beta_i\gamma_i}{(E - \alpha_{iL})(E - \alpha_{iR}) - \beta_i^2}\right] \quad (10b)$$

For $|\epsilon| > 0.4$, corrections for backscattering must be incorporated in the calculation of ϵ itself (*12, 31*). Here, E is the tunneling energy, L and R refer to the left and right hybrid atomic orbitals in the bonds, $(N + 1)$ is the total number of bonds in the bridge, β is the interaction within bonds, γ is the interaction between bonds, α is the energy of the (hybrid) orbitals in the bonds, and β_D and β_A are the coupling matrix elements between the donor and acceptor and the first and last bridge units, respectively. As an example, in a linear extended hydrocarbon chain $\gamma/\beta \simeq 0.25$ and $\beta \simeq -9$ eV. Equations 10a and 10b are generalized in the next section for the case in which the bond types in the bridge may be chemically different.

Most of the electron-transfer model compounds aimed at testing the distance dependence of the transfer rate are of the form DB_nA, where n is variable. The potential in such linkers is, to a good approximation, periodic (*12, 28–30*). The boundary conditions on the periodic potential contain the details of the donor and acceptor structure, but the periodic nature of the

bridge allows relatively simple calculations to make predictions about the energy and symmetry dependence of the coupling within broad classes of linkers. These predictions, which typically include more details than were used to calculate eqs 10a and 10b, are reliable as long as the decay within the bridge is sufficiently rapid and the net mixing onto the bridge is weak. Predictions for σ-bond-coupled electron transfer included pointing out the enhanced mediation properties of bridges with convergent pathways of equal length, such as exist in corner-fused rings (vs. edge-fused rings) and other effects (*28–30*). Although the theoretical calculations seem to be in fair agreement with experiment, there are several questions begging to be addressed synthetically.

1. For fixed reaction free energy, ΔG, but donor and acceptor energies varied in an absolute sense, will the decay length of H_{rp} change the parameter $\beta/2$ in $H_{rp} \propto \exp[-R\beta/2]$ (where R is the donor–acceptor separation distance)? Does a hole or electron-transfer mechanism dominate in chemical systems?

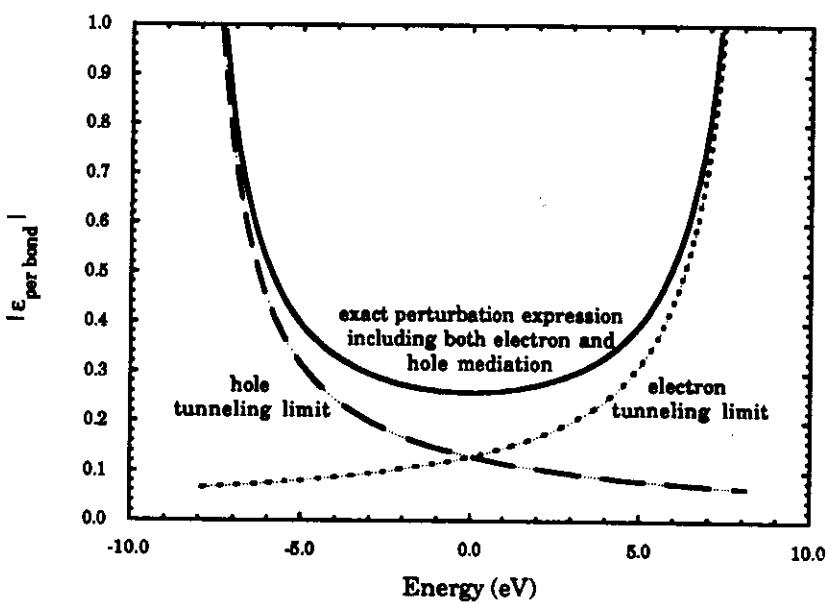


Figure 1a. ϵ (decay per bond) vs. E plots are shown for a C–C chain with $\beta = -8.5$, $\alpha_C = 0$, *and* $\gamma_C = -2.2$ *eV. The infinite chain result (U-shaped curve) is shown (28, 29), as well as the hole- and the electron-mediation limits (eq 13). The approximate curves are adequate in energetic regimes expected for typical model compounds (*$|\epsilon| \sim 0.4$–0.6*).*

2. For fixed donor and acceptor but varied bridge, will the net coupling show the predicted topological effects (*28–30*)?
3. In saturated systems coupling π-donors, does σ or π symmetry coupling into the bridge dominate the net interaction, H_{rp}?
4. How important are hydrogen bonds for mediating electron transfer? Surely there is a role for model building here. Is the picture of hydrogen bonds as preferentially assisting hole mediation (*12*) accurate?
5. How costly are the symmetry demands of σ/π interactions in proteins? Do π groups assist transfer or not? Our current thinking is that the π systems must be aligned in special ways for significant enhancements.
6. The distance dependence of ΔG and λ (reorganization energy) complicate the interpretation of bridge and tunneling energy dependence studies because these parameters cannot be held fixed with transfer distance. Can ΔG and λ studies be per-

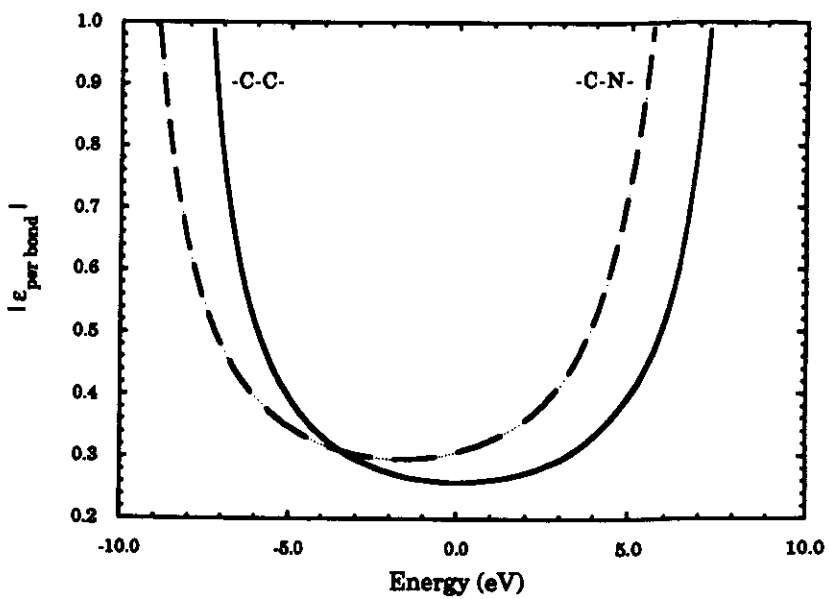


Figure 1b. The energy dependence of ϵ is shown for a C–C vs. C–N chain with $\beta = -8.5$, $\alpha_N = -3.3$, $\gamma_N = -3.1$, *and* $\gamma_C = -2.2$ *eV. The C–N plot is centered at lower energy because of the greater electron affinity of nitrogen. The "U" shape of the curves is characteristic, where ϵ shows distinct electron- and hole-mediation regimes.*

formed as a function of distance to unambiguously deconvolute the bridge structure dependence of the rates?

Answers to these questions are within synthetic and spectroscopic reach, but obtaining them will require a coordinated effort.

The model compounds and the theoretical studies have taught us about the typical length scales for decay of tunneling interactions in saturated and unsaturated organic bridges. Counting bonds along the shortest path from donor to acceptor in well-characterized model compounds suggests that the decay of H_{rp} is about a factor of 0.4–0.6 per bond. Reference 3 summarizes experimentally measured values of these parameters and their dependence on structural details of the bridge. We have learned from the theory that "decay per bond" strategies work rather well for these typical decays, with some qualifications (*12*, *28–30*). The propagation of the donor and acceptor states can be built up by sequentially introducing single bonds (or groups of bonds in strongly delocalized systems) to the chain of orbitals. Following the addition of each bond, the amplitude leaking onto it is calculated as a 2×2 problem. Interference effects can be treated within this strategy (*12*, *28–32*) if intersecting pathways bearing similar amplitudes are handled carefully.

Protein-Mediated Electron Transfer Interpreted with Decay-per-Bond Methods

Although intriguing questions remain in the model compound area, our aim in pursuing that work was to learn how to piece together and parameterize a model for protein-mediated electron transfer in photosynthetic and respiratory reactions. A *physical tunneling pathway* is defined as a collection of interacting bonds in a protein around and between the donor and acceptor that make some contribution to the donor–acceptor interaction. A few specific physical pathways may or may not dominate the electronic coupling between donor and acceptor. Whether a relatively small number of pathways is adequate to describe the coupling in proteins is actually a deep theoretical issue. We argue (on the basis of rapid decay of through-space interactions for typical tunneling energies, the relatively low density of residues, and the anisotropic packing of bonds) that a relatively small number of pathways is likely to be important.

The decay-per-bond approach leads to eq 11 for the contribution to the tunneling matrix element arising from a single physical pathway with N_B covalent couplings between bridge bonds, N_S through-space contacts, and N_H hydrogen bonds (*12*, *31*, *32*)

$$t_{da} = \frac{\beta_A \beta_D \beta_1}{(E - \alpha_L^1)(E - \alpha_R^1) - \beta_1^2} \prod_{i=2}^{N_B} \epsilon_i^B \prod_{j=1}^{N_S} \epsilon_j^S \prod_{k=1}^{N_H} \epsilon_k^H \qquad (11)$$

β_D or β_A couples the donor or acceptor into the first or last bond of the bridge, respectively; t_{da} is the contribution to H_{rp} arising from a single pathway; and β_1 is the coupling between orbitals in the first bond. Values for ϵ can often be approximated by using perturbation theory. As an example, the lowest order contribution to ϵ is given by eq 10b. This limit totally neglects backscattering between bonds, and corrections to it need to be included for large ϵ. For a particular interaction, ϵ can be dissected into contributions from electron and hole mediation across a bond (*12*) as

$$\epsilon = \epsilon^e + \epsilon^h \qquad (12)$$

In the limit where hole mediation through the bond dominates, for example, and the two coupled covalent bonds are the same

$$\epsilon \simeq \epsilon^h = \frac{\gamma}{E - \alpha_{bond}} \qquad (13)$$

One can also write the propagator G_{1M}, which is proportional to H_{rp}, for a donor coupled to site 1 and acceptor coupled to site M, as

$$G_{1M} \propto \prod_{i=1}^{M} \epsilon_i \qquad (14a)$$

The exact expressions for ϵ_i can be written (*20*, *21*)

$$\epsilon_i = \frac{\gamma_i}{E - \alpha_i - \Delta_i} \qquad (14b)$$

where Δ_i is a site-energy correction that takes into account the influence of all residues off (as well as on) the physical pathway between sites 1 and M (*20*, *21*). Strategies that include the influence of all of the higher order corrections to the coupling neglected in our decay-per-bond (eq 11) formulation exist for the calculation of the Δ values. The exciting aspect of these methods is that they provide a way of interpreting the impact of specific residues anywhere in the protein on the coupling between two sites. The challenge now is to implement calculations of Δ values and related quantities for realistic but tractable protein Hamiltonians such as that of ref. 32. This approach still neglects interference between physical pathways. A new approach that includes multiple interacting pathways has been developed and will be used to test the present assumption that a few pathways dominate the coupling in many proteins.

Software that will include the calculation of interaction parameters combined with a search algorithm using these strategies is under development in our group and in other groups that are using somewhat different ap-

proaches (33). We recently wrote software (34) in an effort to understand the dependence of electron-tunneling mediation in proteins on details of the primary, secondary, and tertiary structure. The software makes the following assumptions:

1. All covalent bonds in the path cause equivalent decay, $\bar{\epsilon}_B$. A typical value of this factor is 0.6. Model compounds would suggest typical per-bond decay factors of 0.4–0.6 (this is the decay of H_{rp} per bond; square it to see the effect on the transfer rate).
2. All through-space interactions have the same orientation prefactor (σ) and decay length (β'), $\epsilon_S = \sigma\bar{\epsilon}_B \exp[-\beta'(R - R_C^{eq})]$ (where R is the through-space distance and R_C^{eq} is the reference covalent separation distance). Typically, σ is fixed at 0.5 for pathway surveying and β' is fixed between 1.7 $Å^{-1}$ (10-eV binding energy for transferring electron) and 1.0 $Å^{-1}$ (5-eV binding energy).
3. Hydrogen bonds couple as strongly as two covalent bonds, when scaled to reference covalent bond lengths, $\epsilon_{HB} = \bar{\epsilon}_B^2 \exp[-\beta'(R - R_{HB}^{eq})]$.
4. Interactions between pathways are neglected during the search. Interference effects due to the addition of amplitude arriving at the acceptor from multiple pathways can be included by summing the contributions independently.

We choose to neglect orientation factors in hydrogen bonds. Discussion of these parameters is found in ref. 12. In an extended-Hückel calculation, the β values in eqs 10a, 10b, and 11 depend on the orbital binding energies. References 12 and 27 show that it is actually more appropriate to use the electron tunneling energy to calculate β. In any case, the pathways are not strongly dependent on the particular chosen tunneling energy, as long as a realistic value is selected (34).

This strategy for pathway mapping intentionally neglects differences among bond types and orientations. The method for including bond differences and angular effects is described in ref. 12. Although angular effects require greater attention, the differences among decay factors for different bond types should not cause gross changes in the pathways. The strategy presented here would be meaningless if the qualitative aspects of the predictions were dependent on fine details of the decay parameters.

The σ value can be purposely varied to find pathways that exclude through-space segments. The decay factors include the qualitative aspects of the coupling, such as the similarity between covalent and hydrogen-bonded coupling (12), as opposed to through-space coupling. The rough

choice of parameters is sufficient for a qualitative understanding of dominant pathways.

Realistic values of $\bar{\epsilon}_B$ are defined by the resonance integral for the bond and tunneling electron energy relative to the bond energy. As discussed in refs. 27 and 34, typical values of $\bar{\epsilon}_B$ are 0.4–0.6 for the bonds of interest. The value of β, the decay length of the through-space interactions, is determined by the binding energy of the tunneling electron.

The density of physical pathways found is sufficiently low that identification of individual paths is sensible (i.e., key residues can have an impact on the net coupling). The limitations of these admittedly simplistic assumptions are discussed in refs. 28–32. Now we justify, or at least explain, the approach. Square barrier models of protein-mediated electron tunneling (9, 35) are cruder than the calculation described here because they neglect the atomic graininess and the inhomogeneity of the bridging medium (36, 37). The present model includes these features. However, all of these simple models contain the essential physics of the electron-mediation problem and have provided excellent guidance for designing and interpreting experiments. We are confident that the models presented here will be supplanted by less naive ones in the future. In the meantime, we hope that they will allow the rational design of target proteins for site-directed mutagenesis and semi-synthesis-based electron-transfer studies, along with the interpretation of experimental results not anticipated by existing structureless barrier tunneling models.

Using these assumptions, we searched for the physical pathways with a graph search algorithm (34) in well-characterized proteins with known donor–acceptor couplings and transfer rates. We focus here on ruthenated myoglobin and cytochrome *c* (38–40) because these systems are so well-defined and well-characterized, and the coupling is clearly polypeptide-mediated (41–57). Detailed discussion of this work is presented in ref. 34. Here we will review some of the qualitative conclusions of the ruthenated protein studies and present the strongest evidence in support of the pathway search method.

The pathway search algorithm is not sensitive to special orientation or aromatic residue effects. For this reason, it is useful to look at the family of best pathways to draw qualitative conclusions and to compare relative path lengths between isomers for the best paths found. Myoglobin is a highly helical protein. The best family of pathways (34) between His 81 and the porphyrin are shown in Figure 2. Pathways follow the α-helix from the His to the porphyrin. In the His 12 derivative, the physical paths are roughly orthogonal to two portions of α-helix between His and porphyrin. Important paths follow prominent secondary structures only to the extent that they provide rather direct connections between donor and acceptor. In myoglobin, there seem to be abundant "good" pathways differing from one another in only minor ways. Hence, induction of a rate change in myoglobin by

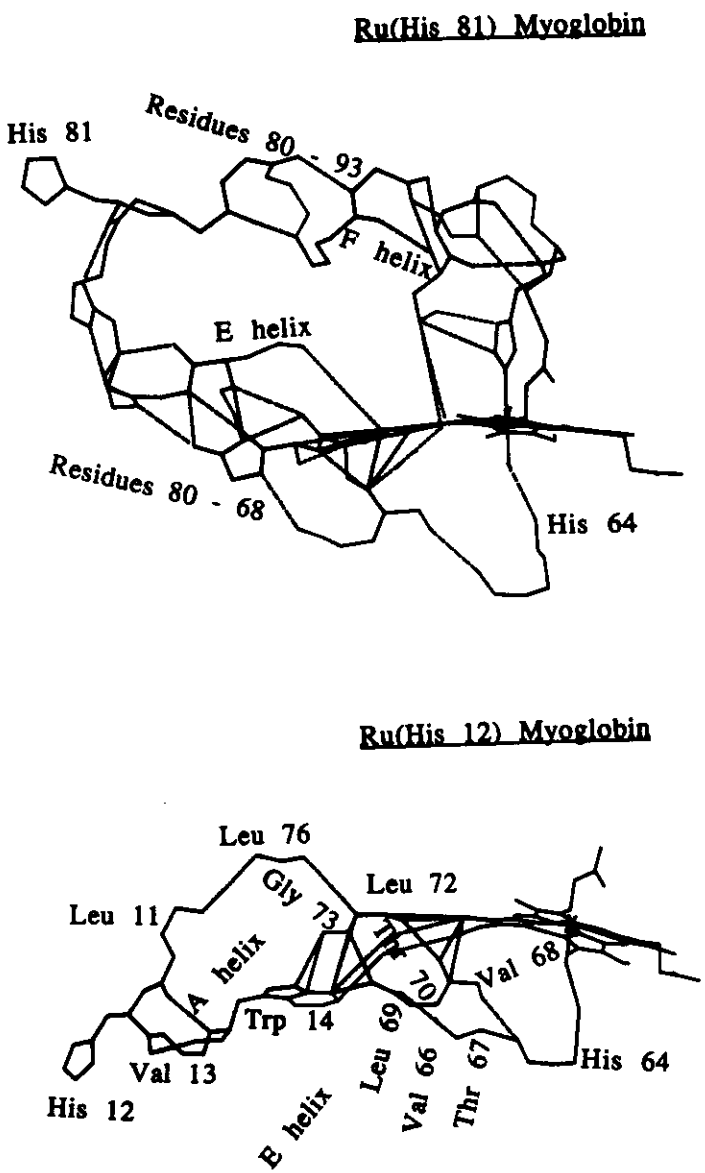


Figure 2. Top: The best family of pathways (34) between His 81 and the porphyrin are shown (myoglobin). The paths follow the α-helix from the His to the porphyrin and the through-space connections onto the ring. Bottom: In the His 12 derivative, the Trp 14 ring bridges most of the pathways between two α helices that are hard to identify because the paths move orthogonally to the helix axes.

changing specific protein atoms or bonds would probably require exquisite planning.

Another interesting aspect of the myoglobin pathway analysis is that the His 116 and His 12 derivatives have through-space contacts in all of their best pathways, as opposed to the His 48 and His 81 derivatives that have purely bond-mediated paths available. This observation led to reanalysis of the experimental data and questioning of whether the quenching in these isomers results from intramolecular electron transfer. The experiments are now being carefully reexamined (*58*).

The relatively large number of paths in myoglobin (hundreds within a factor of 10 coupling of the best) is not found in other proteins. In cytochrome *c* (Figure 3) only about 10–30 strongly coupled pathways are found, most without any through-space connections. This is probably a result of the less helical and less compact nature of cytochrome *c*. The measured rates in cytochrome *c* are known with greater certainty because they are sufficiently fast [the ^{3}Zn–porphyrin experiments in particular (*38–40*)] and provide an interesting study of the utility of the pathway model. Table I reports $\Pi_i \epsilon_i$ for the pathways and the equivalent calculated effective number of sequential covalent bonds.

These effective transfer distances track quite well with the measured rates. However, the measured rates do not track well with structureless-medium models that predict simple exponentially decaying rates proportional to $\exp[-\beta R_{DA}]$. For typical choices of β, the simple exponential scaling is off by at least an order of magnitude for these isomers. This result is the best experimental evidence so far for the importance of the pathways. Not only does the pathway analysis predict the proper ordering of the rates in the three isomers, it also predicts the relative couplings rather well. The factor of 0.6 decay per bond was chosen to give the ratio of His 48/His 81 myoglobin rates of $\sim 10^3$ for paths that differ by approximately five bonds (*41*).

To summarize the results of the pathway analysis, we are finding that (1) there are qualitative differences in the kinds of coupling pathways in different proteins, (2) transfer rates in cytochrome *c* seem to correlate well with the effective number of steps in the pathway but not with the through-space distance, and (3) hydrogen bonds appear to be crucial for linking covalent legs of the paths. Although the method does not differentiate between saturated and aromatic residues, there is no evidence from this family of experiments that aromatic residues provide any special rate enhancement. This finding may reflect the fact that the orbital overlap cost of mixing onto and off such groups can be rather large. Tests of these calculations can be carried out by site-directed mutagenesis of the protein pathways and careful temperature-dependence studies in ranges where fluctuations that facilitate through-space coupling (*28*) interactions (not gating) are shut down.

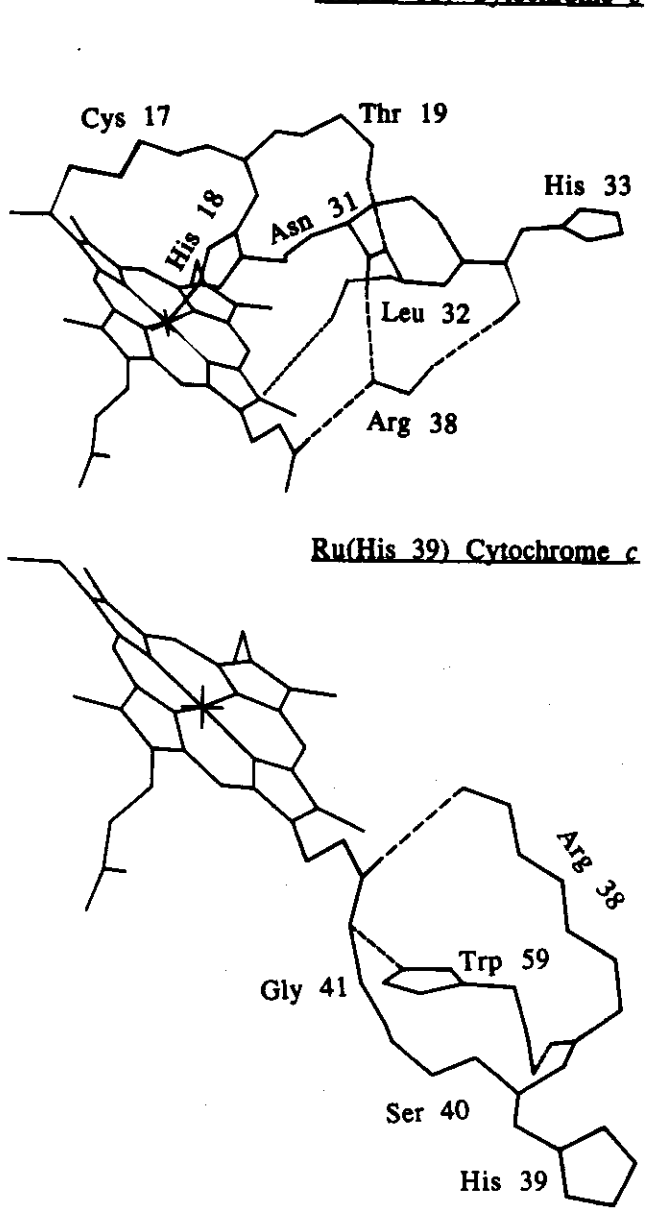


Figure 3. Pathways in His 33 (top), His 39 (bottom), and His 62 modified cytochrome c *(next page). The His 33 derivative is from horse heart, His 39 is from* Candida krusei, *and His 62 is from* Saccharomyces cerevisiae. *Table I correlates the experimental values of the electronic couplings for these isomers with the pathway predictions.*

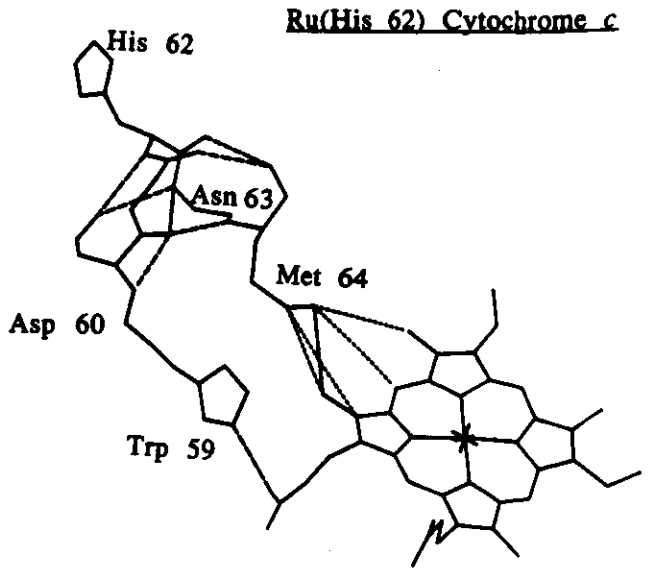


Figure 3. Continued. Pathway in His 62 modified cytochrome c.

Table I. Dominant Paths in Cytochrome *c*

Isomer	*Through-Bond Links*[a]	*Effective Through-Bond Links*	$\Pi_i \epsilon_i$[2b] *(Relative)*	*Relative* $\lvert H_{rp} \rvert^2$ *Fit for* $Ru^{2+} \rightarrow FeP^+$ *and Zn* $P^* \rightarrow Ru^{3+}$ *Reactions*	R_{DA} *(Å)*
His 39[c]	15 (1 H bond)	16.3	664	441	13.0
His 33[c]	18 (1 H bond)	17.6	180	144	13.2
His 62[d]	20 (3 H bond)	22.1	1	1	15.5

NOTE: All results were calculated with $\sigma = 0.5$, $\beta' = 1.7$ Å^{-1}, and counting hydrogen bonds as two through-bond connections from heteroatom to heteroatom. *See* ref. 38 for description of the experiments. None of the isomers had through-space links.
[a]Bonds were counted from Ru to the porphyrin ring edge or to the porphyrin metal atom for paths involving a ligand of the porphyrin metal (His 33 cytochrome *c* only).
[b]Relative coupling squared gives predicted relative transfer rates, assuming equal activation parameters, donor–bridge couplings, and acceptor–bridge couplings.
[c]*C. krusei.*
[d]*S. cerevisiae.*

Conclusions

We conclude with a summary of where the bridge-mediated problem now stands for complex bridges. Typical binding energies of π-electron systems are roughly 6–10 eV. Optical excitation of 2 eV or less decreases this effective binding energy somewhat. However, binding energies like this would result

in values of β, the distance decay of the rate (proportional to the square root of the tunneling energy), of roughly 2–3 $Å^{-1}$, roughly twice that observed in any well-characterized model compound or protein. Apparently, coupling mediated by the bridge dominates the coupling that would result at equal distance in the absence of the bridge. This order-of-magnitude argument has been confirmed by several groups in more detailed calculations. In the weak coupling, relatively rapid decay regime, perturbation-theory approaches are adequate and expressions for the decay per bond in saturated linkers can be estimated. For relevant tunneling energies, these predictions are in very good agreement with calculations that consider the entire bridge. However, predictions of energetic and topological effects on the transfer rates remain to be unambiguously tested.

Experimental and theoretical studies have made rapid progress over the last 5 years. Of particular theoretical interest has been the synthesis and physical study of donor–bridge–acceptor systems with fixed separation distances. The theoretical framework for understanding bridge-mediated coupling in model compounds seems to be well in place; yet many key experimental tests of generic predictions remain to be performed as described in the foregoing discussion. As experimental tests emerge, more detailed theoretical analysis will undoubtedly be warranted.

The situation with proteins is more complex. Site-directed mutagenesis techniques, redox-active protein labeling, semisynthesis techniques, and the solution of the photosynthetic reaction center structure have introduced unanticipated theoretical challenges. Simple questions like, "How well does the ubiquitous hydrogen bond mediate electron transfer?" remain nearly unaddressed. Meanwhile, strategies are still needed to reliably treat coupling mediated by large and complex bridges, as are more systematic experimental tests of the theory. Although the protein tunneling pathway model proposed here seems compelling with existing data, the model makes real predictions that can be directly tested experimentally. Such experimental work might involve the design of systems with particularly weak or strong coupling at fixed transfer distance, analysis of pathway-induced temperature dependences, and tunneling energy effects on the coupling.

Acknowledgments

This work was performed, in part, at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored in part by the Department of Energy's Energy Conversion and Utilization Technologies Division–ECUT, through an agreement with the National Aeronautics and Space Administration. We thank the National Science Foundation and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Brazil) for a binational research grant that allowed international visits during which much of this work was performed and the Brazilian agencies Fianciadora de Estudos e Projetos (FINEP) and CNPq for additional support. We thank our collaborators at Caltech, J. J. Hopfield and H. B. Gray, for many years of exciting collaborative work in this area. We are also grateful to M. J. Therien and H. B. Gray for discussion of the cytochrome *c* and other protein electron-transfer experiments prior to their publication. We also thank E. Canel of Rockefeller University for enjoyable discussions of this problem.

References

1. Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265.
2. Newton, M. D.; Sutin, N. *Annu. Rev. Phys. Chem.* **1984**, *35*, 437.
3. Mikkelson, K. V.; Ratner, M. A. *Chem. Rev.* **1988**, *87*, 113.
4. Onuchic, J. N.; Beratan, D. N.; Hopfield, J. J. *J. Phys. Chem.* **1986**, *90*, 3707.
5. Bialek, W.; Bruno, W. J.; Joseph, J.; Onuchic, J. N. *Photosynth. Res.* **1989**, *22*, 15.
6. Garg, A.; Onuchic, J. N.; Ambegaokar, V. *J. Chem. Phys.* **1985**, *83*, 4491.
7. Onuchic, J. N. *J. Chem. Phys.* **1987**, *86*, 3925.
8. Onuchic, J. N.; Wolynes, P. G. *J. Phys. Chem.* **1988**, *92*, 6495.
9. Hopfield, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3640.
10. Jortner, J. *Biochim. Biophys. Acta* **1980**, *594*, 139.
11. Ballhausen, C. J.; Gray, H. B. *Molecular Orbital Theory*; Benjamin/Cummings: Reading, MA, 1964.
12. Onuchic, J. N.; Beratan, D. N. *J. Chem. Phys.* **1990**, *92*, 722.
13. Joseph, J.; Bialek, W., private communication, 1990.
14. Bialek, W.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5908.
15. Caldeira, A. O.; Leggett, A. J. *Ann. Phys. (N.Y.)* **1983**, *149*, 374.
16. Beratan, D. N.; Hopfield, J. J. *J. Chem. Phys.* **1984**, *81*, 5753.
17. Schiff, L. I. *Quantum Mechanics*, 3rd ed.; McGraw Hill: New York, 1984; Chapter 8.
18. Riemers, J. R.; Hush, N. S. *Chem. Phys.* **1989**, *134*, 323.
19. Larsson, S. *J. Am. Chem. Soc.* **1981**, *103*, 4034.
20. da Gama, A. A. S. *J. Theor. Biol.* **1990**, *142*, 251.
21. de Andrade, P. C. P.; Onuchic, J. N.; Beratan, D. N., unpublished results.
22. Ratner, M. A. *J. Phys. Chem.* **1990**, *94*, 4877.
23. Balaji, V.; Ng, L.; Jordan, K. D.; Paddon-Row, M. N.; Patney, H. K. *J. Am. Chem. Soc.* **1987**, *109*, 6957.
24. Falcetta, M. F.; Jordan, K. D.; McMurry, J. E.; Paddon-Row, M. N. *J. Am. Chem. Soc.* **1990**, *112*, 579.
25. Farazdel, A.; Dupuis, M.; Clementi, E.; Aviram, A. *J. Am. Chem. Soc.* **1990**, *112*, 4206.
26. Cave, R. J. ; Baxter, D. V.; Goddard, W. A., III; Baldeschwieler, J. D. *J. Chem. Phys.* **1987**, *87*, 926.
27. Beratan, D. N.; Onuchic, J. N.; Hopfield, J. J. *J. Chem. Phys.* **1985**, *83*, 5325.
28. Onuchic, J. N.; Beratan, D. N. *J. Am. Chem. Soc.* **1987**, *109*, 6771.
29. Beratan, D. N.; Hopfield, J. J. *J. Am. Chem. Soc.* **1984**, *106*, 1584.
30. Beratan, D. N. *J. Am. Chem. Soc.* **1986**, *108*, 4321.
31. Beratan, D. N.; Onuchic, J. N.; Hopfield, J. J. *J. Chem. Phys.* **1987**, *86*, 4488.
32. Beratan, D. N.; Onuchic, J. N. *Photosynth. Res.* **1989**, *22*, 173.
33. Kuki, A.; Wolynes, P. G. *Science (Washington, D.C.)* **1987**, *236*, 1647.

34. Beratan, D. N.; Onuchic, J. N.; Betts, J.; Bowler, B. E.; Gray, H. B. *J. Am. Chem. Soc.*, **1990**, *112*, 7915.
35. Siders, P.; Cave, R. J.; Marcus, R. A. *J. Chem. Phys.* **1984**, *81*, 5613.
36. Davydov, A. S. *Phys. Status Solidi B* **1987**, *90*, 457.
37. McConnell, H. *J. Chem. Phys.* **1961**, *35*, 508.
38. Therien, M. J.; Bowler, B. E.; Selman, M. A.; Gray, H. B.; Chang, I-J.; Winkler, J. R. In *Electron Transfer in Inorganic, Organic, and Biological Systems*; Bolton, J. R.; Mataga, N.; McLendon, G. L., Eds.; Advances in Chemistry Series 228; American Chemical Society: Washington, DC, 1991; Chapter 12.
39. Bowler, B. E.; Meade, T. J.; Mayo, S. L.; Richards, J. H.; Gray, H. B. *J. Am. Chem. Soc.* **1989**, *111*, 8757.
40. Therien, M. J.; Selman, M. A.; Gray, H. B.; Chang, I-J.; Winkler, J. R. *J. Am. Chem. Soc.* **1990**, *112*, 2420.
41. Cowan, J. A.; Upmacis, R. K.; Beratan, D. N.; Onuchic, J. N.; Gray, H. B. *Ann. N.Y. Acad. Sci.* **1988**, *550*, 68.
42. Moore, J. M.; Case, D. A.; Chazin, W. J.; Gippert, G. P.; Havel, T. F.; Powls, R.; Wright, P. E. *Science (Washington, D.C.)* **1988**, *240*, 314.
43. Bowler, B. E.; Raphael, A. L.; Gray, H. B. *Prog. Inorg. Chem.*, in press.
44. Gray, H. B.; Malmström, B. G. *Biochemistry* **1989**, *28*, 7499.
45. Liang, N.; Pielak, G. J.; Mauk, A. G.; Smith, M.; Hoffman, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 1249.
46. Liang, N.; Mauk, A. G.; Pielak, G. J.; Johnson, J. A.; Smith, M.; Hoffman, B. M. *Science (Washington, D.C.)* **1988**, *240*, 311.
47. Elias, H.; Chou, M. H.; Winkler, J. R. *J. Am. Chem. Soc.* **1988**, *110*, 429.
48. Conrad, D. W.; Scott, R. A. *J. Am. Chem. Soc.* **1989**, *111*, 3461.
49. Pan, L. P.; Durham, B.; Wolinska, J.; Millett, F. *Biochemistry* **1988**, *27*, 7180.
50. Durham, B.; Pan, L. P.; Long, J. E.; Millett, F. *Biochemistry* **1989**, *28*, 8659.
51. Farver, O.; Pecht, I. *FEBS Lett.* **1989**, *244*, 379.
52. Jackman, M. P.; McGinnis, J.; Powls, R.; Salmon, G. A.; Sykes, A. G. *J. Am. Chem. Soc.* **1988**, *110*, 5880.
53. Osvath, P.; Salmon, G. A.; Sykes, A. G. *J. Am. Chem. Soc.* **1988**, *110*, 7114.
54. Faraggi, M.; Klapper, M. H. *J. Am. Chem. Soc.* **1988**, *110*, 5753.
55. Hazzard, J. T.; McLendon, G.; Cusanovich, M. A.; Das, G.; Sherman, F.; Tollin, G. *Biochemistry* **1988**, *27*, 4445.
56. Cusanovich, M. A.; Meyer, T. E.; Tollin, G. *Adv. Inorg. Biochem.* **1987**, *7*, 37.
57. Raphael, A. L.; Gray, H. B.; Raphael, A. L.; Gray, H. B. *Proteins* **1989**, *6*, 338.
58. Upmacis, R.K., unpublished results.

RECEIVED for review April 27, 1990. ACCEPTED revised manuscript August 16, 1990.

Reprinted from Advances in Chemistry No. 228
Electron Transfer in Inorganic, Organic, and Biological Systems
James R. Bolton, Noboru Mataga, and George McLendon, Editors